



TESI DOCTORAL

**Títol INGENIERÍAS TECNOLÓGICAS VERSUS INGENIERÍAS
MULTIMEDIA: ESTUDIO DE PERFILES APLICANDO LA
METODOLOGÍA DE LOS SISTEMAS CONSISTENTES DE
CLÚSTERES**

Realitzada per Xavier Senmartí i Senmartí

en el Centre ETSEEI d'Enginyeria i Arquitectura La Salle

i en el Departament d'Electrònica

Dirigida per Dr. Guillem Bou Bauzá

Sinopsi

El present treball de recerca tracta de determinar las diferències acadèmiques existents entre els estudiants d'enginyeries purament tecnològiques i els d'enginyeria multimèdia, que combina tecnologia amb continguts de l'àrea d'humanística i ciències socials. En particular, es restringeix al cas dels plans d'estudi de La Salle, que estan emmarcats en l'àmbit de las tecnologies de la informació i la comunicació, el seu sistema de tutories, l'evolució dels alumnes durant el curs i el rendiment manifestat en les diferents proves d'avaluació.

Davant del poc rendiment de proves estadístiques de correlació i factorització, la novetat metodològica d'aquesta tesi consisteix en la creació d'una metodologia denominada Sistemes consistents de clústers. Aquesta metodologia realitza estudis d'evolució de clústers mitjançant diverses mesures. Es diferencia d'estudis tipus clústers dinàmics per la seva visió transversal i no longitudinal de les variables que han evolucionat en el temps i, a més, considera uns estats de referència separats en el temps, per la qual cosa es basa en un sistema discret de mesures.

La metodologia dels Sistemes consistents de clústers redueix el problema de l'evolució dels clústers resultants en mesures successives, a un problema de cadenes de Markov i a proves simples de significació (Chi quadrat), per tant es pot exportar a recerques de ciències socials al poder ser explicat en termes de proves estadístiques habituals en les recerques d'aquest tipus.

El resultat de la tesi ens mostra com la dinàmica evolutiva dels estudiants d'enginyeries anomenades "tecnològiques pures" és totalment diferent dels estudiants de multimèdia. S'obtenen sistemes de clústers completament diferents encara que comparteixen certa evolució comú a causa de la intervenció correctora i assistencial del Centre en el seu sistema de tutories. De la recerca resulten també directrius per a l'assistència a l'estudiant en aquestes titulacions i es constaten algunes sospites pel que fa al procediment d'aprenentatge que fins ara eren només conjectures.

Resumen

La presente investigación trata de determinar las diferencias académicas existentes entre los estudiantes de ingenierías puramente tecnológicas y los de ingeniería multimedia, que combina tecnología con contenidos del área de humanística y ciencias sociales. En particular, se restringe al caso de los planes de estudio de La Salle, que están enmarcados en el ámbito de las tecnologías de la información y la comunicación, su sistema de tutorías, la evolución de los alumnos durante el curso y el rendimiento manifestado en las diferentes pruebas de evaluación.

Ante el poco rendimiento de pruebas estadísticas de correlación y factorización, la novedad metodológica de esta tesis consiste en la creación de una metodología denominada Sistemas consistentes de clústeres. Dicha metodología realiza estudios de evolución de clústeres a lo largo de diversas mediciones. Se diferencia de estudios tipo clústeres dinámicos por su visión transversal y no longitudinal de las variables que han evolucionado en el tiempo y, además, considera unos estados de referencia separados en el tiempo, por lo que se basa en un sistema discreto de mediciones.

La metodología de los Sistemas consistentes de clústeres reduce el problema de la evolución de los clústeres resultantes en mediciones sucesivas, a un problema de cadenas de Markov y a pruebas simples de significación (Chi cuadrado), por lo que puede exportarse en investigaciones de ciencias sociales al poder ser explicado en términos de pruebas estadísticas habituales en las investigaciones de este tipo.

El resultado de la tesis nos muestra como la dinámica evolutiva de los estudiantes de ingenierías llamadas “tecnológicas puras” es totalmente diferente de los estudiantes de multimedia. Se obtienen sistemas de clústeres completamente diferentes aunque comparten cierta evolución común debida a la intervención correctora y asistencial del centro en su sistema de tutorías. De la investigación resultan también directrices para la asistencia al estudiante en estas titulaciones y se constatan algunas sospechas sobre el procedimiento de aprendizaje que hasta ahora eran sólo conjeturas.

Abstract

The aim of the present research work is to determine the academic differences between technological-engineering students and multimedia students, who combine technology with humanistic and social contents. It specifically involves La Salle study plans, which are focused on information and communication technologies, their tutorial system and the students evolution and their academic performance along the year.

Looking at the low performance of correlation and factorization statistical tests, this thesis presents an innovation in methodology that consists in what is called Systems of clusters. This methodology performs tests in clusters evolution through different kind of measures. The difference between them and dynamic clusters is caused by the transversal and not by the longitudinal vision of the variables that have evolved through time and, moreover, it considers a reference status separately from time, so it is based on a discreet system of measures.

The high potential of this methodology is the advantage of reducing the problem of clusters evolution that comes from successive measures to a problem of Markov chain and simple test of signification (Chi-square), so there is the open possibility to export it to social research as it is possible to be explained in terms of statistical methods usually used in these research fields.

The thesis results show that the dynamic evolution of engineering students of the strong technological branch is totally different from multimedia students. The Systems of clusters obtained are completely different, even though they share some common evolution due to the corrective and mentoring intervention of the University in their tutorial system. From the research, it is also important to highlight the directives obtained in order to assist the student in the degree and some suspects about the learning process are confirmed.

Índice

Sinopsi	I
Resumen	III
Abstract	V
Índice	VII

Capítulo 1: Ingenierías Tecnológicas versus Ingenierías Multimedia: Descripción del Panorama académico 1

1.1 Interés de la multimedia en el mundo empresarial	8
1.2 Descripción de la situación problemática	16
1.3 Aproximación al problema de investigación	18
1.3.1 Precedentes.....	18
1.3.2 Enunciado informal del problema	19
1.3.3 Importancia de la investigación.....	19

Capítulo 2: Bases metodológicas: Análisis del rendimiento en Ingeniería con la metodología orientada al clúster 21

2.1 Precedentes metodológicos: Cluster Dynamics y Cadenas de Markov	23
2.1.1 La idea de evolución en las investigaciones basadas en agrupaciones.....	23
2.1.2 Una idea intuitiva del método (metáfora de las dunas)	24
2.2 Fundamentos conceptuales del enfoque centrado en el clúster.....	25
2.2.1 Definición de sistema consistente biunívoco	25
2.3 Detalle de los aspectos metodológicos	30
2.3.1 Eliminación de clústeres espurios: Tratamiento cualitativo o experiencia del investigador	31
2.3.2 La reordenación con criterios correlacionales.....	33
2.3.3 La intensidad de transiciones basada en el coeficiente de Contingencia.....	34
2.4 Gráficas descriptivas de evolución de los clústeres de un sistema consistente	34
2.5 Método de combinación con cadenas de Markov para la evaluación de la intervención académica institucional	37
2.5.1 Esquema general de las investigaciones realizadas con el presente método	38

Capítulo 3: Definición del problema de investigación 41

3.1 Introducción: una investigación de doble vertiente	43
3.2 Objetivos de la investigación	44
3.3 Enunciado del problema	44
3.5 Formulación del problema de investigación	45
3.6 Consideraciones sobre las variables y las hipótesis derivadas del problema.....	45

Capítulo 4: Aplicaciones básicas de los sistemas consistentes de clústeres. Predicción y seguimiento 51

4.1 Detección de los clústeres predictores en ingeniería multimedia	53
---	----

4.1.1	Descripción de los clústeres	55
4.1.2	Etiquetación de los clústeres	56
4.2	Detección de los clústeres predictores en ingenierías tecnológicas.....	56
4.2.1	Una primera corrección a la agrupación.....	56
4.2.2	La agrupación corregida.....	58
4.2.3	Descripción de los clústeres	60
4.3	Clústeres resultantes en las variables de seguimiento.....	61
4.3.1	Aportación descriptiva de los clústeres de seguimiento	61
4.3.2	Comentario metodológico a los clústeres de seguimiento.....	64
4.4	Aproximación por evolución de clústeres predictores.....	65
4.4.1	Una metáfora conceptual.....	65
4.4.2	Lectura evolutiva de las variables predictoras.....	65

Capítulo 5: Construcción de los sistemas consistentes de clústeres en ingenierías multimedia y tecnológicas..... 71

5.1	Tratamiento de las diferentes variables implicadas en el estudio	73
5.2	Detección de los clústeres de rendimiento en ingeniería multimedia.....	74
5.2.1	Descripción de los clústeres	80
5.2.2	Reordenación de los clústeres de rendimiento	81
5.3	Detección de los clústeres de rendimiento en ingenierías tecnológicas.....	83
5.3.1	Descripción de los clústeres	89
5.3.2	Reordenación de los clústeres de rendimiento	90
5.4	Sistema de clústeres resultante.....	90

Capítulo 6: Estudio descriptivo de los sistemas consistentes de clústeres..... 93

6.1	Objetivo del estudio	94
6.2	Evolución de los sistemas consistentes de clústeres de ingeniería multimedia	94
6.2.1	Estudio preliminar: clústeres predictores y clústeres de rendimiento	94
6.2.2	Análisis de evolución con Crosstabs.....	95
6.2.3	Gráficas de la evolución de los clústeres de rendimiento.....	101
6.2.4	Análisis de correlación directa sobre las notas de matemáticas	104
6.3	Evolución de los sistemas consistentes de clústeres de ingenierías tecnológicas..	105
6.3.1	Estudio preliminar: clústeres predictores y clústeres de rendimiento	105
6.3.2	Análisis de evolución con Crosstabs.....	106
6.3.3	Gráficas de la evolución de los clústeres de rendimiento.....	112

Capítulo 7: Análisis bajo el enfoque de las cadenas de Markov..... 117

7.1	Recapitulación.....	119
7.2	Evolución de los clústeres en ingeniería multimedia.....	119
7.2.1	Transiciones entre el primer y segundo parcial: Grafos y matrices de transición	120
7.2.2	Transiciones entre el segundo y tercer parcial: Grafos y matrices de transición.....	122
7.2.3	Cálculo de la Matriz Esperada ME a partir de la Matriz primero-segundo	124
7.2.4	Transiciones entre el primer y tercer parcial: Grafos y matrices de transición	125
7.2.5	Intervención del Centro: Planteamiento estadístico del problema	126
7.2.6	Resultados de la prueba de comparación	129
7.3	Evolución de los clústeres en ingenierías tecnológicas. Primer intento de ajuste..	131
7.3.1	Transiciones entre el primer y segundo parcial: Grafos y matrices de transición	131
7.3.2	Transiciones entre el segundo y tercer parcial: Grafos y matrices de transición.....	134
7.3.3	Cálculo de la Matriz esperada ME a partir de la Matriz primero-segundo.....	135
7.3.4	Transiciones entre el primer y tercer parcial: Grafos y matrices de transición	136
7.4	Desarrollo de la reducción a 2 clústeres en el tercer parcial para las ingenierías tecnológicas.....	138

7.4.1	Diagrama resultante de evolución de los clústeres.....	141
7.4.2	Comparación final entre el estado esperado y el pronosticado.....	143
7.5	Corrección a sistema biunívoco de 3 clústeres	144
7.5.1	Obtención de los clústeres para el segundo y tercer parcial, forzando a 3 grupos	144
7.5.2	Transiciones entre el primer y segundo parcial.....	145
7.5.3	Cálculo de la Matriz esperada ME a partir de la Matriz primero-segundo.....	146
7.6	Modelo explicativo de reducción a 4 clústeres	149
7.6.1	Construcción de los 4 clústeres de cada parcial	149
7.6.2	Transiciones entre el primer y segundo parcial: Grafos y matrices de transición	151
7.6.3	Prueba de significación Chi cuadrado para pronunciarse sobre la intervención académica del centro.....	153

Capítulo 8: Conclusiones..... 155

8.1	Recapitulación: esquema general de la tesis	157
8.2	Diferencias entre alumnos de Ingeniería multimedia y alumnos de Ingenierías tecnológicas.....	158
8.3	Semejanzas entre alumnos de Ingeniería multimedia y alumnos de Ingenierías tecnológicas.....	159
8.4	Consecuencias derivadas de nuestro análisis: Titulaciones adaptadas al nuevo EEES	162
8.5	Directrices para el futuro.....	163

Bibliografía..... 165

Capítulo 1

Ingenierías Tecnológicas versus Ingenierías Multimedia: Descripción del Panorama académico

En Ingeniería i Arquitectura La Salle de la Universitat Ramon Llull es posible cursar estudios enmarcados en diferentes Ingenierías. Las que enumeramos a continuación las agrupamos bajo el epígrafe *ingenierías tecnológicas* (en el sentido de *estrictamente tecnológicas*) y son las siguientes:

- a) Ingeniería Técnica de Telecomunicaciones en:
 - Sistemas Electrónicos
 - Imagen y Sonido
 - Sistemas de Telecomunicación
 - Telemática
- b) Ingeniería Técnica en Informática de:
 - Sistemas
 - Gestión
- c) Ingeniería Superior de Telecomunicaciones
- d) Ingeniería Superior en Electrónica
- e) Ingeniería Superior en Informática
- f) Máster universitario en Ingeniería electrónica y automática
- g) Máster universitario en Ingeniería de redes y telecomunicaciones
- h) Máster universitario en Informática y su gestión

Por otro lado, en dicho Centro también se ofrece otra Ingeniería, ya dentro del nuevo escenario del Espacio Europeo de Educación Superior (EEES según el tratado de Bolonia), verificada y aprobada por el Ministerio e inscrita en el Registro de Universidades, Centros y Títulos (ver BOE núm. 7 del jueves 8 de enero de 2009, resolución de 23 de diciembre de 2008), que obedece a una fuerte demanda laboral bajo el título de *Grado en Ingeniería Multimedia*. Con diferentes enfoques esta área de conocimiento también se imparte en otros Centros universitarios.

En el momento que se empezó a realizar la tesis se ofrecía dicha titulación como título propio de Universidad pendiente de homologación con el nombre de Graduado técnico en Tecnologías Multimedia y estaba en proceso de evaluación/acreditación por parte de la AQU (Agencia para a la Calidad del Sistema Universitario de Catalunya) ya que formaba parte de un Plan Piloto para la adaptación al nuevo escenario EEES (ver en http://www10.gencat.net/dursi/es/un/eees_pla_pilot.htm). A 10 de octubre de 2008 la AQU certificó su adaptación al EEES. Además como continuación también se ofrece el Máster oficial en Creación, Diseño e Ingeniería multimedia.

En el curso 1996/1997 nació en La Salle como una nueva iniciativa, única en España y pionera en Europa, con el nombre de “Ingeniería en Multimedia”. Esta carrera universitaria, que por tanto ya se viene impartiendo desde hace más de 10 años con resultados excelentes, está configurada en 2 ciclos: Ingeniero Técnico, de 3 años de duración, e Ingeniero Superior, de 2 años adicionales. Los estudios de Multimedia representan la síntesis de las tecnologías Electrónica, Informática y

Telecomunicaciones, con la incorporación, además de la Imagen y el Sonido. En los capítulos de la realidad virtual, la animación, teleservicios ergonómicos y teleformación, entre otros, desarrollan un servicio absolutamente fundamental para el nuevo tejido empresarial, moderno e innovador, que se instaura en estos comienzos del siglo XXI.

Como se puede ver el Graduado en Tecnologías Multimedia se caracteriza en La Salle por combinar una fuerte base de Ingeniería con contenidos procedentes del diseño interactivo y la narrativa audiovisual. Esto le confiere un carácter singular en la orientación, que queda reflejado en la siguiente tabla donde se puede comparar con los otros Centros que también la imparten:

Centro	Orientación de la asignatura
Universitat Politècnica de Catalunya	Apuesta clara por una orientación al Diseño, de manera que sus alumnos pueden incluso realizar pasarelas finales hacia esta disciplina. En este sentido se premian asignaturas como el Diseño Gráfico, la Usabilidad y el Diseño centrado en el Usuario o el Guión Multimedia. http://www.citm.upc.edu/multimedia.php?seccion=2&subseccion=3&idomaNou=esp
Universitat Oberta de Catalunya	Similar al graduado anterior poniendo más énfasis, si cabe, en la comunicación interactiva y en la producción de contenidos. Debe tenerse en cuenta que el modelo de enseñanza así como el perfil de entrada del alumnado es muy distinto puesto que se trata de un centro de educación a distancia. http://www.uoc.edu/web/esp/estudios/estudios_uoc/gmmd/gmmd_cuadro.htm
Universitat Autònoma de Barcelona – Tomàs Cerdà	Perfil tecnólogo-informático en el que prima la programación, los sistemas operativos, la ingeniería del software y los sistemas digitales, sin dejar de lado el resto de materiales referidos al audiovisual Multimedia. Se trata por tanto de un “hermanamiento” con la Informática y no tanto con el Diseño de interactivos. http://www.uab.es/servlet/Satellite/FUTUROS-ESTUDIANTES/Estudios-de-1--y-2--ciclo/Plan-de-estudios-1099409749337.html?param1=1089612449223
Universitat de Girona	Este es un graduado muy orientado a la comunicación audiovisual y multimedia, lejano al resto de programas, en el que incluso se tratan perfiles artísticos y humanísticos, distintos de los pertenecientes a ámbitos como el Diseño y la Informática. Importancia en lo que se refiere a la producción de contenido “clásico” audiovisual como la televisión, el video y el audio, por supuesto también en sus vertientes y soportes actuales. http://www.eram.cat/gram.php?id=4
La Salle (Universitat Ramon Llull)	Base ingenieril importante “heredada” de la orientación del centro para con todas sus especialidades. Un cierto “hermanamiento” con las Informáticas de Gestión y de Sistemas aunque debido esencialmente a contenidos de programación pura. Otros ámbitos de peso como el procesado: de señal, del habla, del audio y de la imagen, también desde un punto de vista teórico. Sin descuidarse, por supuesto, las disciplinas que integran el ciclo de vida de cualquier producto multimedia (Diseño, Usabilidad y DCU, Contenidos, Gestión, Tecnología y Guión).

Hay que considerar que el peso específico de La Salle en este estudio se derivó de ser el primer centro universitario de Europa que había puesto en funcionamiento los estudios de Ingeniería Multimedia (1996) tal como ya se ha comentado. Por otra parte, otros factores importantes a tener en cuenta, que dieron autoridad al centro, fueron: la adopción del modelo académico de integración en la empresa, la autonomía del Centro en temas curriculares y la adopción del modelo de seguimiento personalizado de los alumnos.

a) Adopción del modelo académico de integración en la empresa

Los estudios se concibieron sin olvidar que emanaban de una fuerte demanda en el mercado de las ingenierías. Por ello se facilitó a los alumnos la posibilidad de realizar no sólo prácticas en las empresas, sino también realizaciones de “estadías” en los centros de desarrollo de La Salle (Centro de Transferencia de Tecnología) y colaboraciones en los departamentos de la Escuela. Por otra parte se realizaron importantes convenios para la cesión de tecnología por parte de las empresas y, además, se fomentó la participación de miembros de las mismas en los currículums (ya fuera en cursos completos o seminarios).

Esta colaboración en proyectos reales es habitual en La Salle y se fomenta en todas las carreras universitarias. Uno de los objetivos fundamentales para la creación del Centro de Transferencia de Tecnología fue el poder ofrecer a los alumnos la participación en proyectos reales de empresa para su formación con el tutelaje de los propios profesores.

Todos aquellos alumnos que participan de esta oferta (colaboraciones en los departamentos, en el Centro de Transferencia de Tecnología o en empresas) acaban con una formación muy superior al resto de sus compañeros.

b) Autonomía del centro en temas curriculares

En el caso de multimedia por el hecho de ser un título propio de universidad tiene la desventaja inherente de que los alumnos egresados no pueden acceder a organismos públicos y no tienen título reconocido por el ministerio, pero en cambio ha permitido una evolución muy rápida e interesante. En estos 11 años se ha podido cambiar de manera muy ágil algunas asignaturas y sus contenidos para adaptarlos a los cambios que se han ido desarrollando en el área de la multimedia. Los más significativos se produjeron durante el curso 2003-2004 y fueron:

- Introducción a la electrónica se sustituyó por física con contenidos básicos de electrónica y de física mecánica.
- Procesado de señal se pasó a segundo curso porque a los alumnos les faltaban conocimientos de matemáticas.
- Medios estáticos y dinámicos se pasó de segundo a primer curso cambiando el nombre por Usabilidad y también sus contenidos.
- Algoritmos y manipulación de datos se sustituyó por Estructura de datos y Programación avanzada. Se aumentaron los conocimientos de programación.
- Se substituyó Microelectrónica por Guionaje.

- Procesado digital de la imagen y Procesado digital del habla se pasaron de segundo a tercer curso.
- Diseño VLSI para multimedia y Gestión de proyectos multimedia se substituyeron por Ingeniería del software en multimedia.

c) Adopción del modelo de Seguimiento personalizado de los alumnos

La Salle empezó a impartir estudios universitarios a principios del siglo XX en la modalidad de electricidad y mecánica. Al ser La Salle una institución en el que su principal objetivo es la educación de los jóvenes, el inicio universitario vino enmarcado por una gran preocupación no solamente para formar jóvenes en el conocimiento de las diferentes materias sino en su formación humana. Así es que desde siempre los profesores dedicaban una parte de su tiempo al seguimiento del alumnado.

Hasta el curso 1985-1986 los profesores disponían de una ficha personal de cada alumno de su asignatura y procedían a realizar su propio seguimiento.

Durante el curso 1986-1987 aparece la figura de tutor para los alumnos de primer curso debido al incremento importante del número de estudiantes y sobre todo para reducir el fracaso universitario que aparecía en el primer curso de carrera. A partir de este momento el doctorando empezó a responsabilizarse de dicha labor.

Hasta el día de hoy el doctorando se ha encargado de coordinar dicha tarea y se ha especializado en mejorar el seguimiento de los alumnos sobretodo para los alumnos de primer año. En estos momentos existe en el Centro un equipo de tutores que estudian, coordinan y mejoran el rendimiento de los alumnos con las siguientes funciones:

- Procurar la plena integración del alumno a la dinámica del Centro.
- Planificar un sistema de evaluación continuada que permita controlar el trabajo de los alumnos, conocer su rendimiento y mejorarlo.
- Entrevistar a los alumnos a su cargo par su seguimiento y ayudarle en el caso que sea necesario.
- Realizar encuestas
- Realizar estadísticas
- Proponer a la dirección mejoras

Para ello se ha facilitado al equipo de tutores una serie de herramientas, primero a través de fichas en papel, más adelante a través de una base de datos gestionada por el coordinador del equipo y a día de hoy con una aplicación informática integrada en la base de datos del Centro accesible para todos los profesores a través de la intranet.

Gracias a dicha aplicación todos los profesores pueden acceder a la ficha telemática del alumno para poner sus notas y sus comentarios. En todo momento su tutor tiene perfectamente conocimiento del rendimiento académico de su alumno, sus problemas y evolución.

Gestió de notes - Tutor
Notes Parcial > Veure Not

Notes Parcial - NOM DE L'ALUMNE

Notes Parcial en el curs 2008/2009				
(11021) Càlcul (I)				
Mates bàsiques				
1a Conv				
7				

Notes Parcial en el curs 2007/2008				
(11011) Àlgebra (1-B)				
1r trimestre				
Ac octubre	Test novembre	Ac 1r trimestre	Examen 1r trimestre	Final 1r trimestre
5	3.3	5	8	7.1
2n trimestre				
Ac 2n trimestre	Examen 2n trimestre	Final 2n trimestre		
8	8	8		
3r trimestre				
Ac 3r trimestre	Examen 3r trimestre	Final 3r trimestre		
5	5.95	5.95		
Juny				
1r trimestre	2n trimestre	3r trimestre	Final juny	Observacions
7.1	8	5.95	7	-
Setembre				
Examen 3r trimestre	Examen tot el curs	Final setembre		
-	-	-		

(11021) Càlcul (1-B)							
1r trimestre							
Convocatòria	Nota Mates Bàsiques	Ac octubre	Test novembre	Ac 1r trimestre	Examen 1r trimestre	Final 1r trimestre	
2	6	6	-	6	2.2	3.4	
2n trimestre							
Ac 2n trimestre	Examen 2n trimestre	Final 2n trimestre					
7	3	4.2					
3r trimestre							
Ac 3r trimestre	Examen 3r trimestre	Final 3r trimestre					
4	2.5	2.95					
Juny							
1r trimestre	2n trimestre	1r trimestre	2n trimestre	3r trimestre	JUNY ACTA	Observacions	
5.75	NP	5.75	4.2	2.95	4	Pots presentar-te només al 3r P	
Setembre							
3r trimestre	Tot el curs	Nota Final					
0.5	-	3					

(14021) Tècniques de representació (1-B)						
Pràctiques						
Pràctica	Pràctica 1	Pràctica 2	Pràctica 3	Pràctiques juny	Pràctiques setembre	
7	8	7	8.5	7.775	-	
1r trimestre						
Ac octubre	Test novembre	Ac 1r trimestre	Examen 1r trimestre	Final 1r trimestre		
8	2.5	7	3.95	4.9		
2n trimestre						
Ac Febrer	Ac 2n trimestre	Examen 2n trimestre	Final 2n trimestre			
6	6	0.5	2.15			
3r trimestre						
Ac 3r trimestre	Examen 3r trimestre	Final 3r trimestre				
4	0.01	1.21				
Juny						
1r trimestre	2n trimestre	3r trimestre	Teoria juny	Final juny	Nota modificada per junta	

(95011) Anglès I (1-B)					
1r trimestre					
Final avanKcat	Ac octubre	Test novembre	Ac 1r trimestre	Examen 1r trimestre	Final 1r trimestre
6	-	-	-	-	-
2n trimestre					
Ac febrer	Examen 2n	Final 2n			
-	-	-			
3r trimestre					
Ac maig	Ac 3r trimestre	Examen 3r trimestre			
-	-	-			
Juny					
3r trimestre	Final	Final curs			
-	6	6			

Comentaris

Data	Professor	Comentari
16/10/2008	Xavier Senmartí Senmartí	D'aquesta manera es visualitzen els comentaris que introdueix el professor-tutor. No són visibles per l'alumne

© 2005 Copyright de tots els continguts. Centre de Serveis Informàtics (CSI) - Sistema d'informació
 Enginyeria i Arquitectura La Salle - Universitat Ramon Llull

Ingenierías Tecnológicas versus Ingenierías Multimedia: estudio de perfiles aplicando la metodología de los sistemas consistentes de clústeres

7

1.1 Interés de la multimedia en el mundo empresarial

Como ya hemos comentado desde 1996 se está ofreciendo una titulación de Ingeniería Multimedia pero es en estos momentos cuando se plasma mejor el impacto que tiene en el mercado laboral.

Los nuevos tiempos exigen de la aparición de titulaciones con carga interdisciplinaria para alimentar al mercado laboral. Este punto de manera indiscutible ha sido objeto de exposición reiterada en documentos de la Comunidad Europea desde inicios de la década de los noventa.

La titulación recoge una idea apuntada en varias ocasiones: la de que dentro del contexto de las Humanidades actuales se requiere de un refuerzo por parte de las Ciencias y de las Tecnologías. Es el caso de los filósofos que dialogan sobre Bioética (reclamando la implicación de biólogos genetistas) o de los sociólogos en su investigación de los grandes grupos sociales (reclamando modelos estadísticos más potentes) o de los periodistas que, ante la complejidad de la actualidad, reclaman una especialización por áreas (economista-periodista, psicólogo-periodista, etc.).

De hecho se trata de recoger una aspiración de las ciencias humanas que, de alguna manera, nos recuerda a los inicios del pensamiento científico tal y como lo entendemos hoy en día, es decir, en la concepción de las ciencias por parte de la antigua Grecia, en la que hay una fuerte relación entre filosofía y física, por ejemplo. Hace falta construir a humanistas que dispongan de una base sólida en un cierto campo científico o tecnológico. La historia está plagada de precedentes de este tipo: Piaget, sin ir más lejos, es el padre de la pedagogía evolutiva actual aunque era médico; Arthur Clarke que relee y divulga a Nietzsche para el público del siglo XX, es físico; Einstein renueva todo el pensamiento filosófico futuro desde su posición de eminente físico; Marvin Minsky es el referente de la inteligencia artificial en muchas disciplinas, como en el caso de la psicología cognitiva, pero es matemático; Lee Iacocca, un alto ejecutivo de Ford y padre del popular modelo Mustang, estudió ingeniería industrial... y un informático como Bill Gates es prácticamente el hombre más rico del mundo porque fue una de las primeras personas en entender la nueva economía derivada de la revolución de las TIC.

Por otro lado, hace falta recordar que esta idea que se va exponiendo se recogió en el diseño de los planes de estudios universitarios mediante la introducción de los llamados créditos de libre elección. Es decir que estos se pensaron para que el titulado o titulada en una disciplina científica o técnica, pudiera acercarse a las Humanidades, completando su currículum. Podría ser que en la práctica, no todos los estudiantes hayan entendido esta filosofía. El Ingeniero en Multimedia de La Salle da un paso firme en este sentido ya que se define como “tecnología completada con contenido de carácter humanístico”.

No se puede olvidar que la rápida evolución de la nueva economía está comportando la aparición de nuevos modelos de negocio que a menudo ponen en peligro la existencia de los modelos tradicionales. Con el objetivo de dotar a la empresa de la capacidad de anticipación necesaria para convertir estas amenazas en oportunidades, el ingeniero multimedia se forma en la exploración de nuevas vías de negocio mediante el uso de los nuevos canales de comunicación "pull" (Internet, WAP/GPRS/UMTS, TV interactiva, etc.).

A grandes rasgos, podríamos caracterizar a la Multimedia por:

- a) Ser una disciplina que utiliza, y aglutina, herramientas de trabajo y modelos propios que la diferencian del resto de titulaciones de Ingeniería ya existentes. De hecho los Ingenieros multimedia entienden de productos distintos a los de las titulaciones afines, y diferenciados de lo clásico, por tanto innovadores, como por ejemplo:
- La integración de conceptos del Guionaje, Gráficos por Ordenador, Animación y Usabilidad para conseguir una producción cultural consistente en una visita guiada por un edificio histórico, ya inexistente y por lo tanto inexplorable. Sería un caso real de recuperación de memoria histórica, válido tanto a nivel cultural como docente.
 - Un simulador quirúrgico basado en técnicas de telemedicina, que a partir de imágenes médicas reconstruye el órgano del paciente y permite que futuros y futuras especialistas interactúen con él en tiempo real, utilizando periféricos específicos a tal efecto, emulando técnicas laparoscópicas.
 - Un sistema de audioguías en formato mp3 descargables a cualquier periférico de bolsillo (móviles, asistentes personales digitales o pda's) gracias a puntos Bluetooth situados estratégicamente en edificios y construcciones notables de una ciudad, realizado para el departamento de turismo del ayuntamiento de turno.
 - Un viaje de Realidad Virtual (RV) dedicado a la promoción de un determinado artículo de consumo de una multinacional, basado en periféricos específicos (guantes de RV o Datagloves, cascos de RV o Head Mounted Displays, pantallas inmersivas de tipo Workbench, etc.) que garantizan un altísimo grado de inmersión y de entendimiento, por parte del usuario/a final, de aquello que se pretende vender,
 - Un locutor virtual fotorealista basado en móvil que es capaz de leer textos así como de interpretar comandos orales del usuario/a. El locutor es personalizable en cuanto a apariencia y tono de voz y está basado en tecnología de procesado de señal (imagen y habla).
 - Un gestor de contenidos que permite que personas sin ningún tipo de experiencia en las TIC manejen sin problemas, y revisen y actualicen, un complejo portal Web de cualquier compañía o servicio público, por ejemplo. Además es altamente usable (ergonómico y centrado en el usuario/a final) y accesible (para todos/as).
 - Mecanismos de e-Business y Marketing Interactivo que fidelizan al cliente tomando nota de sus preferencias cuando navega por la red y adaptando el mensaje a enviar a éstas, concretándolo y haciendo que sea de verdad operativo y útil. Tecnologías que implementan pequeños “robots software” capaces de adaptarse, gracias a su inteligencia artificial, a estos/as.
 - Mecanismos de interacción avanzada y natural, huyendo de esquemas que sí, son los clásicos como en el caso del teclado, pero nada ergonómicos ni adaptados a los discapacitados. Se trata de interactuar en tiempo real con las aplicaciones gracias a la sombra, al seguimiento de la retina o del movimiento de la cabeza o extremidades, mediante lecturas por parte de cámaras de video o de sensores específicos (acelerómetros por ejemplo).
 - Y otros muchos productos claramente innovadores...

- b) Ser complementaria con otras disciplinas dentro del campo de las TIC, como las Telecomunicaciones (Hardware) o la Informática (Software), o al margen de estas, como las Bellas Artes o las Humanidades. Entiéndase que comparte pero se diferencia a su vez, de las anteriores, puesto que emplea sus propios modelos y procesos de trabajo.
- c) Formar parte de cualquier problema que requiera de un análisis e integración de disciplinas bien diferentes. Sin ir más lejos, en aspectos relativos a la producción de contenidos audiovisuales, es necesario integrar materiales de todo tipo como guiones, vídeo, audio, fotografía, imagen sintética y texto, y transmitirlos de forma eficiente por canales diversos como pueden ser la red o los dispositivos móviles (teléfonos, asistentes digitales o pda's, etc.), por ejemplo.
- d) Estar reconocida por prestigiosos organismos.
- e) Existir grupos de investigación oficiales que basan sus líneas de investigación en esta disciplina. Por mencionar unos cuantos, el Grupo de Investigación en Tecnologías Audiovisuales y Multimedia (TAM) de Ingeniería La Salle (URL); los Grupos de Investigación en Comunicación Audiovisual y Cultura Digital: Creación, Participación, Industria y Usos Sociales (COMCAD), en Comunidades Virtuales (GIRCOM), en Personalización en Entornos de Aprendizaje Virtual (PERSONAL) o en Trabajo y Aprendizaje Cooperativo en Entornos Virtuales (TACEV) de la UOC; el Grupo de Investigación en Didáctica y Multimedia (DIM) de la UAB; el Grupo de Investigación en Enseñanzas y Aprendizaje Virtual (GREAV) de la UB; el Grupo de Investigación en Interacción Persona-Ordenador y Bases de Datos (IPO) de la Universidad de Lleida; el Grupo de Investigación en Aplicaciones Multimedia Distribuidas (DMAG) de la UPF; el Grupo de Procesado Multimedia de la Universidad Carlos III de Madrid; el Grupo de Sistemas en Internet, Multimedia y Bases de Datos (SIMBAD) de la Universidad de Valencia; el Virtual Laboratories in Science and Technology Group de la Universidad de Córdoba; el Grupo de Herramientas Interactivas Avanzadas de la Universidad Autónoma de Madrid; el Grupo de Comunicación Audiovisual, Nuevas Tecnologías e Industrias Culturales (INTERMEDIATIC) de la UPV o el Grupo de Tratamiento de Imágenes de la Universidad Autónoma de Madrid, entre otros.
- f) Estar dentro de los objetivos que dictamina el 7º Programa Marco Europeo para la investigación y el desarrollo a nivel europeo.
- g) Ser valorada por la empresa debido a su perfil multidisciplinar. Claro está que actualmente cualquier proyecto de envergadura incorpora conocimiento de áreas bien diversas. Es en estos casos cuando es fundamental una figura que conozca las tecnologías y las canalice por un hilo conductor común que las integre de manera coherente. Hace falta una "vista de pájaro" de todo el sistema para llegar a un buen diseño y funcionamiento final, eficiente por encima de todo.
- h) El carácter generalista de la titulación. El Ingeniero multimedia puede introducirse en diferentes ámbitos profesionales ya que las afinidades con otras disciplinas son evidentes. Las Telecomunicaciones, la Informática, la Imagen y

el Sonido, el Diseño Gráfico y la Comunicación son otras áreas íntimamente ligadas con la Multimedia que, en cierta manera, las aglutina y gestiona eficientemente y de forma conjunta.

- i) Ofrecer estudios de segundo ciclo en el área desde hace más de 10 años. Aunque no se ha gozado de reconocimiento oficial, el tejido empresarial acoge a los titulados/as a sus plantillas. El grado de inserción laboral es del 100%. De hecho la integración laboral es tan alta que a veces retrasa la finalización de los estudios dada la tendencia de nuestros alumnos a recibir ofertas laborales desde etapas muy tempranas de los estudios.
- j) Acoger a muchos alumnos que se encuentran trabajando en el sector Multimedia para completar su formación, en este caso por necesidad.

El interés empresarial por la multimedia también queda reflejado por la oferta académica de La Salle en ciclos formativos de grado superior y en másteres profesionales como: Management con MCDM – Master en Creación y Diseño Multimedia y MEB – Master en e-Business o guión con MCPAD - Master en Comunicación y Producción Audiovisual Digital y tecnología con MGA3D - Master en Guión y Animación 3D entre otros; y cursos específicos en materia de Animación, Videojuegos, Programación Gráfica, Herramientas de Autor y Lenguaje/Narrativa Audiovisuales u otros, gestionados desde nuestro centro de formación especializada.

La multidisciplinariedad de la multimedia la podemos apreciar en una parte de un estudio de investigación de perfiles realizada entre Agosto del 2007 y Febrero del 2008 por parte de Interactiva del Instituto de Recursos e Investigación para la Formación IRIF, SL., con la colaboración del Consortium for Multimedia Studies (CM), formado por los siguientes centros: Ingeniería i Arquitectura La Salle, Universitat de Girona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Oberta de Catalunya y la Asociación empresarial AENTEG. Dicho estudio dio los resultados que a continuación mostramos.

En el ámbito docente, como se puede apreciar en la figura 1.1, podemos observar una mayor concentración de las titulaciones en los ámbitos de Informática y Bellas Artes de los docentes que imparten las titulaciones de multimedia. Sin embargo también podemos observar la gran cantidad de titulaciones que pueden tener los docentes que imparten multimedia.

En el ámbito profesional, en la figura 1.2, podemos apreciar los sectores donde se desarrollan los profesionales multimedia. Es importante significar que en un porcentaje elevado los profesionales multimedia desarrollan su trabajo en el ámbito de las TIC como era de esperar, no obstante refleja la cantidad de sectores en las que participa.

En cuanto al perfil y el rol que desempeñan, como se puede ver en la figura 1.3, hay que destacar que la mayor concentración la tenemos en Informática y Web, pero es muy importante la cantidad de roles diferentes que pueden desempeñar.

En la figura 1.4 podemos observar que la mayoría de los profesionales de multimedia son propiamente graduados multimedia. De aquí vemos la gran demanda del mercado.

Finalmente en la figura 1.5 observamos a partir de que muestra de titulados se han obtenido todos estos resultados y a la vez cuan joven es esta titulación, ya que el gran numero de titulados se concentra en los últimos años.

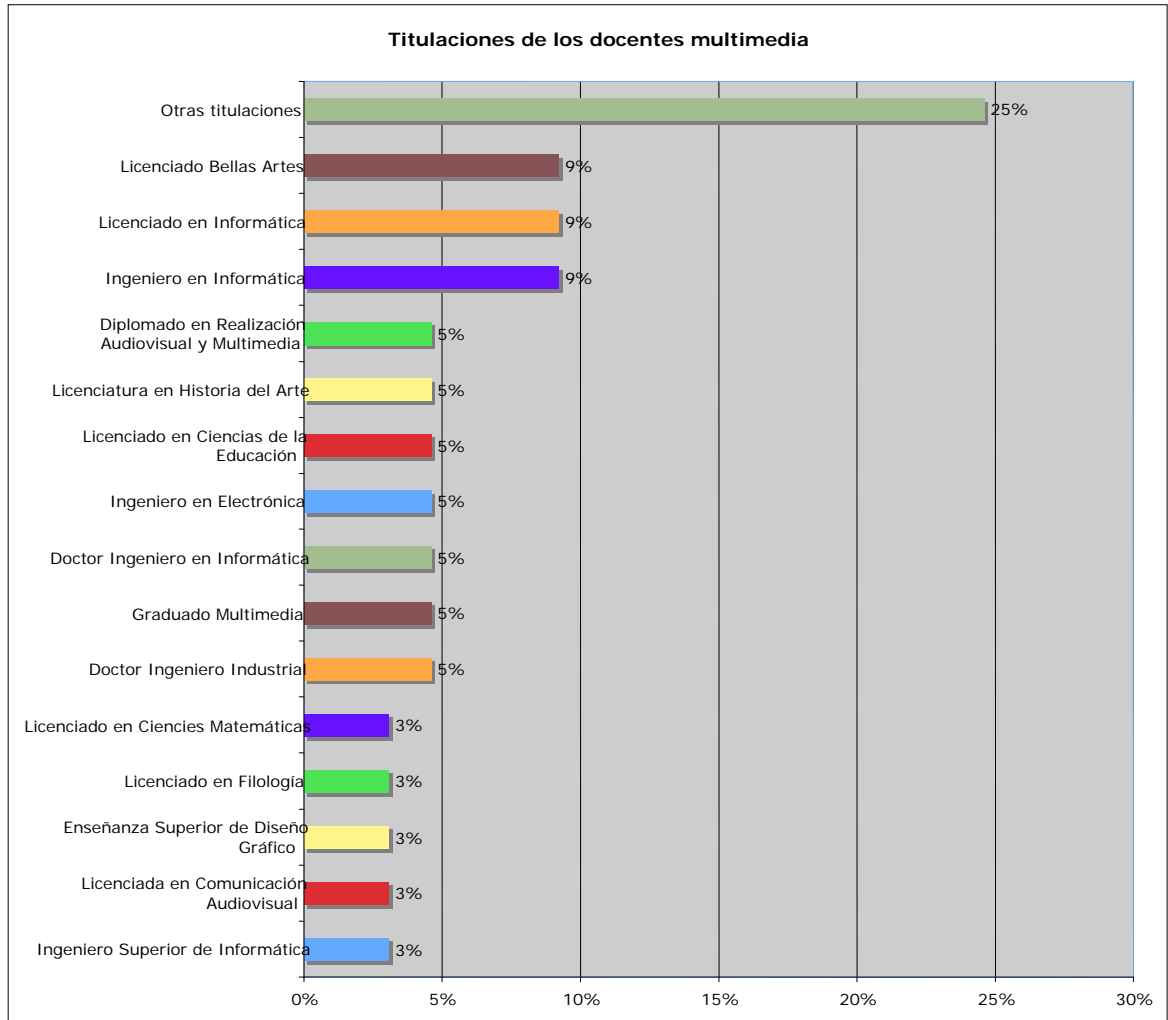


Figura 1.1 Titulaciones que ostentan los docentes multimedia.

	FRECUENCIA	PORCENTAJE
Consultoría TIC (usabilidad, programación y desarrollo, 3D)	33	17%
Docencia / universidad / investigación	24	12%
Aplicación Web	18	9%
Empresas basadas en Internet (punto com)	16	8%
Tecnologías de la Información y la Comunicación (TIC)	15	8%
Proyectos audiovisuales	10	5%
Aplicaciones Multimedia	9	5%
Diseño y Publicidad	9	5%
Industria informática	7	4%

1. Ingenierías Tecnológicas versus Ingenierías Multimedia: Descripción del Panorama académico

Diseño gráfico	7	4%
Informática	7	4%
Marketing / Publicidad	6	3%
Televisión	5	3%
Automoción	4	2%
Telecomunicaciones	4	2%
Sector de banca	4	2%
Entidades y asociaciones públicas	3	2%
Biomecánica	2	1%
Arquitectura	2	1%
Industria del hogar	1	1%
Eventos artísticos y culturales	1	1%
Artes Gráficas	1	1%
Instalaciones eléctricas	1	1%
Prensa	1	1%
Ferias, congresos, eventos	1	1%
Ingeniería informática	1	1%
Alquiler de equipos de luz y sonido	1	1%
Sector Construcción	1	1%
Telefonía móvil	1	1%
Servicios	1	1%
Salud / Enfermería	1	1%
Empresas químicas multinacionales	1	1%
Radio	1	1%
Sector Editorial	1	1%
TOTAL	200	100%

Figura 1.2 Sectores donde se desarrollan los profesionales multimedia.

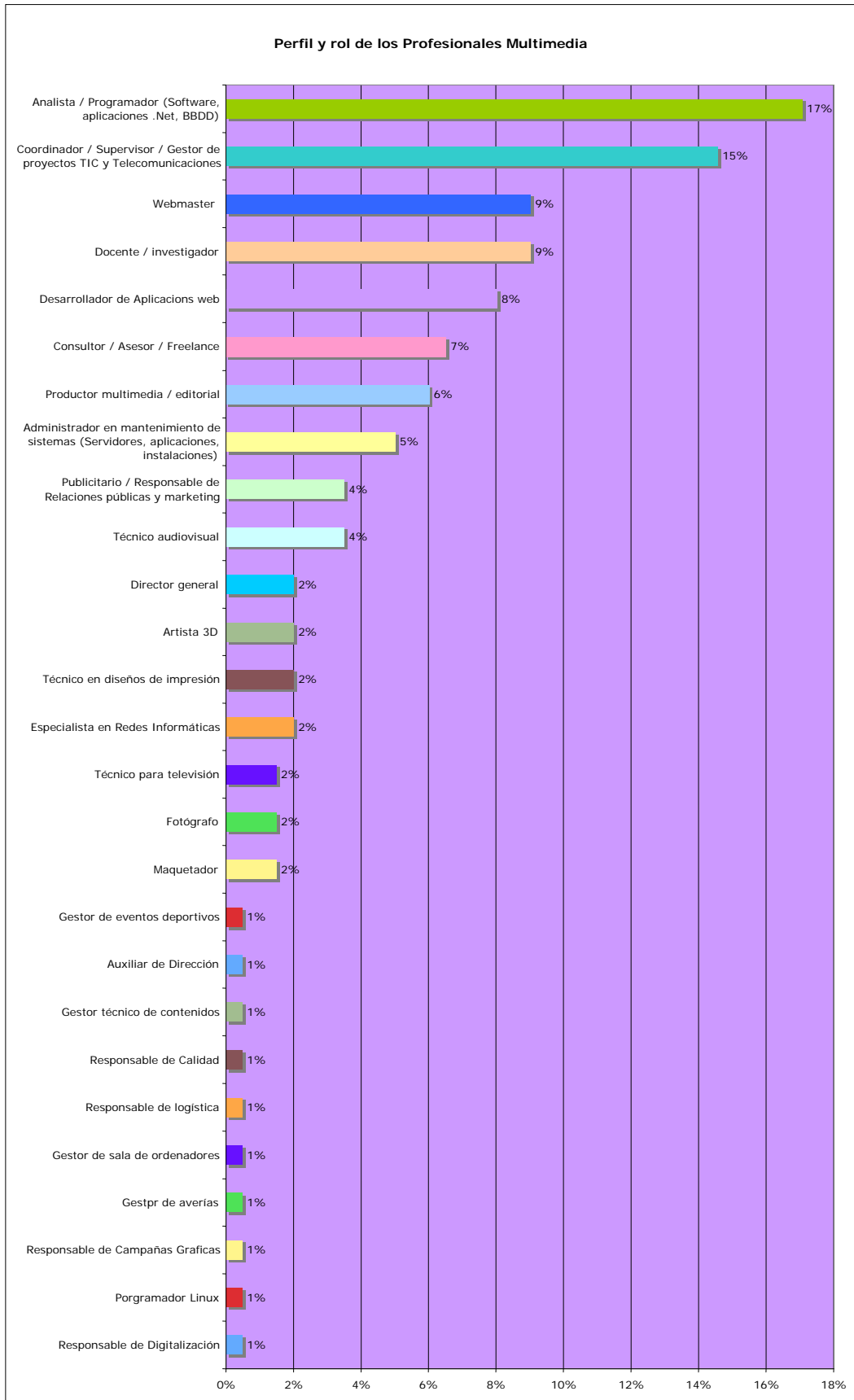


Figura 1.3 Perfil y rol de los profesionales multimedia.

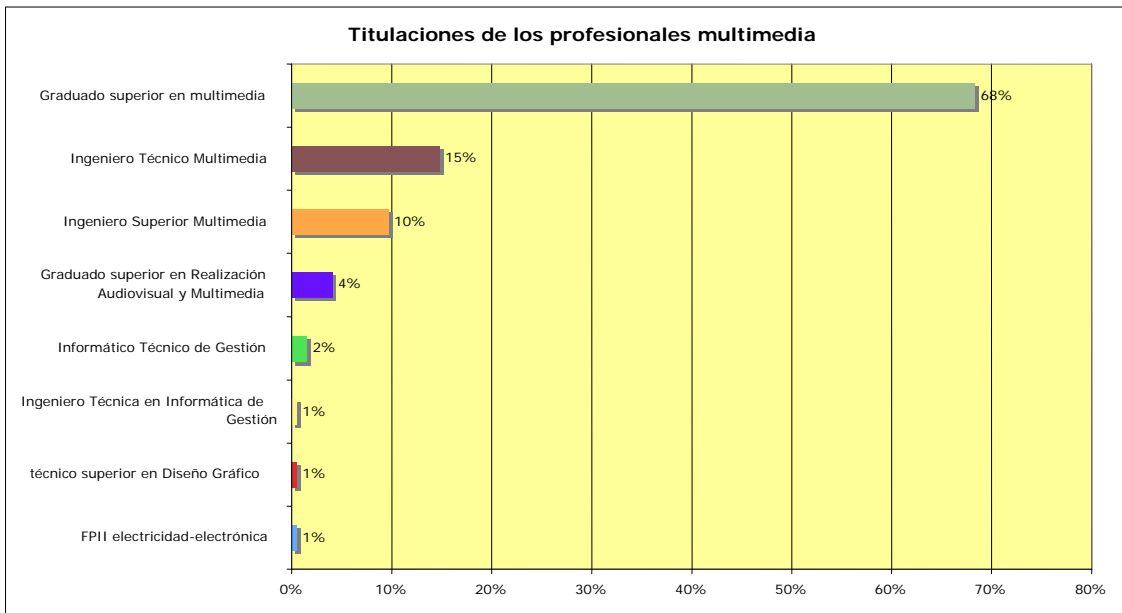


Figura 1.4 Titulación del profesional multimedia.

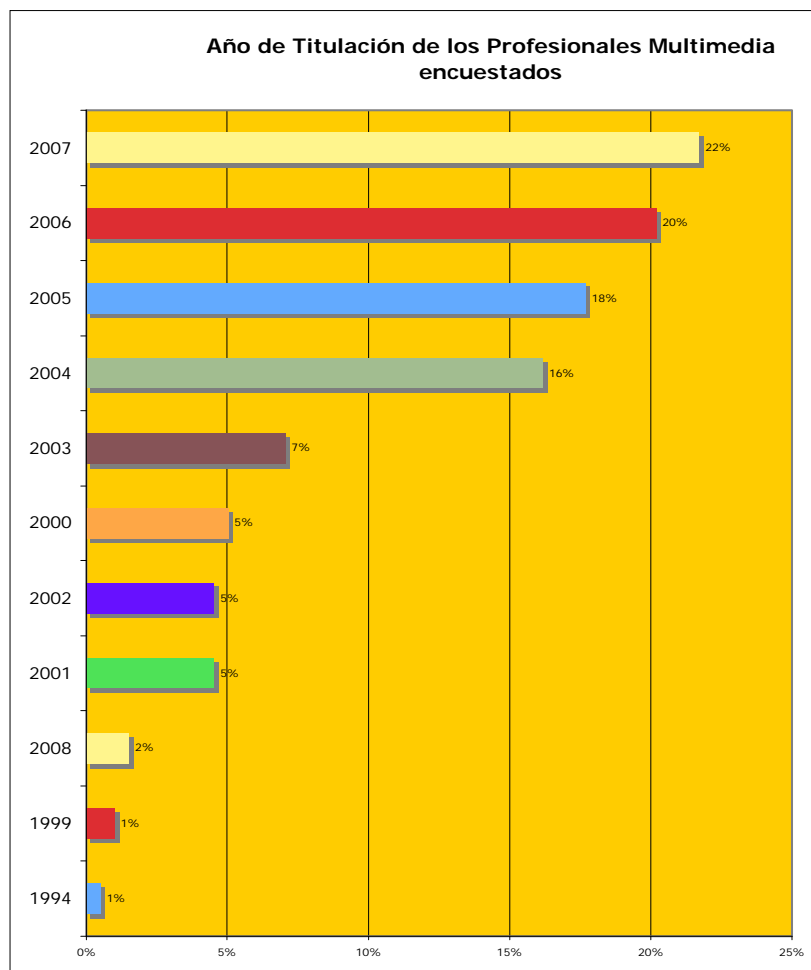


Figura 1.5 Año de titulación de la muestra.

1.2 Descripción de la situación problemática

En esta investigación, nos acogemos a la acepción de McGuigan de problema de investigación como situación emanada de una inquietud. Además se ha hecho un análisis que según Ary y otros (1987) Bartolomé (1982) Van Dalen y Meyer (1983) se debe hacer junto con el planteo del problema. Nos hemos de asegurar que el problema sea:

- a) Real: Debe partirse de un problema percibido o sentido. En nuestro caso claramente lo es. Tratamos sobre hechos reales de un grupo de estudiantes.
- b) Factible: Debe de reunir las condiciones suficientes para ser estudiado y que esté al alcance del investigador. En nuestro caso, es un problema al que queremos dar solución por la proximidad a la realidad vivida día a día en el centro.
- c) Relevante: El investigador debe exigirse rigor a la hora de elegir el problema y sobretodo verificar que es un problema que interesa, que es actual y que tiene una relevancia práctica tangible. Nuestro problema es actual y de gran relevancia ya que de ello depende la calidad de la enseñanza en el centro.
- d) Resoluble: Debe de ser susceptible a verificación empírica (kerlinger, 1985). Respecto a nuestro problema planteado es resoluble ya que es el objetivo de dicho trabajo comprobar la resolubilidad del mismo.
- e) Generador de conocimiento: Se debe reflexionar de si la solución del problema contribuirá a crear conocimiento pedagógico o cubrirá alguna laguna que exista en el conocimiento del momento. Gracias a este estudio, se le dará una nueva visión a problemas que se plantean desde hace años respecto a la enseñanza.
- f) Generador de nuevos problemas: La solución a dicho problema debe conllevar a nuevos horizontes donde se planteen problemas de nueva índole. La nueva índole a la que nos referimos, en nuestro caso, sería la creación de nuevos planes de estudio más acordes con la realidad.

Podemos decir que nuestro problema de investigación emana de los siguientes hechos. Por una parte:

- a) Se ha realizado un esfuerzo de diseño curricular para la creación de la ingeniería multimedia.
- b) Se han puesto recursos de tutoría al servicio de los alumnos, de igual modo que existen en las otras ingenierías.
- c) Se han supuesto unas características vocacionales e intelectuales de los alumnos a partir de sondeos y pequeños estudios estadísticos.

Pero, por otra:

- a) No se ha realizado un procesamiento de la información relativa a los rendimientos académicos para constatar los resultados de este diseño curricular.
- b) No se puede afirmar que el tipo de tutoría para unos ingenieros y para otros (tecnológicos y multimedia) deba ser el mismo, considerar los mismos aspectos o incidir en las mismas orientaciones, salvo por lo que parece emanar de la experiencia.
- c) No se han realizado pruebas estadísticas de contrastación de las supuestas percepciones vocacionales e intelectuales

Así pues, a la pregunta “¿Quiénes son nuestros alumnos?” o “¿Cómo son nuestros alumnos?” el centro puede responder, con cierta autoridad, en los casos de las ingenierías tecnológicas. Escribimos “cierta autoridad” porque no se han realizado tratamientos estadísticos en el marco de investigaciones como la presente pero sí, en cambio, tratamientos de aspectos parciales. Además, existen datos de seguimiento e informes de conclusiones (elaborados frecuentemente en el marco de la orientación para la matrícula y creados a partir de estadísticas descriptivas de rendimiento).

En cambio, si esta misma pregunta se formula en relación a los alumnos de Ingeniería Multimedia, obviamente la autoridad de la respuesta del centro será mucho menor, y lo será por dos factores:

- a) Por una menor edad académica de la titulación, lo cual es un motivo evidente, pero se está haciendo un gran esfuerzo por parte de todo los profesionales.
- b) También por un mayor desconocimiento por parte del profesorado de las materias, ya que se incluyen contenidos de diseño interactivo o audiovisuales.

Diversos estudios (como el ya clásico de Gardner de 1983) indican que el profesorado tiende a repetir la manera de enseñar con la que entendió las materias cuando era estudiante. En nuestro caso, por lo que se refiere a las ingenierías clásicas, y debido a su mayor homogeneidad de contenidos, los profesores del centro tienen una idea precisa del esfuerzo que se exige al estudiante. En el caso de la ingeniería multimedia, en cambio, pueden medir este esfuerzo por indicadores finales (como dedicación y rendimiento) pero no por un conocimiento directo de las materias o las metodologías de enseñanza.

En la figura 1.6 podemos apreciar el plan de estudios de la titulación de Ingeniería Técnica Multimedia que han seguido los alumnos a los que el doctorando ha realizado el seguimiento personalizado a través de los tutores correspondientes.

PLA D'ESTUDIS D'ENG. TÈCNICA EN MULTIMÈDIA - PLA NOU				
codi	nom oficial	nom comú	crèdits PLA	crèdits matric
PRIMER CURS				
11011	Algebra	Algebra	9	9
95011	Anglès I	Anglès I	9	9
14021	Circuits digitals	TR-IO	9	9
11021	Càlcul	Càlcul	9	9
14041	Disseny d'interfícies multimèdia	Disseny artístic	9	9
TM019	Física	Física	12	12
50011	Programació d'Ordinadors	Programació I	9	9
TM013	Usabilitat en sistemes interactius	Usabilitat	6	6
SEGON CURS				
95022	Anglès II	Anglès II	9	9
50072	Estructures de dades	Estructura de dades	9	9
TM005	Gràfics per ordinador	Introducció als gràfics per ordinador	9	9
TM018	Guionatge (llenguatge)	Guionatge I	6	6
11032	Matemàtiques i càlcul numèric	Matemàtiques i Càlcul Numèric	9	9
50022	Programació II	Programació II	9	9
49063	Telemàtica	Telemàtica I	9	9
48171	Tractament del senyal	Tractament del senyal	9	9
TERCER CURS				
TM008	Animació per ordinador	Animació per ordinador	9	9
50063	Bases de dades	Arxius i bases de dades	6	6
TM009	Compressió de dades	Compressió de dades	6	6
50043	Enginyeria del software en multimèdia	Enginyeria del software	9	9
TM010	Producció multimèdia	Producció multimèdia	6	6
TM015	Programació hipermèdia	Programació hipermèdia	6	6
TM007	Realitat virtual	Realitat virtual	6	6
TM014	Sistemes i entorns distribuïts	Sistemes i entorns distribuïts	9	9
TM002	Tractament Digital de la Imatge	Tractament digital de la imatge	9	9
TM003	Tractament Digital de la Parla	Tractament digital de la parla	9	9
TM020	Treball Final de Carrera	TFC	9	10

Figura 1.6 Plan de estudios de Ingeniería técnica multimedia

Si se acepta una simplificación de los términos, puede decirse que la situación problemática que da pie al problema de investigación gira en torno a una cuestión simple: ¿Son diferentes los alumnos de ingeniería multimedia del resto de estudiantes de las diferentes especialidades de ingeniería? Formulada esta pregunta, cabe preguntarse en qué aspectos son diferentes y cómo va a traducirse en la intervención educativa del centro. Puede decirse, pues, que se ha enfocado el problema de investigación tal como se expone en el siguiente apartado.

1.3 Aproximación al problema de investigación

1.3.1 Precedentes

La Escola Universitària d'Enginyeria Tècnica de Telecomunicació La Salle ha realizado diversas sondeos sobre los alumnos de los estudios de Ingeniería de Telecomunicaciones, Informática y Multimedia.

También se han realizado tesis doctorales sobre rendimiento académico en asignaturas y campus virtuales como *La docència d'enginyeria electrònica: directrius per al disseny de la instrucció presencial y a distància* de Jordi Margalef i Marrugat y *Propuesta*

e-learning para titulaciones de ingeniería en el espacio europeo de educación superior. El campus virtual mínimo de Lluís Vicent Safont.

De la gestión de matrícula se dispone de grandes cantidades de datos sobre rendimiento, por lo que se pueden realizar tratamientos estadísticos más potentes. El doctorando en todos sus años de profesional como educador en la universidad ha liderado la acción tutorial a los alumnos de primer curso dirigiendo un equipo de profesores que han desarrollado grandes bases de datos sobre los resultados y evolución del alumnado. Gracias a ello se ha ido mejorando el seguimiento del alumnado y los sistemas de evaluación continua. Como ya se ha comentado el doctorando ha desarrollado una serie de paquetes informáticos que permiten un mejor tratamiento de la información.

1.3.2 Enunciado informal del problema

En el capítulo 3 de la presente tesis se realizará el enunciado del problema de investigación según los cánones que establece la metodología. Sin embargo, con el escenario que se ha descrito en los apartados anteriores, ya podemos definir de manera informal nuestro problema de investigación del siguiente modo:

¿En qué difieren, por lo que se refieren a rendimiento conducta académico (entendido de manera amplia) los estudiantes de ingeniería multimedia del resto de estudiantes de las otras especialidades de ingeniería?

Decimos “entendido de manera amplia” porque no nos centraremos sólo en unas calificaciones finales, sino en todo el proceso que nos ha llevado a ellas, y teniendo en cuenta las variables predictoras y las de seguimiento, ya que son medidas que se han tomado con mucha frecuencia sobre cada individuo y retratan muy bien su evolución.

En el procedimiento de investigación se han tenido en cuenta unas variables predictoras que consisten en datos sobre conocimientos académicos que poseían los alumnos antes de entrar en la universidad extraídos de la nota de las pruebas de selectividad, de las notes de bachillerato (sobre todo las de ciencias) y de la prueba de acceso que realiza La Salle (matemáticas y física). En cuanto a las variables de seguimiento consisten en datos sobre los conocimientos que iban adquiriendo los alumnos a medida que avanzaba el curso incluyendo los resultados finales.

1.3.3 Importancia de la investigación

En el ámbito institucional el interés queda acentuado debido al proceso de Bolonia en el cual están inmersos todos los Centros Universitarios Europeos. Uno de objetivos fundamentales que marca dicho proceso es el seguimiento que conviene hacer a los estudiantes durante todo el proceso de su aprendizaje.

Todo el desarrollo tutorial y sus mejoras han sido conseguidas a través de bibliografía y de la propia experiencia personal y del equipo de colaboradores. Evidentemente se han tenido en cuenta estadísticas de los resultados obtenidos por los estudiantes pero se ha echado siempre en falta un trabajo de investigación global, ajustado a un modelo estadístico, madurado y que ofreciera un esquema explicativo (no sólo descriptivo) de la vida académica de nuestro centro. Este hueco es el que viene a cubrir la presente tesis doctoral y, para ello, se fundamenta un doble trabajo:

- a) Elaboración de una nueva metodología de tratamiento de los estudios basados en clústeres, que hemos denominado *Sistemas consistentes de clústeres*.
- b) Unión a la metodología anterior de todo el saber del doctorando derivado de su experiencia académica, en el conocimiento de las tutorías y de los estudiantes de las ingenierías tecnológicas y multimedia desde hace más de una década.

Como valor añadido de dicha investigación hay que señalar que nos va a permitir no solamente mejorar el proceso de seguimiento y ayuda a los alumnos, sobre todo en los primeros cursos universitarios, sino también en el proceso de selección. Sabíamos que era importante la base matemática y física de los candidatos a estudiar en la universidad pero por primera vez gracias a este trabajo podremos valorar la validez de la pruebas de entrada y la selectividad. De esta forma podremos evaluar y mejorar las directrices y criterios que se aplican en la entrada de nuevos alumnos al centro.

Finalmente, cabe realizar un apunte personal sobre la presente investigación. He sido responsable del sistema tutorial universitario de La Salle durante veinte años, tengo pleno conocimiento de las variaciones en el mismo y de las causas de tales variaciones, he visto colaborar en él a diferentes profesores que defendían diferentes puntos de vista y, quizás lo más importante, sé que se ha dado marcha atrás cuando una medida no ha funcionado. Para las personas que han trabajado durante estos años y para mí en particular, como responsable, se da cohesión y solidez a la labor desarrollada con la elaboración de esta investigación.

Capítulo 2

Bases metodológicas: Análisis del rendimiento en Ingeniería con la metodología orientada al clúster

2.1 Precedentes metodológicos: Cluster Dynamics y Cadenas de Markov

2.1.1 La idea de evolución en las investigaciones basadas en agrupaciones

Los estudios basados en clústeres nos facilitan una aproximación por grupos para entender un determinado fenómeno en un momento dado. Estos algoritmos de clasificación suelen acompañarse de otras pruebas estadísticas o, alternativamente, suelen desarrollarse en profundidad (justificación de las distancias puestas en juego, justificación de los métodos de agrupación, interpretaciones después de realizar iteraciones, etc.). Sin embargo, mientras no se introduzca la idea de *evolución* de los grupos, toda la información que se obtiene se entiende referida a un instante determinado, esto es, a una foto estática del fenómeno.

Conscientes de lo señalado en el párrafo anterior, se han desarrollado herramientas estadísticas para introducir esta idea de “evolución de los grupos”. Una de las más conocidas es la que agrupa las investigaciones bajo el epígrafe de *Cluster Dynamics*. Con esta denominación, se designan unas determinadas pruebas que estudian series temporales y, una vez determinadas éstas, intentan clasificarlas.

Una definición clara de lo que son las técnicas de *Cluster Dynamics* y otras de tratamiento avanzado de datos la encontramos en *Innovation, learning and cluster dynamics* de Bart Nooteboom:

“cluster dynamics is the development of clusters in time. There are three reasons for this. First, dynamics stands high on the scientific research agenda. Often, studies of business systems, networks, inter-firm relations and organizations profess the importance of dynamics but remain stuck in, at best, comparative statics.”

Por otra parte, es una técnica muy conocida en estadística la de las cadenas de Markov para el estudio, en general, de un sistema de estados. Una cadena de Markov, según James R. Norris es:

*“Si tenemos un conjunto de estados $S=\{s_1, s_2, \dots, s_r\}$. El proceso empieza cuando uno de estos estados se mueve sucesivamente de un estado a otro. Cada movimiento lo llamamos *step*. Si la cadena está actualmente en el estado s_i , entonces este se mueve a s_j en el siguiente *step* con una probabilidad descrita por p_{ij} , y su probabilidad no depende de que estado de la cadena estuviera antes del estado actual. Las probabilidades de p_{ij} se llaman probabilidades de transición. El proceso puede permanecer en el estado actual y esto ocurre con la probabilidad de p_{ii} . Una primera distribución de probabilidad, definida sobre S, especifica el estado de inicio. Normalmente esto se hace especificando un estado particular como el Estado inicial.”*

Ahora bien, si miramos en conjunto el enfoque tipo *Cluster Dynamics* y el de las Cadenas de Markov, podemos pensarlos como dos formas matemáticamente “perpendiculares” de acercarse a un fenómeno. Ello se plasma gráficamente en la figura 2.1.

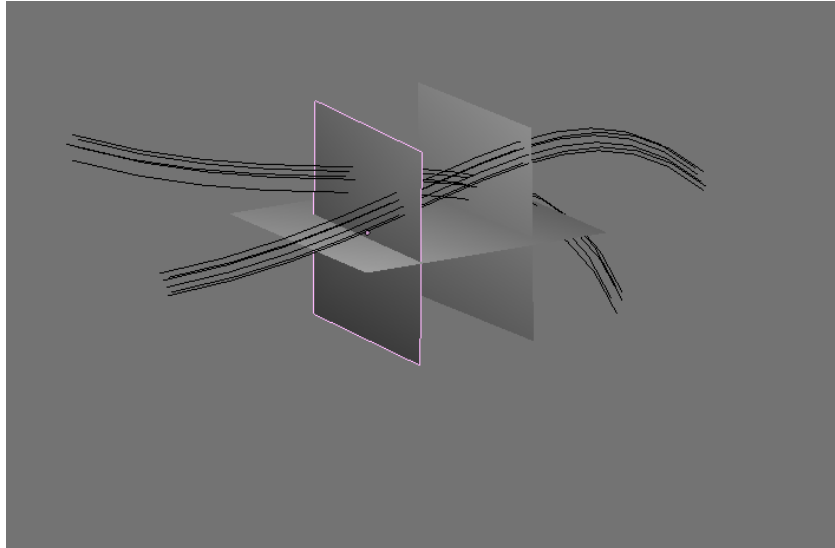


Figura 2.1 Las dos maneras de aproximarse al estudio de la evolución de los clústeres

Cuando revisamos la bibliografía sobre estos dos enfoques, observamos que, mientras sí había numerosos artículos sobre el primer enfoque (*Cluster Dynamics*), se notaba un vacío de aplicación bajo el prisma en el segundo enfoque. Es decir, en primer lugar observamos que ciertamente hay muchos artículos sobre evolución basados en Cadenas de Markov, pero raramente se refieren a evolución de clústeres. Por otra parte, en segundo lugar, los estudios en las Cadenas de Markov, quizá por deformación profesional matemática, tienden a preocuparse única y exclusivamente por el *límite*¹ al cual tienden los estados. Y, finalmente, en tercer lugar, muchos estudios sobre evolución de clústeres se centran en los individuos (sucesivos clústeres a los que pertenecen) pero no en los clústeres en sí (tomados como entidades que evolucionan).

A raíz de estas tres observaciones, pudimos enfocar nuestro estudio a este campo, ya que se trataba de una técnica novedosa, de aplicación inteligible intuitivamente, de alta potencia de resultados, y que se caracteriza por:

- a) Estudiar la evolución de los clústeres mirados como un sistema dentro de una Cadena de Markov.
- b) No considerar el paso al límite, sino las deducciones que se podían hacer observando dos o tres estados consecutivos.
- c) No centrarse en los individuos sino en los clústeres como entidades objeto de estudio.

2.1.2 Una idea intuitiva del método (metáfora de las dunas)

De los tres puntos señalados en el apartado anterior, el último es el de mayor originalidad por cuanto trabaja con los clústeres como entidades. Debido a ello, a los clústeres que entraban en juego se les exigió ciertas condiciones de regularidad, y se bautizó a un conjunto de clústeres que las cumpliera como un *Sistema consistente de clústeres*.

Con la construcción de un *sistema consistente de clústeres* buscamos, hablando metafóricamente, unas “dunas estadísticas”. Es decir, unas agrupaciones de individuos

¹ Aquí la palabra “Límite” se refiere a la operación matemática de cálculo de límites.

que cambian de intensidad (diferentes valores de las variables) y de integrantes (se incorporan y se pierden miembros del grupo) pero que son percibidas como entidades. La imagen física de las dunas del desierto es, pues, altamente ilustrativa: varían en forma (altura, pendiente, volumen, etc.) y se alimentan cada vez de nuevas partículas de arena en su desplazamiento, sin embargo, una filmación del desierto nos permite identificarlas como entidades.

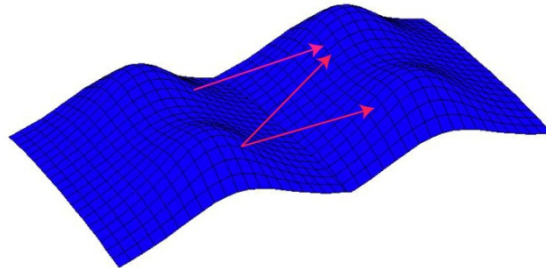


Figura 2.2 Metáfora de las dunas para ilustrar los sistemas consistentes de clústeres.

Para el tratamiento metodológico de nuestras entidades, siendo un tema de creación reciente, no fuimos demasiado restrictivos. Sin embargo, a efectos de consistencia, es obvio que conceptual y matemáticamente había que exigir las siguientes condiciones:

- a) Los clústeres espúreos deben eliminarse del sistema consistente.
- b) Los clústeres de aparición única deben etiquetarse y explicarse.
- c) Los clústeres admitidos deben superar las pruebas habituales de significación.
- d) Los clústeres deben reordenarse de manera que la interpretación sea cómoda y lo más directa posible.
- e) El número de variables de cada prueba, que produce los clústeres, debe permanecer constante.

Como puede observarse, no se ha exigido que el número de clústeres permanezca constante en las diferentes etapas de su evolución. Es más, el punto b) habla explícitamente de clústeres que aparecen en un paso y desaparecen en el siguiente. Se ha optado por esta opción para no dejar de lado muchos estudios en los que no existe una regularidad en el número de clústeres. Como veremos, ello no impide que podamos seguir, sin embargo, el rastro de las “dunas”.

2.2 Fundamentos conceptuales del enfoque centrado en el clúster

2.2.1 Definición de sistema consistente biunívoco

Sea una muestra de n individuos sobre la que realizamos una serie de m mediciones. Para cada medición i , tenemos un cierto número fijo de variables j , que tomamos como variables clasificadoras para construir un cierto número de clústeres k_i .

En tales condiciones, decimos que los clústeres resultantes forman un sistema consistente biunívoco si:

- a) El número de clústeres resultantes en cada medición es constante. Es decir, $k_i = k_{i+1}$, para $i=1\dots m-1$
- b) Las diferencias entre las puntuaciones de los clústeres son significativas para al menos una de las j variables

Observaciones

Si la cantidad de clústeres permanece constante, entonces podemos establecer una correspondencia biyectiva entre cada conjunto de clústeres con el resultante de la siguiente prueba de agrupación estadística. Si dicha correspondencia biyectiva tiene un significado estadístico y conceptual en el contexto del experimento general, entonces obtendremos una explicación satisfactoria del fenómeno estudiado.

La condición de existir al menos una diferencia significativa en los “perfiles” de los clústeres nos permite diferenciarlos como entidades. De hecho, es deseable que sean significativamente diferentes cuantas más variables mejor, pero en estudios multivariantes puede suceder que algunas variables no presenten diferencias significativas.

Un primer ejemplo

Como ejemplo de esta situación podemos pensar en un grupo de aprendizaje en un campus virtual y una serie de mediciones sobre la frecuencia de sus consultas en tres períodos lectivos (a efectos de fijar ideas, podemos suponer primer, segundo y tercer examen parcial). Supongamos que al observar una serie de variables indicativas de sus hábitos de consulta, observamos la aparición de 3 clústeres para el primer parcial:

- a) El primer clúster lo componen alumnos que consultan frecuentemente los materiales puestos a su disposición en el campus virtual.
- b) El segundo clúster lo componen alumnos que consultan frecuentemente sólo cierto tipo de materiales (por ejemplo, ciertos apuntes y determinadas prácticas, obviando los de otra naturaleza o incluso ignorando apuntes concretos).
- c) El tercer clúster, lo componen alumnos que raramente consultan ningún tipo de material.

Puede suceder que, al repetir el tratamiento estadístico en los otros dos parciales (segundo y tercero), obtengamos igualmente estos 3 clústeres. Imaginemos que al dibujar los perfiles de los mismos observamos que los 3 clústeres disminuyen su actividad con el tiempo; tal situación sería:

- a) El primer grupo consulta algo menos los materiales,
- b) El segundo grupo sigue consultando selectivamente materiales, pero con frecuencia menor a medida que pasan los parciales,
- c) El tercer grupo deja de consultar materiales a medida que pasan los parciales

Sin embargo, se distinguen claramente los 3 grupos en cada prueba de clústeres.

En tal situación, tenemos unos clústeres parecidos (posiblemente de la misma “forma” al dibujar su perfil de medias de variables de agrupación) y distinguibles. Esto nos permite tratarlos como entidades, es decir, hablar de *consistencia* de los clústeres. Como, además, tenemos el mismo número y una equivalencia perfecta entre ellos (correspondencia biyectiva), decimos que el sistema es *consistente biunívoco*.

Un ejemplo detallado

En el caso que seguiremos para ilustrar el método, disponemos de las notas de una muestra de alumnos de ingeniería (especialidad multimedia), tomadas en tres parciales. Para cada examen parcial, resultan 3 clústeres, con los perfiles que se muestran en las figuras siguientes. La asignación de los individuos a los diferentes clústeres se guarda en las variables CL_NOT1, CL_NOT2 y CL_NOT3 (Clúster de notas de cada uno de los parciales). Las barras representan la nota media del grupo para cada una de las asignaturas. Siendo P1 el primer parcial, P2 el segundo y P3 el tercero. Definimos ALG como Álgebra, CAL como Cálculo, FIS como Física, IO como Introducción a los Ordenadores, PGM como Tecnología y metodología de la Programación, ANG como Inglés, DIS como Diseño y USA como Usabilidad.

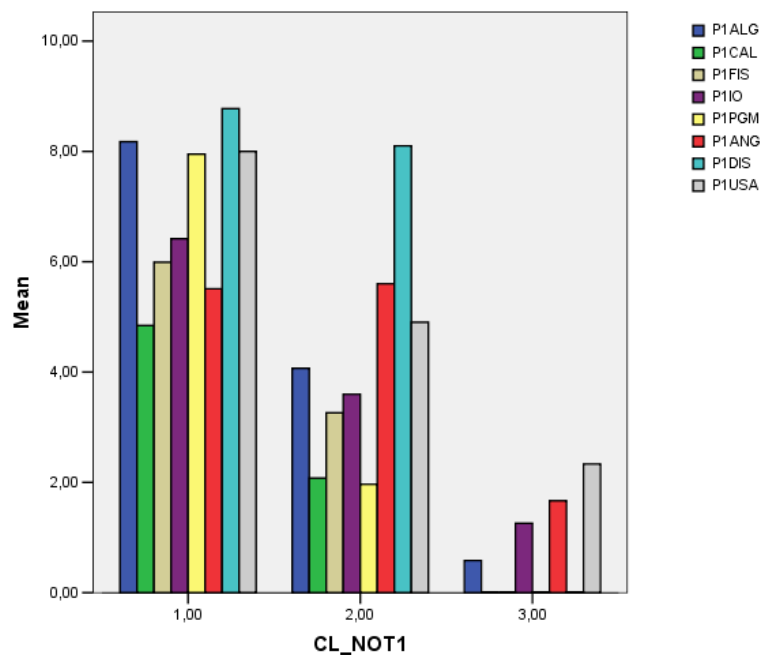


Figura 2.3 Puntuaciones resultantes para los 3 clústeres del primer parcial

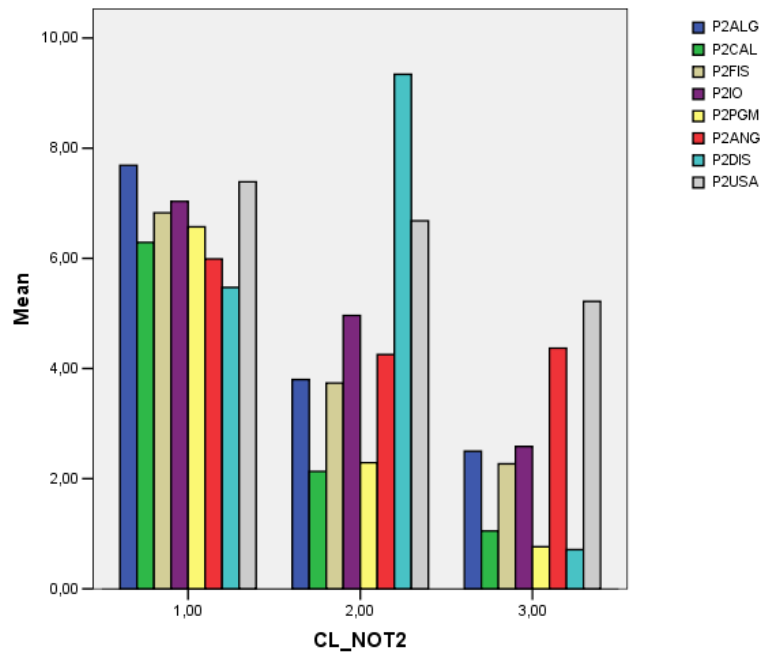


Figura 2.4 Puntuaciones resultantes para los 3 clústeres del segundo parcial

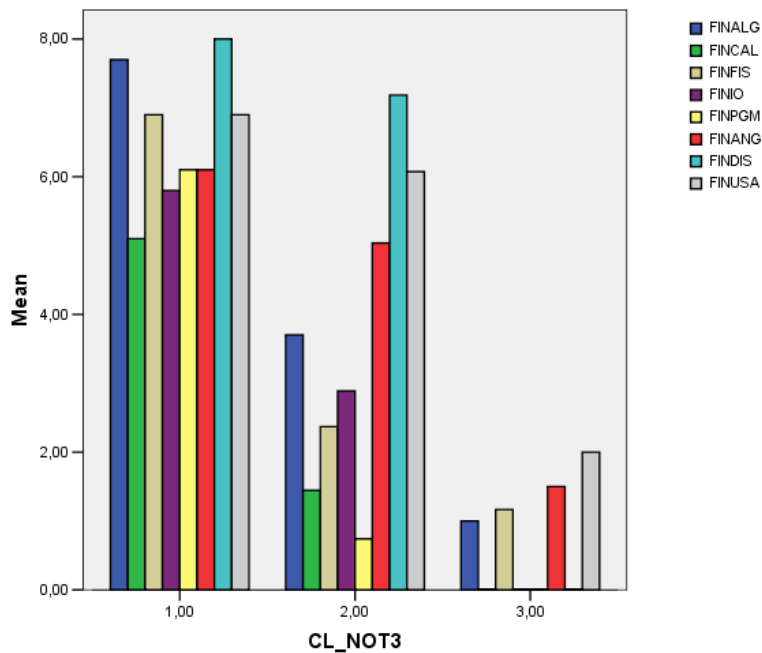


Figura 2.5 Puntuaciones resultantes para los 3 clústeres del tercer parcial

Para cada clasificación tenemos que asegurarnos que al menos una de las variables presenta diferencias significativas. Tal como puede verse en la siguiente tabla, para la primera clasificación lo son todas las variables (y lo mismo sucede para las otras clasificaciones):

		Sum of Squares	df	Mean Square	F	Sig.
P1ALG	Between Groups	183,290	2	91,645	22,465	,000
	Within Groups	163,175	40	4,079		
	Total	346,466	42			
P1CAL	Between Groups	78,841	2	39,421	22,366	,000
	Within Groups	70,500	40	1,762		
	Total	149,341	42			
P1FIS	Between Groups	99,271	2	49,635	13,983	,000
	Within Groups	141,988	40	3,550		
	Total	241,259	42			
P1IO	Between Groups	85,504	2	42,752	11,537	,000
	Within Groups	148,226	40	3,706		
	Total	233,730	42			
P1PGM	Between Groups	302,085	2	151,042	37,330	,000
	Within Groups	161,846	40	4,046		
	Total	463,931	42			
P1ANG	Between Groups	42,743	2	21,372	4,292	,020
	Within Groups	199,176	40	4,979		
	Total	241,919	42			
P1DIS	Between Groups	194,223	2	97,112	68,441	,000
	Within Groups	56,756	40	1,419		
	Total	250,980	42			
P1USA	Between Groups	103,130	2	51,565	10,748	,000
	Within Groups	191,896	40	4,797		
	Total	295,026	42			

Figura 2.6 Análisis de la varianza para las puntuaciones de los clústeres del primer parcial

A la vista de estas clasificaciones, vemos que resultan 3 grupos de estudiantes:

- a) Los que presentan un rendimiento notable (primer clúster de las 3 clasificaciones).
- b) Los que presentan un rendimiento bajo pero buscan puntuar en las asignaturas concretas de inglés, diseño o usabilidad (segundo clúster).
- c) Los que presentan un rendimiento muy bajo (tercer clúster).

En tal situación, tenemos unos clústeres parecidos (posiblemente de la misma “forma” al dibujar su perfil de medias de variables de agrupación) y distinguibles. Esto nos permite tratarlos como entidades, es decir, hablar de *consistencia* de los clústeres. Como, además, tenemos el mismo número y una equivalencia perfecta entre ellos (por lo que podemos establecer una correspondencia biyectiva), decimos que el sistema es *consistente biunívoco*.

Ejemplos de aplicación a la investigación educativa

La aparición de un *sistema consistente biunívoco*, en tanto que nos permite hablar de los clústeres como entidades, nos permite realizar preguntas sobre los individuos en relación a los mismos. Véase la siguiente figura 2.7:

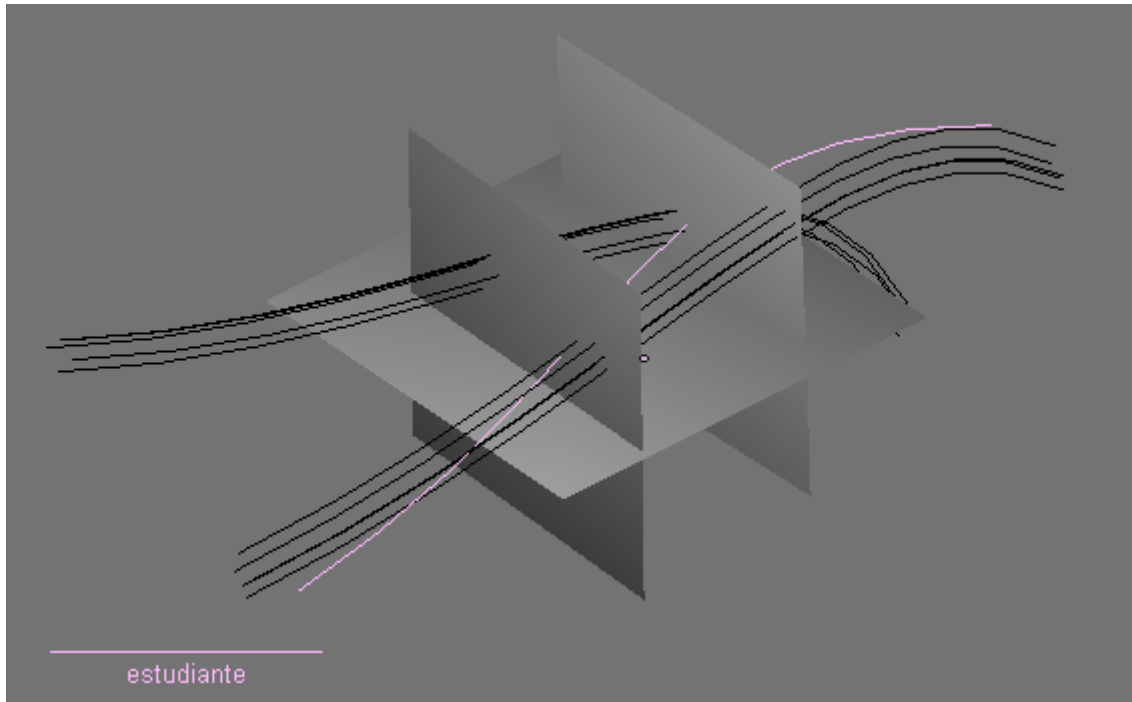


Figura 2.7 Sistema consistente biunívoco y observación de un individuo particular

La Figura 2.7 nos indica que el alumno en cuestión ha cambiado cada vez de clúster para ir a peor. La inspección de este caso y otros similares (o de sentido opuesto) nos permitiría sacar conclusiones sobre la acción tutorial, la dinamización del campus virtual o de aspectos generales de la vida en el centro. Es claro que se puede combinar una metodología cuantitativa que se ha tomado de partida con un estudio en profundidad del individuo, ya fuera también cuantitativo (observaciones sobre muchas variables hechas sobre dicho sujeto) o cualitativa (metodologías de profundización en la realidad educativa).

2.3 Detalle de los aspectos metodológicos

Una vez que hemos ilustrado el método con estos ejemplos de aplicación, vamos a volver al tema de los detalles metodológicos para tratar las cuestiones clave de los cálculos y decisiones al trabajar con sistemas consistentes de clústeres.

En el apartado 2.1.2 hemos señalado 5 características de los sistemas consistentes de clústeres:

- a) Los clústeres espurios deben eliminarse del sistema consistente.
- b) Los clústeres de aparición única deben etiquetarse y explicarse.
- c) Los clústeres admitidos deben superar las pruebas habituales de significación.
- d) Los clústeres deben reordenarse de manera que la interpretación sea cómoda y lo más directa posible.
- e) El número de variables de cada prueba que producen los clústeres debe permanecer constante.

La condición e), como puede entenderse, es una restricción de naturaleza axiomática. Es decir, quizá se podría hablar en cierto sentido de “consistencia” con un número de variables clasificatorias variable, pero intuitivamente se hace difícil en una primera aproximación al concepto. En la presente tesis, y en el artículo que sirve de base para el marco teórico, se optó por una definición que incluyera tal restricción. Se deja, en todo caso, para posteriores investigaciones un desarrollo posible de los sistemas consistentes de clústeres con un número de variables clasificadoras no constante a lo largo de las diferentes etapas de la evolución.

La condición b) no es de naturaleza estadística sino conceptual, ya que se refiere al rigor en la interpretación. Es decir, en un sistema no biunívoco de clústeres (en el que, al no tener un número constante de grupos, unos grupos pueden aparecer y desaparecer en el siguiente paso) es importante explicar, precisamente, por qué se trata de un sistema de tal tipo. Por tanto, la condición b) es simplemente una advertencia para que no se interpreten tales clústeres a la ligera. Es decir, es una condición de rigor que nos recuerda que los clústeres de aparición única no pueden obviarse, ya que se trata de elementos fundamentales en el sistema consistente, e imprescindibles para explicar el fenómeno estudiado.

Hechas estas observaciones, las otras 3 condiciones se explican con más detalle en los 3 apartados siguientes.

2.3.1 Eliminación de clústeres espurios: Tratamiento cualitativo o experiencia del investigador

En general, en estadística hablamos de un resultado “espurio” cuando se produce causado por alguna excepcionalidad. Normalmente nos referimos con esta denominación a puntuaciones especiales de ciertos individuos que no tienen que ser consideradas en el estudio del fenómeno.

Por ejemplo, en un estudio sobre rendimiento podríamos encontrarnos con que un alumno hubiera hecho trampas en la prueba de evaluación, de manera que su calificación contradijera todas las tendencias. En este caso, sería un resultado doblemente “espurio” (estadística y legalmente). Normalmente, en una investigación como la presente (en la que por su carácter ex post facto se toman datos de la realidad sin manipulación por parte del experimentador) aparecen resultados espurios debido a condiciones especiales de los alumnos, tales como: cursar una sola asignatura porque el estudiante proviene de otra especialidad (libre elección), cursar una sola asignatura porque es la que le queda para conseguir el título, no presentarse a unas pruebas por enfermedad, abandonar a mitad de curso por traslado de domicilio u otros motivos, etc.

Las investigaciones en las que se usan clústeres pueden dar buenos resultados con muestras no excesivamente grandes (decenas de individuos). Pueden funcionar también en grupos pequeños de individuos si el número de variables es grande (puesto que la matriz de datos es numerosa y se dispone, por tanto, de un elevado número de muestras). Sin embargo, en estos estudios aparecen fácilmente clústeres formados por un individuo o un número reducido de ellos (dos, tres, cuatro...). Este resultado se da porque, al tener un carácter especial, los valores del individuo diferente discrepan de todos los demás individuos, y discrepan en una magnitud mucho mayor que las

discrepancias entre grupos. Con tales longitudes en las distancias, el sistema estadístico le asigna, obrando coherentemente, un clúster aislado cuando aplica el algoritmo de agrupación. Puede decirse que la aparición de estos clústeres sin significado grupal es el precio que se paga por disponer de un tratamiento estadístico aplicable a conjuntos reducidos de individuos.

En estos casos, es responsabilidad del investigador tomar una decisión sobre los clústeres minoritarios basándose en su conocimiento de campo donde se ha realizado el experimento. Metodológicamente, deberían usarse técnicas cualitativas de exploración para poder pronunciarse, en rigor, sobre la validez o no de cada clúster singular. Sin embargo, es costumbre académica conceder un margen de confianza al investigador para que justifique su decisión frecuentemente sin acudir a ninguna técnica de soporte. En general, además, la singularidad suele quedar bastante justificada a los individuos de los clústeres aislados.

Adjuntamos un ejemplo de decisión de este tipo, en una investigación sobre patrones de rendimiento académico de la Universidad Pontificia Comillas de Madrid, por su valor ilustrativo. En ella, el investigador ataca el problema de 2 clústeres minoritarios que, una vez inspeccionados, califica de espurios:

“Estos dos conglomerados están integrados cada uno por tan solo un alumno cada uno. Realmente lo que el análisis detecta es una situación atípica o fuera de lo común. En este sentido, una vez identificados los alumnos que integran cada uno de los conglomerados se entienden las razones que dan lugar a esta situación. Así el conglomerado 7 se caracteriza por obtener unos resultados excepcionales en los factores 2 (Factor Lingüístico Humanístico) y 3 (Factor Empresarial), y unos resultados pésimos en los factores 1 (Factor Cuantitativo) y 4 (2º Idioma). La razón de esta situación es que se trata de un alumno que tuvo problemas personales que le impidieron presentarse a las asignaturas de estadística, Macroeconomía, Matemáticas Financieras y 2º Idioma, en la convocatoria de junio, por lo que la nota que se le ha adjudicado en dicha convocatoria ha sido de un cero.

Por su parte el conglomerado 2 se caracteriza por presentar unos resultados excepcionales en los factores cuantitativo y 2º idioma, aceptables en el factor cuantitativo y pésimos en el factor empresarial, quedando esto último justificado por el hecho de que tampoco se presentó en junio a Microeconomía, Marketing y Gestión Empresarial.

Estos dos conglomerados se corresponden, por lo tanto, con dos tipologías atípicas, formados por alumnos que en general no deberían de tener problemas para aprobar en primera convocatoria pero que por diferentes razones han decidido dejar algunas asignaturas para septiembre.” (Rúa, 2001)

Sin embargo, un tercer clúster minoritario sí es incluido en el conjunto final de clústeres que deben considerarse para la interpretación:

“Este conglomerado está integrado por 3 alumnos. Dos del Programa Francés y uno del programa Hispano-Americano. Se trata del grupo con los peores resultados académicos. Los tres alumnos van mal en todas las asignaturas, obtienen resultados pobres en el factor cuantitativo, y pésimos en el resto de los factores. Se trata del conglomerado con el peor rendimiento académico.

Esta tipología esta identificando aquellos alumnos que por las razones que sea han obtenido unos resultados muy malos en todas las asignaturas. Se trataría del grupo de alumnos que prácticamente no han estudiado nada a lo largo del curso.” (Rúa, 2001)

2.3.2 La reordenación con criterios correlacionales

Debe tenerse en cuenta que la ordenación de clústeres en los algoritmos jerárquicos de clasificación se produce según establecen las distancias de los individuos. Este hecho debe considerarse a efectos de lectura e interpretación de los resultados. Es decir, no siempre obtendremos unos esquemas “ordenados” de clústeres como los de las figuras 2.3, 2.4 y 2.5.

En el caso que nos ocupa, se ha ordenado la variable CL_NOT1 de acuerdo a un diagrama “ascendente” de clústeres de rendimiento (figura 2.3). Sin embargo, las figuras 2.4 y 2.5 se han elaborado reordenando la etiquetación de los clústeres de CL_NOT2 y CL_NOT3, de manera que la correlación entre los clústeres de las tres variables fuera máxima. Como se trata de un sistema consistente con 3 clústeres, operativamente, dada la variable CL_NOT? (el “?” es 2 ó 3 según el examen parcial que se trate), se construyen cinco variables nuevas, correspondientes a las 5 permutaciones diferentes de la Identidad:

Permutación	Inicial	Final	Inicial	Final	Inicial	Final
ABC	1	1	2	2	3	3
ACB	1	1	2	3	3	2
BAC	1	2	2	1	3	3
BCA	1	3	2	1	3	2
CAB	1	2	2	3	3	1
CBA	1	3	2	2	3	1

Con esta nomenclatura, la variable CL_NOT2BAC, por ejemplo, será la variable CL_NOT2 en la que se ha hecho un RECODE según los valores de la tercera línea de la tabla. La tabla de correlaciones de Pearson presenta cierta simetría en los coeficientes por motivos obvios, ya que algunas permutaciones son inversas a las otras. La redundancia en estas pruebas puede eliminarse, si bien es un buen recurso para detectar errores en la recodificación.

2.3.3 La intensidad de transiciones basada en el coeficiente de Contingencia

Llegados a este punto, el lector entenderá que la información suministrada por los sistemas consistentes de clústeres se basa, esencialmente, en observar y explicar su evolución. Cuestiones como “en qué clústeres concretos se *rompe* un clúster determinado”, “de qué clústeres anteriores procede un clúster en la clasificación actual” o bien “cómo se *desplaza* cierto clúster” son las que se pueden responder con las gráficas y pruebas asociadas a este método.

Uno de los estadísticos que utilizaremos para medir la intensidad de las relaciones entre los clústeres será el coeficiente de contingencia, aplicado a las variables de pertenencia a cada clúster dado.

En el caso que nos ocupa, vemos que la asociación entre el paso 2-3 es mayor que el 1-2, ya que el coeficiente de contingencia entre CL_NOT1 y CL_NOT2 es de 0,518 mientras que entre CL_NOT2 y CL_NOT3 es de 0,652:

Symmetric Measures CL_NOT1 * CL_NOT2

	Value	Approx. Sig.
Nominal by Nominal Contingency Coefficient	,518	,003
N of Valid Cases	43	

Symmetric Measures CL_NOT2 * CL_NOT3

	Value	Approx. Sig.
Nominal by Nominal Contingency Coefficient	,652	,000
N of Valid Cases	43	

2.4 Gráficas descriptivas de evolución de los clústeres de un sistema consistente

Para muestras no demasiado grandes, de entre 30 y 100 individuos, las gráficas de líneas dan una idea de la estabilidad y la evolución de los clústeres de un sistema consistente.

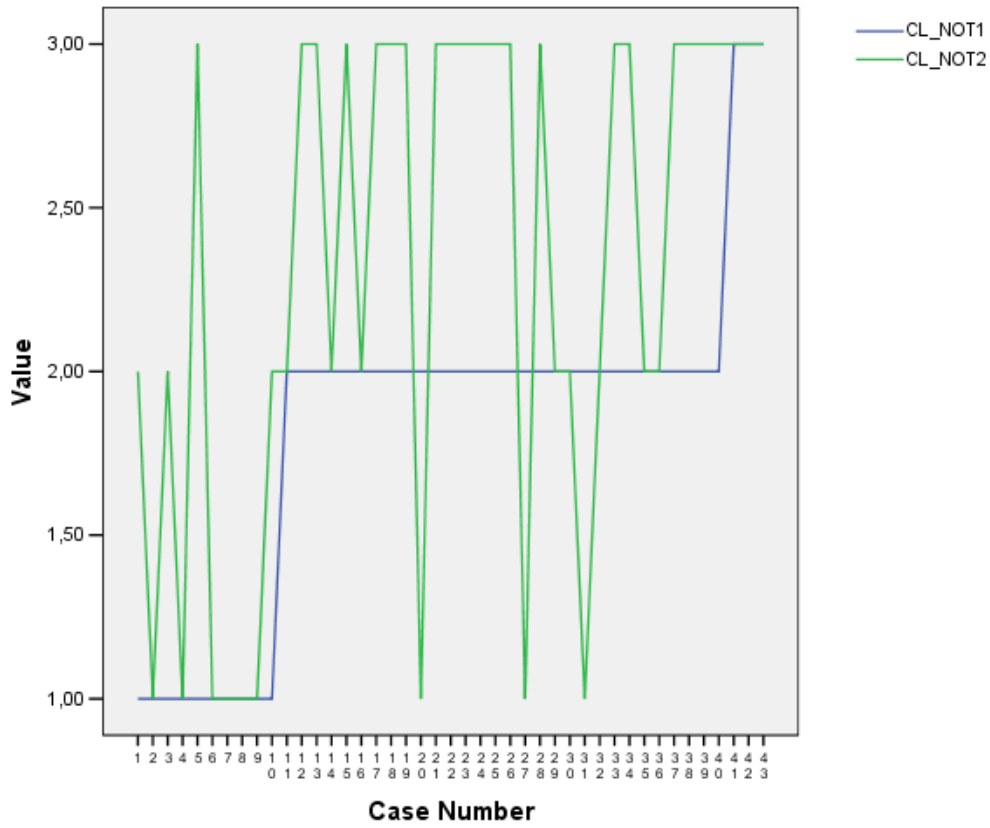


Figura 2.8. Gráfica de pertenencia a los clústeres del primer parcial superpuesta con la del segundo

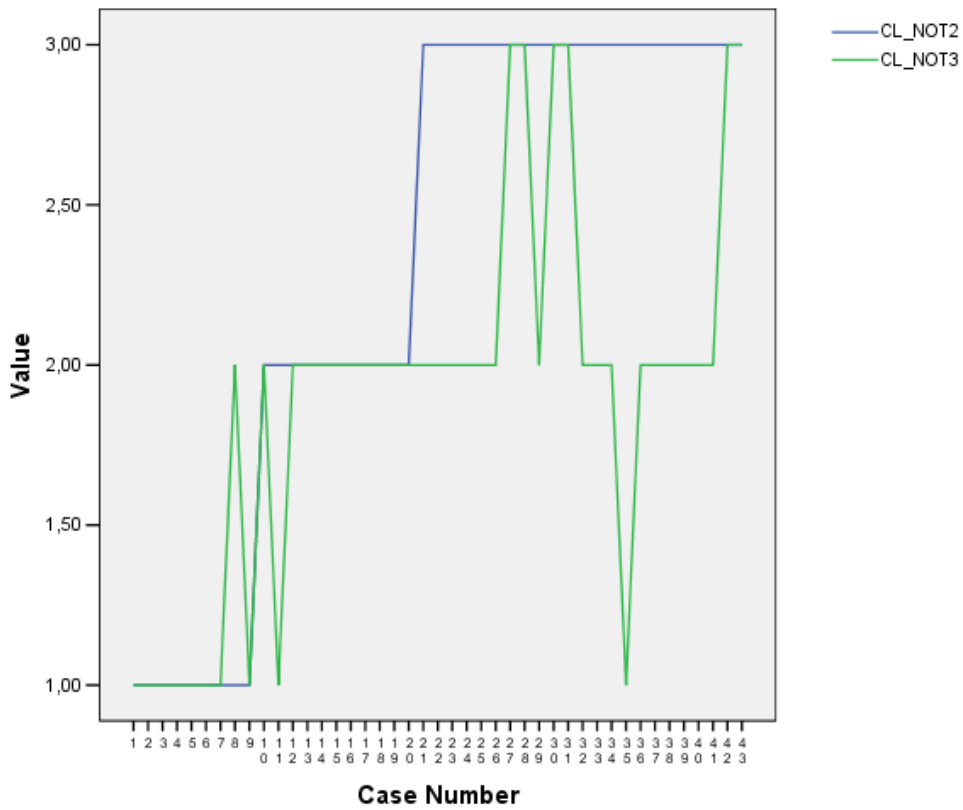


Figura 2.9 Gráfica de pertenencia a los clústeres del segundo parcial superpuesta con la del tercero

Para completar la información suministrada por estas gráficas, se aplica una prueba de tablas de contingencia en la que se incluyen los desgloses de la evolución de los clústeres, es decir, las cantidades de individuos de cada clúster que van a parar a los de la siguiente clasificación.

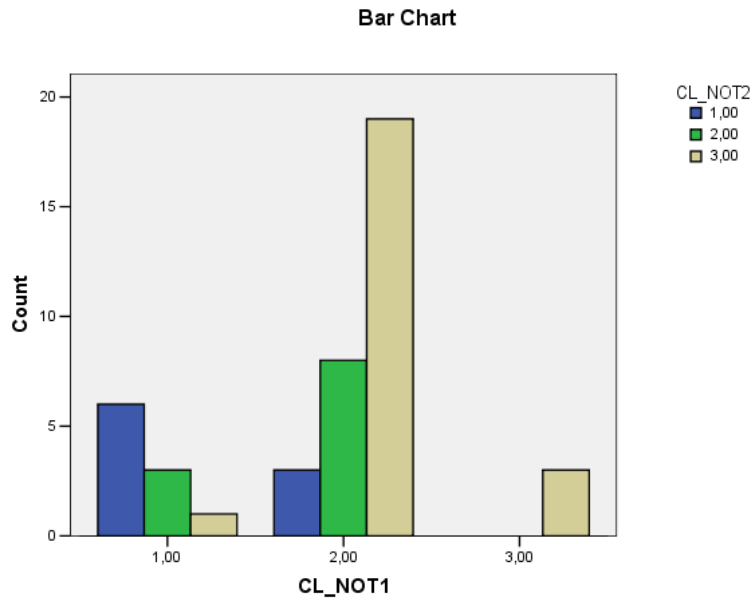


Figura 2.10 Evolución de los clústeres del primer parcial hacia el segundo, en cifras absolutas

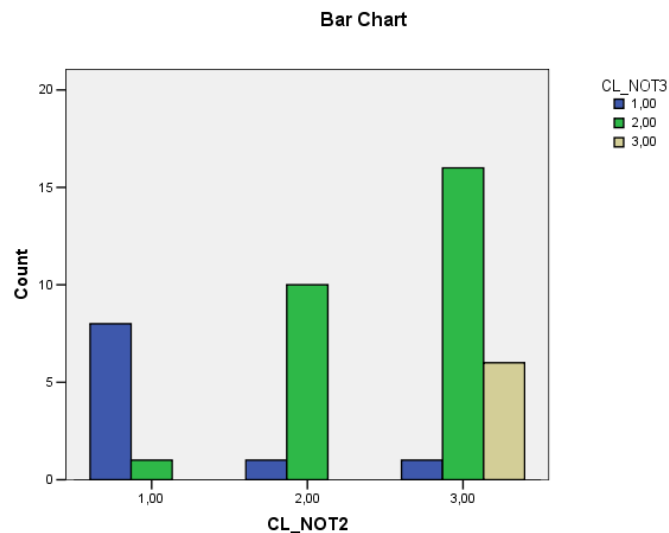


Figura 2.11 Evolución de los clústeres del segundo parcial hacia el tercero, en cifras absolutas

2.5 Método de combinación con cadenas de Markov para la evaluación de la intervención académica institucional

Las figuras anteriores anuncian la idea del método. Se trata simplemente de mirar el resultado final esperado, mediante la aplicación sucesiva de una matriz de estados y medir la discrepancia con el resultado final.

Las suposiciones experimentales del método son:

- Existen tres pruebas de evaluación, de las cuales resultan siempre 3 clústeres emparentados (sistema consistente).
- Del apartado a), por tanto, resultan 2 matrices de transiciones.
- El centro pone en marcha sus mecanismos de tutoría y ayuda al alumno de manera activa a mediados del segundo parcial y, de manera intensa, después de las notas de este segundo parcial (para encarar el final de curso).
- Del apartado c), por tanto, podemos suponer que la matriz de transición entre el primer y segundo parcial es la correspondiente a una intervención leve del centro o a una no intervención.
- El resultado final obtenido, si no hubiera habido intervención intensa del centro, hubiera sido, por tanto, la aplicación de la matriz de transición entre el primer y segundo parcial por dos veces consecutivas (deriva académica natural del grupo).
- Así pues, la discrepancia entre este final esperado y el real nos mide la actuación del centro.

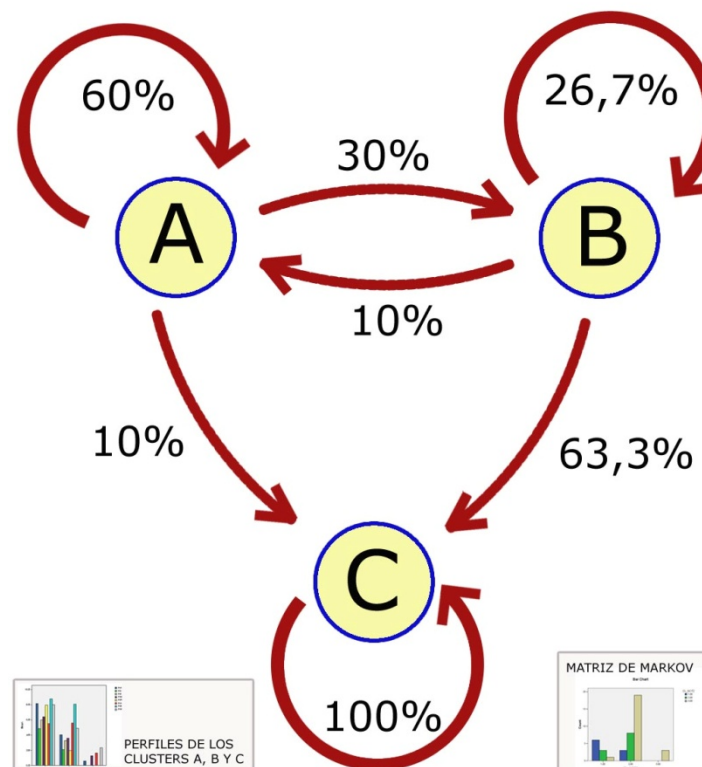


Figura 2.12 Dígrafo de probabilidades de transición entre clústeres del primer y segundo parcial

En la figura 2.12 hemos dibujado el dígrafo de probabilidades de transición entre los clústeres de la primera y segunda prueba para visualizar mejor el método.

La hipótesis de investigación de que la intervención del centro no es significativa equivale, estadísticamente, a la aplicación por segunda vez de la matriz de transiciones, para pronosticar los clústeres del tercer parcial. De este modo, bajo esta suposición, resultaría:

$$\begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.267 & 0.633 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.267 & 0.633 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.39 & 0.2601 & 0.34099 \\ 0.0867 & 0.101289 & 0.821011 \\ 0 & 0 & 1 \end{pmatrix}$$

Multiplicando el vector de estados iniciales por ésta matriz, obtenemos el estado final esperado. Ahora, con una simple prueba Chi cuadrado, podemos saber la diferencia entre este estado final y el que se da en la realidad, junto con su grado de significación. Por tanto, se obtiene con este método simple de dos pasos una estimación de la eficiencia de la intervención académica del centro.

El caso que nos ha servido de ejemplo forma parte de una investigación en la que se quiere observar las diferencias académicas entre ingenieros multimedia e ingenieros de otras especialidades. La investigación nos informará de la capacidad formativa de recuperación y corrección de errores del centro para cada uno de los estudios.

2.5.1 Esquema general de las investigaciones realizadas con el presente método

Al considerar no los individuos en sí como objeto de estudio, sino las agrupaciones de ellos, hemos visto que se obtiene un retrato explicativo de lo que sucede en el rendimiento de los alumnos. Este retrato es el punto de partida para una investigación multivariante más compleja, pero que puede ser planteada según la figura 2.13.

Es decir, las tres clasificaciones que se obtienen en los parciales son las “entradas” y “salidas” de la acción formativa del centro (con diferente intensidad).

La desviación de la tendencia natural es el indicador del éxito formativo del centro.

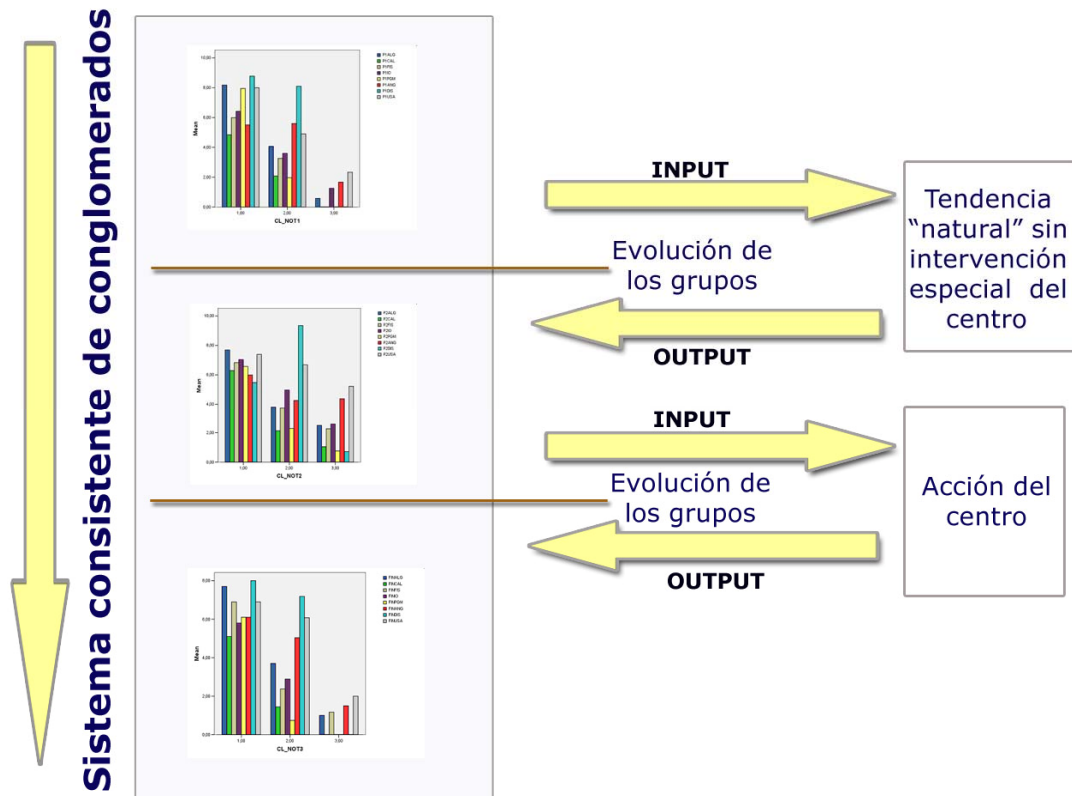


Figura 2.13 Esquema de una investigación con el método de los sistemas consistentes de clústeres

Capítulo 3

Definición del problema de investigación

3.1 Introducción: una investigación de doble vertiente

Hemos desarrollado un apunte metodológico en la presente investigación por tener éste, precisamente, un peso significativo en ella. Es decir, la presente tesis no sólo se ocupa de un tema inexplorado (las diferencias académicas constatables entre dos tipos de estudiantes de ingeniería) sino también de poner en práctica una metodología novedosa, simple de entender y potente en los resultados, que puede ser aplicada para evaluar la capacidad formativa de los centros docentes.

Con estas consideraciones, a la hora de definir el problema de investigación, hay consenso, en general, que nace de tres posibles causas:

- a) Existencia de una laguna en el conocimiento.
- b) Percepción de aparentes contradicciones en investigaciones anteriores.
- c) Observación de un fenómeno nuevo.

En nuestro caso la razón principal de la tesis, por lo que se refiere al contenido, es la primera causa, ya que no se dispone de suficientes estudios (tal como se ha explicado en el capítulo 1) comparativos entre la ingeniería multimedia y el resto de ingenierías. No obstante, esta misma laguna en el conocimiento se da en la metodología de investigación puesto que, tal como hemos indicado en el capítulo 2, los métodos correlacionales son insuficientes en determinados casos e, igualmente, las simples clasificaciones estadísticas producen resultados pobres.

Por tanto, admitida la dualidad del conocimiento científico en contenido y método, puede decirse que es la presente una “investigación de doble vertiente”, por cuanto su problema central tiene estas dos dimensiones.

Por otra parte, la argumentación anterior se refuerza si consideramos las condiciones que debe reunir un problema de investigación. Hay que señalar que las condiciones de realidad, factibilidad, relevancia y delimitación, exigibles a los planteamientos de los problemas en las investigaciones, emanan de lo expuesto en el capítulo 1; sin embargo, la resolubilidad (McGuigan, 1977; Kerlinger, 1985) emana en esta tesis de las aportaciones metodológicas expuestas en el capítulo 2. Del mismo modo, si miramos esta investigación inmersa en una teoría de fondo o valoramos su capacidad de generación de conocimiento, debemos convenir que la aportación metodológica es importante en ambos puntos (de hecho, en el momento de leer esta tesis, está llegando a su fase final una segunda investigación que nace de la misma metodología y la desarrolla de diferente manera).

Hechas estas observaciones, podemos pasar a señalar los objetivos de la investigación y a exponer la formulación del problema.

3.2 Objetivos de la investigación

Tal como señalan Arnal, Del Rincón, y Latorre (1992), un problema de investigación es “en sentido amplio, cualquier situación sin una solución satisfactoria”. Esta acepción surge de la revisión bibliográfica hecha tanto sobre múltiples investigaciones como sobre diferentes aportaciones de autores de metodología. Así pues, acogiéndonos a esta concepción amplia del concepto, podemos señalar los objetivos en la doble vertiente que se ha señalado en el apartado anterior.

Por tanto, la presente tesis se marca como objetivo general, el formulado de la siguiente manera:

Objetivo general de la investigación

Distinguir los perfiles académicos de los estudiantes de Multimedia y los de ingenierías estrictamente tecnológicas pero aportando una metodología de análisis de la realidad académica novedosa, basada en la evolución de clústeres a lo largo del curso.

3.3 Enunciado del problema

Según Kerlinger (1985) el enunciado del problema debe:

- a) Expresar una relación entre dos o más variables.
- b) Estar claramente planteado y, a ser posible, en forma de pregunta.
- c) Permitir verificación empírica.

Si tenemos en cuenta que Kerlinger es un autor adscrito a la psicología experimental, entenderemos que, a la hora de considerar su aportación en otras ciencias sociales (algunas menos precisas en la experimentación o la naturaleza de los resultados), sea tradicional permitir ciertas licencias en las formulaciones de los problemas de investigación.

En el caso que nos ocupa, señalaremos de qué manera se han aplicado las condiciones de Kerlinger a la formulación del problema:

- a) Expresar una relación entre dos o más variables.

Nuestra investigación se basa, precisamente, en un análisis de las relaciones entre muchas variables y, por añadidura, de manera evolutiva a través del tiempo. Para cada individuo se han tomado 7 u 8 variables de calificaciones en cada una de las 3 pruebas parciales, así como igual número en las diversas pruebas de seguimiento, así como variables ligadas a los grupos de aprendizaje, el sexo, el tipo de titulación y otras.

Estas variables han conformado los clústeres, que son la materia prima de la investigación, y a su vez, se ha estudiado la variación de estos clústeres aplicando modelos evolutivos de Markov.

Un enunciado que plasmase esta complejidad metodológica sería poco breve y difícilmente inteligible. Podemos asumir, por tanto, que incluyendo la referencia a un modelo explicativo basado en clústeres el lector científico entiende que se van a tratar relaciones entre diferentes variables.

b) Estar claramente planteado y, a ser posible, en forma de pregunta.

En aras de la claridad y la concisión se optará por una pregunta de unas pocas líneas, que exprese el concepto nuclear del problema: una diferencia entre dos tipos de alumnos (ingeniería multimedia e ingenierías estrictamente tecnológicas) que necesita de una explicación estadística para ser resuelto.

c) Permitir verificación empírica

La verificación empírica del problema son precisamente los capítulos 4, 5, 6 y 7 de la presente tesis, puesto que en ellos se da respuesta al problema planteado y se sientan las bases estadísticas para redactar un capítulo final de conclusiones.

3.4 Formulación del problema de investigación

Por tanto, el enunciado del problema de investigación que resuelve esta tesis es el siguiente:

Problema de investigación

¿Es posible encontrar un modelo explicativo basado en la evolución de clústeres estadísticos que explique el diferente comportamiento académico, por lo que se refiere a las calificaciones, de los estudiantes de ingeniería multimedia respecto del resto de estudiantes de las otras especialidades de ingeniería?

3.5 Consideraciones sobre las variables y las hipótesis derivadas del problema

Es frecuente en muchas investigaciones enumerar, a partir del problema de investigación, una serie de hipótesis que se derivan del mismo, ya sea para ampliar su enunciado, para señalar su aplicación a grupos concretos de individuos o para precisar aspectos particulares del texto general.

En esta investigación no cabe tal enumeración de hipótesis debido a la naturaleza extensa de la metodología utilizada. Los estudios de evolución de clústeres ofrecen unos modelos explicativos suficientemente exhaustivos y altamente clarificadores. En todo caso, en el apartado de conclusiones se expondrán las ideas principales debidamente clasificadas en los aspectos que emanan del estudio estadístico.

Por lo que se refiere a las variables puestas en juego, no estará de más incluir un listado clarificador para que el lector pueda tener siempre presente el significado de todas las mediciones. Este listado sirve de guía y es el que se reproduce a continuación.

Listado de variables

especialitat: que especialidad de ingeniería cursan.

sexe: alumno o alumna.

grup: clase a la que van.

mitjaexp: nota final del expediente de bachillerato.

mitjapaau: nota final de la selectividad teniendo en cuenta la nota de bachillerato.

npau: nota de las pruebas de selectividad.

nmates: nota del examen de matemáticas de la prueba de entrada a La Salle.

nfísica: nota del examen de física de la prueba de entrada a La Salle.

nfina: nota promedio de los exámenes de matemáticas y física de la prueba de entrada.

ncou: nota media aproximada tomada, por el responsable de admisiones, de las asignaturas de matemáticas y física de bachillerato.

algot: nota de evaluación continuada propuesta por el profesor de álgebra que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

caloct: nota de evaluación continuada propuesta por el profesor de cálculo que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

fisoct: nota de evaluación continuada propuesta por el profesor de física que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

eleoct: nota de evaluación continuada propuesta por el profesor de electrónica que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

iooct: nota de evaluación continuada propuesta por el profesor de introducción a los ordenadores que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

pgmoct: nota de evaluación continuada propuesta por el profesor de programación que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

angoct: nota de evaluación continuada propuesta por el profesor de inglés que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

dissoct: nota de evaluación continuada propuesta por el profesor de diseño que valora el trabajo del alumno en clase y en casa durante los meses de septiembre y octubre.

algtst: nota de un examen tipo test de álgebra de evaluación continuada realizada a principios de noviembre.

caltst: nota de un examen tipo test de cálculo de evaluación continuada realizada a principios de noviembre.

fistst: nota de un examen tipo test de física de evaluación continuada realizada a principios de noviembre.

eletst: nota de un examen tipo test de electrónica de evaluación continuada realizada a principios de noviembre.

-
- iotest: nota de un examen tipo test de introducción a los ordenadores de evaluación continuada realizada a principios de noviembre.
- pgmtest: nota de un examen tipo test de programación de evaluación continuada realizada a principios de noviembre.
- algnov: nota de evaluación continuada propuesta por el profesor de álgebra que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- calnov: nota de evaluación continuada propuesta por el profesor de cálculo que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- fisnov: nota de evaluación continuada propuesta por el profesor de física que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- elenov: nota de evaluación continuada propuesta por el profesor de electrónica que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- ionov: nota de evaluación continuada propuesta por el profesor de introducción a los ordenadores que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- pgmnov: nota de evaluación continuada propuesta por el profesor de programación que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- angnov: nota de evaluación continuada propuesta por el profesor de inglés que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- dissnov: nota de evaluación continuada propuesta por el profesor de diseño que valora el trabajo del alumno en clase y en casa durante todo el primer trimestre.
- algnad: nota del examen de álgebra del primer trimestre.
- calnad: nota del examen de cálculo del primer trimestre.
- fisnad: nota del examen de física del primer trimestre.
- elenad: nota del examen de electrónica del primer trimestre.
- ionad: nota del examen de introducción a los ordenadores del primer trimestre.
- pgmnad: nota del examen de programación del primer trimestre.
- angnad: nota del examen de inglés del primer trimestre.
- dissnad: nota del examen de diseño del primer trimestre.
- alg1p: nota final del primer trimestre de álgebra después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- cal1p: nota final del primer trimestre de cálculo después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- fis1p: nota final del primer trimestre de física después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- ele1p: nota final del primer trimestre de electrónica después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- io1p: nota final del primer trimestre de introducción a los ordenadores después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- pgm1p: nota final del primer trimestre de programación después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- ang1p: nota final del primer trimestre de inglés después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

- diss1p: nota final del primer trimestre de diseño después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.
- usa1p: nota final del primer trimestre teniendo en cuenta la nota del examen trimestral y la nota de las prácticas.
- algggen: nota de evaluación continuada propuesta por el profesor de álgebra que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- calggen: nota de evaluación continuada propuesta por el profesor de cálculo que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- fisggen: nota de evaluación continuada propuesta por el profesor de física que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- eleggen: nota de evaluación continuada propuesta por el profesor de electrónica que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- ioggen: nota de evaluación continuada propuesta por el profesor de introducción a los ordenadores que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- pgmgen: nota de evaluación continuada propuesta por el profesor de programación que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- anggen: nota de evaluación continuada propuesta por el profesor de inglés que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- dissggen: nota de evaluación continuada propuesta por el profesor de diseño que valora el trabajo del alumno en clase y en casa durante el mes de enero.
- algmarch: nota de evaluación continuada propuesta por el profesor de álgebra que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- calmarch: nota de evaluación continuada propuesta por el profesor de cálculo que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- fismarch: nota de evaluación continuada propuesta por el profesor de física que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- elemarch: nota de evaluación continuada propuesta por el profesor de electrónica que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- iomarch: nota de evaluación continuada propuesta por el profesor de introducción a los ordenadores que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- pgmmarch: nota de evaluación continuada propuesta por el profesor de programación que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- angmarch: nota de evaluación continuada propuesta por el profesor de inglés que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- dissmarch: nota de evaluación continuada propuesta por el profesor de diseño que valora el trabajo del alumno en clase y en casa durante todo el segundo trimestre.
- algpasq: nota del examen de álgebra del segundo trimestre.
- calpasq: nota del examen de cálculo del segundo trimestre.
- fispasq: nota del examen de física del segundo trimestre.
- elepaspq: nota del examen de electrónica del segundo trimestre.
- iopaspq: nota del examen de introducción a los ordenadores del segundo trimestre.
- pgmpaspq: nota del examen de programación del segundo trimestre.
- angpaspq: nota del examen de inglés del segundo trimestre.
- alg2p: nota final del segundo trimestre de álgebra después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

cal2p: nota final del segundo trimestre de cálculo después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

fis2p: nota final del segundo trimestre de física después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

ele2p: nota final del segundo trimestre de electrónica después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

io2p: nota final del segundo trimestre de introducción a los ordenadores después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

pgm2p: nota final del segundo trimestre de programación después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

ang2p: nota final del segundo trimestre de inglés después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

diss2p: nota final del segundo trimestre de diseño después de haber sumado una puntuación extra a la nota del examen, entre 0 y 2 puntos, según la nota de evaluación continuada del trimestre.

usa2p: nota final del segundo trimestre teniendo en cuenta la nota del examen trimestral y la nota de las prácticas.

algjny: nota final de junio de álgebra teniendo en cuenta las notas parciales.

caljny: nota final de junio de cálculo teniendo en cuenta las notas parciales.

fisjny: nota final de junio de física teniendo en cuenta las notas parciales.

elejny: nota final de junio de electrónica teniendo en cuenta las notas parciales.

iojny: nota final de junio de introducción a los ordenadores teniendo en cuenta las notas parciales y las prácticas.

pgmjny: nota final de junio de programación teniendo en cuenta las notas parciales y las prácticas.

angjny: nota final de junio de inglés teniendo en cuenta las notas parciales.

dissjny: nota final de junio de diseño teniendo en cuenta las notas parciales y los trabajos.

usajny: nota final de junio de usabilidad teniendo en cuenta las notas parciales y las prácticas.

Nota: Así como las variables de evaluación presentan una regularidad que permite un tratamiento estadístico riguroso, hay que señalar que la recogida de las variables de seguimiento no goza de tal regularidad. Al tratarse de una investigación *expostfacto*¹ aparecen los problemas derivados de la práctica cotidiana. En nuestro caso, ello implica la falta de algunas variables en las mediciones de seguimiento.

¹ La investigación *expostfacto* es aquella realizada una vez ya han ocurrido todos los hechos y no se puede manipular los resultados, ya que no ejercemos ningún control sobre las variables.

Capítulo 4

**Aplicaciones básicas de los sistemas consistentes de clústeres.
Predicción y seguimiento.**

Para la muestra de alumnos de multimedia se realizaron pruebas de clústeres para los tres tipos de variables objeto de estudio: predictoras, de evaluación y de seguimiento. Los alumnos de multimedia tienen una asignatura más que los de las ingenierías tecnológicas (8 en vez de 7 asignaturas).

En este capítulo veremos cómo se puede trabajar el enfoque de los sistemas consistentes de clústeres para formarse una primera idea de las diferencias entre ingeniería multimedia e ingenierías tecnológicas.

También se realizará una prospección sobre los clústeres de seguimiento.

4.1 Detección de los clústeres predictores en ingeniería multimedia

El dendograma de clústeres es el siguiente:

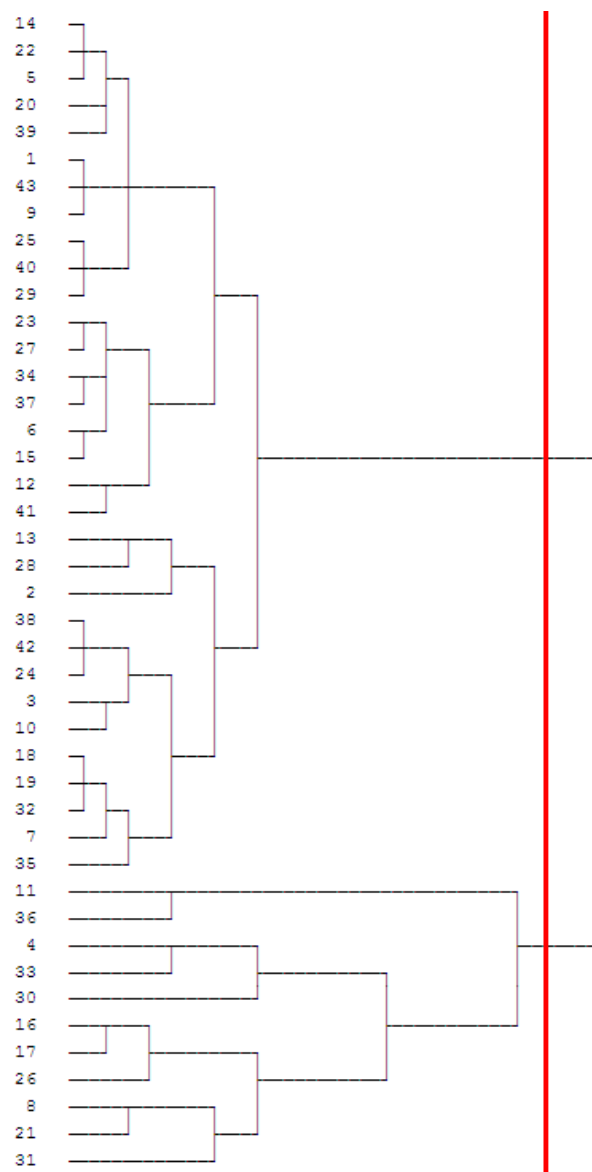


Figura 4.1 Dendograma de los alumnos de multimedia. Con la definición de los dos clústeres.

De donde vemos que resultan 2 grandes clústeres. El perfil académico de estos dos grupos, por lo que se refiere a las variables predictoras, es el siguiente:

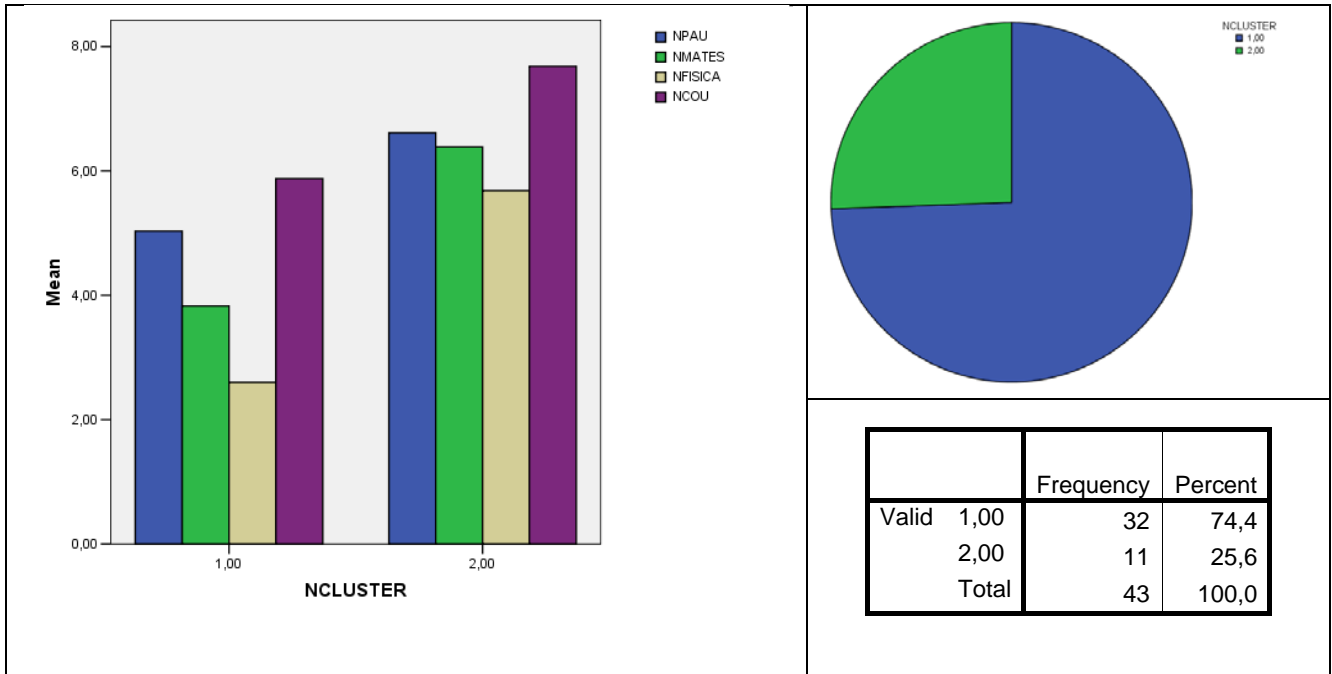


Figura 4.2 Diagrama de pertenencia a los clústeres de las variables predictoras y puntuaciones medias en las variables de agrupamiento.

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
NPAU	Equal variances assumed	6,644	,014	-5,547	41	,000
	Equal variances not assumed			-3,985	11,582	,002
NMATES	Equal variances assumed	2,354	,133	-4,767	41	,000
	Equal variances not assumed			-4,016	13,562	,001
NFISICA	Equal variances assumed	,530	,471	-6,747	41	,000
	Equal variances not assumed			-5,959	14,388	,000
NCOU	Equal variances assumed	1,833	,183	-4,954	41	,000
	Equal variances not assumed			-4,218	13,735	,001

Figura 4.3 Resultado de la prueba t de Student para comparación de medias de dos muestras. Tanto bajo la suposición de igualdad de varianzas como desigualdad, la prueba es significativa.

Por tanto, podemos hablar de dos clústeres con puntuaciones significativamente diferentes para las variables de agrupación en las variables predictoras de multimedia. La descripción de los mismos, se realiza a continuación.

4.1.1 Descripción de los clústeres

El clúster número 1, que es el más numeroso con 32 estudiantes, tiene un rendimiento medio. Con las notas de COU y de las PAU alcanzado o superando el 5, pero sin embargo en las pruebas de la Salle, están por debajo del aprobado.

El clúster número 2 tiene menos estudiantes, contiene 11 miembros. En este caso, tenemos a los alumnos de un rendimiento óptimo. Observamos que en las pruebas externas al centro tienen una nota bastante elevada, se sitúa entre el 6 y el 8. Y en las pruebas del centro observamos que son entre el 5 y el 7.

Semejanzas

Los perfiles tienen en común:

- a) Podemos observar que hay una relación en cuanto a la ordenación de las notas de los clústeres. La ordenación es COU, PAU, matemáticas y física.
- b) Existe una correlación entre las notas de las pruebas de La Salle y las que provienen de medidas externas al centro. Ya que observamos que el clúster de mejor rendimiento en las COU y en las PAU, también los es en las pruebas de la Salle. El clúster de rendimiento medio presenta también el mismo patrón siendo así, el de menores valores en las pruebas internas como en las externas.
- c) Llama la atención que, en ambos clústeres, las notas de matemáticas son sustancialmente mejores que las de física tanto en alumnos de alto rendimiento como en los de rendimiento medio.

Comentario: Debido a que la nota de matemáticas, en cualquier caso, es mejor que la de física, intuimos que los alumnos de ingeniería multimedia son más abstractos que empíricos.

Diferencias

Los perfiles se diferencian por:

- a) Aprobados La Salle contra suspendidos La Salle: la diferencia más significativa que podemos observar, es que el clúster 1 suspende en las pruebas de Salle, contra el clúster 2 que aprueba dicho control.
- b) La segunda diferencia que podemos deducir, es que los alumnos del primer clúster están aprobados rozando la nota mínima, contra los que aprueban con soltura y puntúan por encima del 6 en selectividad, que se corresponden con el segundo clúster.

4.1.2 Etiquetación de los clústeres

Por lo que se refiere a las variables predictoras en la ingeniería multimedia, establecemos 2 clústeres con las siguientes etiquetas y descripciones:

- a) Clúster número 1, Suspendidos Salle, regulares en COU:
 - Escogemos la relación más fuerte que se puede observar. Los alumnos que puntuaron por debajo del aprobado en las pruebas de La Salle, estuvieron justos en sus pruebas de COU y PAU.
- b) Clúster número 2, Aprobados Salle, altos en COU:
 - En este caso la relación se mantiene siendo así los alumnos con mejor puntuación en COU, los que aprueban en las pruebas de La Salle.

Se trata de alumnos con buenas calificaciones globales, sin embargo, mientras la nota de COU es elevada, la de La Salle queda muy por debajo. Las notas de La Salle son mucho más bajas debido a que se preguntan muchos temas que en las diferentes escuelas ajenas no han trabajado a conciencia.

4.2 Detección de los clústeres predictores en ingenierías tecnológicas

Para la muestra de alumnos de ingenierías tecnológicas también se realizaron pruebas de clústeres para los tres tipos de variables objeto de estudio: predictoras, de evaluación y de seguimiento. Ahora los alumnos tienen sólo 7 asignaturas.

4.2.1 Una primera corrección a la agrupación

Reproducimos en la figura 4.4 el dendograma resultante con un individuo en rojo que es la parte que queremos resaltar porque refleja bien un aspecto metodológico que se ha citado en el marco teórico. Se trata de un clúster formado por un solo alumno que debe ser analizado cualitativamente.

Una vez hecha la localización del alumno, como sucede frecuentemente en este tipo de estudios, lo que le confiere categoría de clúster no es más que su singularidad debido a circunstancias particulares. En el caso que nos ocupa, al ser un alumno que el análisis le aleja tanto de los demás que lo asigna un clúster. Esto es debido a que el alumno en concreto obtuvo una buena puntuación en física (6,5) y, sin embargo, muy deficiente en matemáticas (2). Este hecho no sigue el patrón marcado por los demás alumnos del estudio. En estos casos corresponde eliminar el caso puesto que no puede tener el peso de los demás.

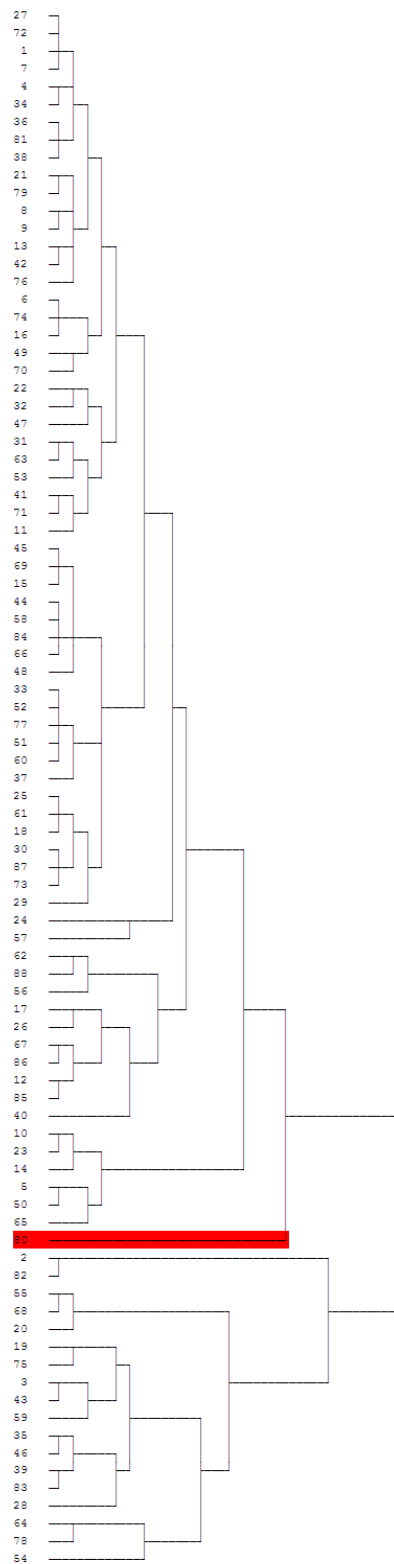


Figura 4.4 Dendrograma para ingenierías tecnológicas, variables predictoras, con clúster espurio.

4.2.2 La agrupación corregida

El dendograma resultante, quitado el caso espurio, nos sugiere una estructura de 4 grupos. Al ser un grupo numeroso, aplicamos un K-MEANS para forzar un modelo explicativo de 4 clústeres. Dado que este modo de proceder es desconocido para los lectores que no estén familiarizados con los análisis de clústeres, expondremos a continuación las decisiones que nos han llevado a elegir dicho tratamiento estadístico:

a) Punto de partida: dendograma con clústeres inclusores

Una ojeada al dendograma de la figura 4.4 nos muestra como existen grupos de semejanza entre los individuos pero, en vez de terminar en diferentes clústeres, pasan a formar parte de un mismo clúster final. Se trata de una evolución “inclusora” el realizar las iteraciones, es decir, los individuos semejantes pasan a ser absorbidos por un grupo precedente antes de desembocar en un grupo propio.

En la figura 4.5 hemos señalado estos grupos semejantes que, como puede verse, terminan los tres primeros de ellos en un clúster y el cuarto en otro.

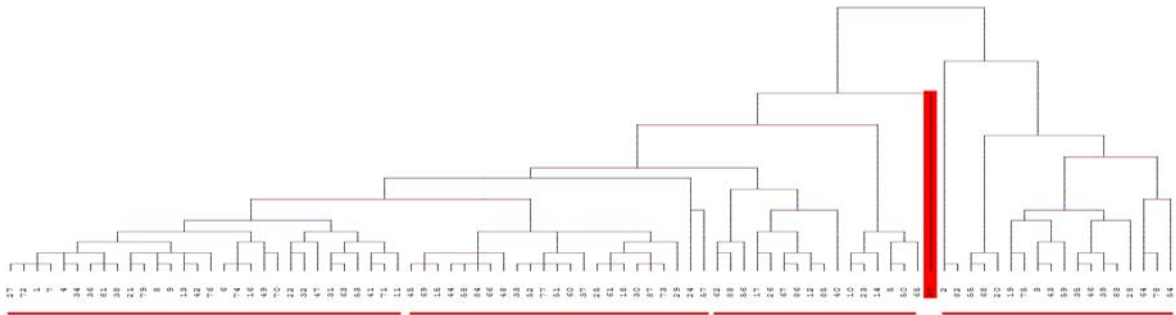


Figura 4.5 Dendograma para ingenierías tecnológicas, variables predictoras, con clúster espurio con grupos semejantes señalados.

b) Solución con clústeres forzados

A la vista de la figura anterior, y considerando las iteraciones que han sido necesarias para formar los grupos (longitud de las líneas de agrupación) nos preguntamos si no sería más adecuada una agrupación final que reflejara estos cuatro grupos de semejanza.

No obstante, para considerar dichos cuatro grupos, no se puede trazar una línea divisoria en el dendograma (que marca la iteración en que se detienen las agrupaciones) puesto que aparecen en diferentes profundidades. Así pues, se impone realizar una prueba K-MEANS en la que se pide al programa estadístico que fuerce un número concreto de clústeres.

c) Tanteo con diferentes K-MEANS

En pruebas de este tipo es habitual solicitar diferentes K-MEANS, con diferentes grupos finales, y posteriormente comprobar la solidez de los clústeres obtenidos mediante pruebas ANOVA aplicadas a los perfiles de los grupos.

En nuestro caso hemos probado agrupaciones de 3, 4 y 5 grupos, pero sólo presentaban perfiles suficientemente diferentes los grupos obtenidos en la clasificación de 4.

d) Elección del K-MEANS de 4 grupos

Por tanto, disponemos de dos argumentos que nos indican la conveniencia y la corrección de realizar una prueba K-MEANS de 4 grupos. El primer, que el dendograma así lo ha sugerido; y el segundo, que sus resultados de perfiles han sido significativos, rechazando las clasificaciones inmediatamente inferior y superior.

Con tal enfoque, obteneos un perfil de clústeres como el siguiente:

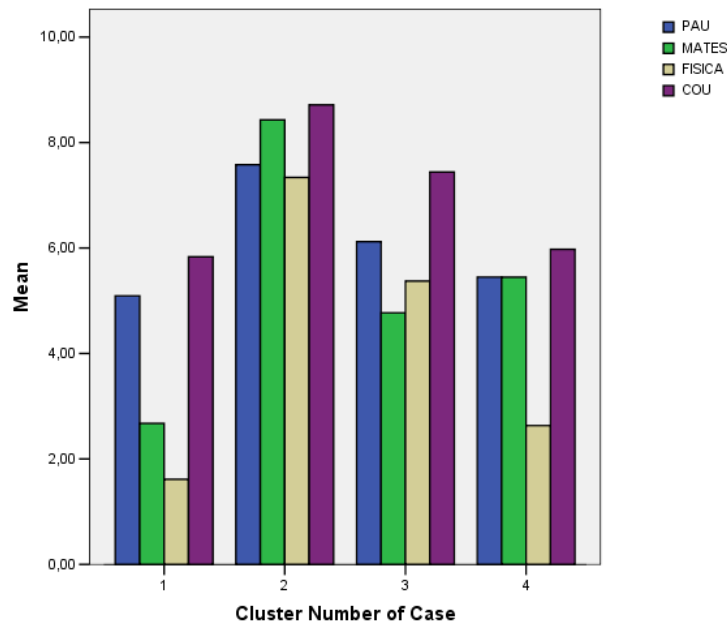


Figura 4.6 Perfiles resultantes de los 4 clústeres de las variables predictivas.

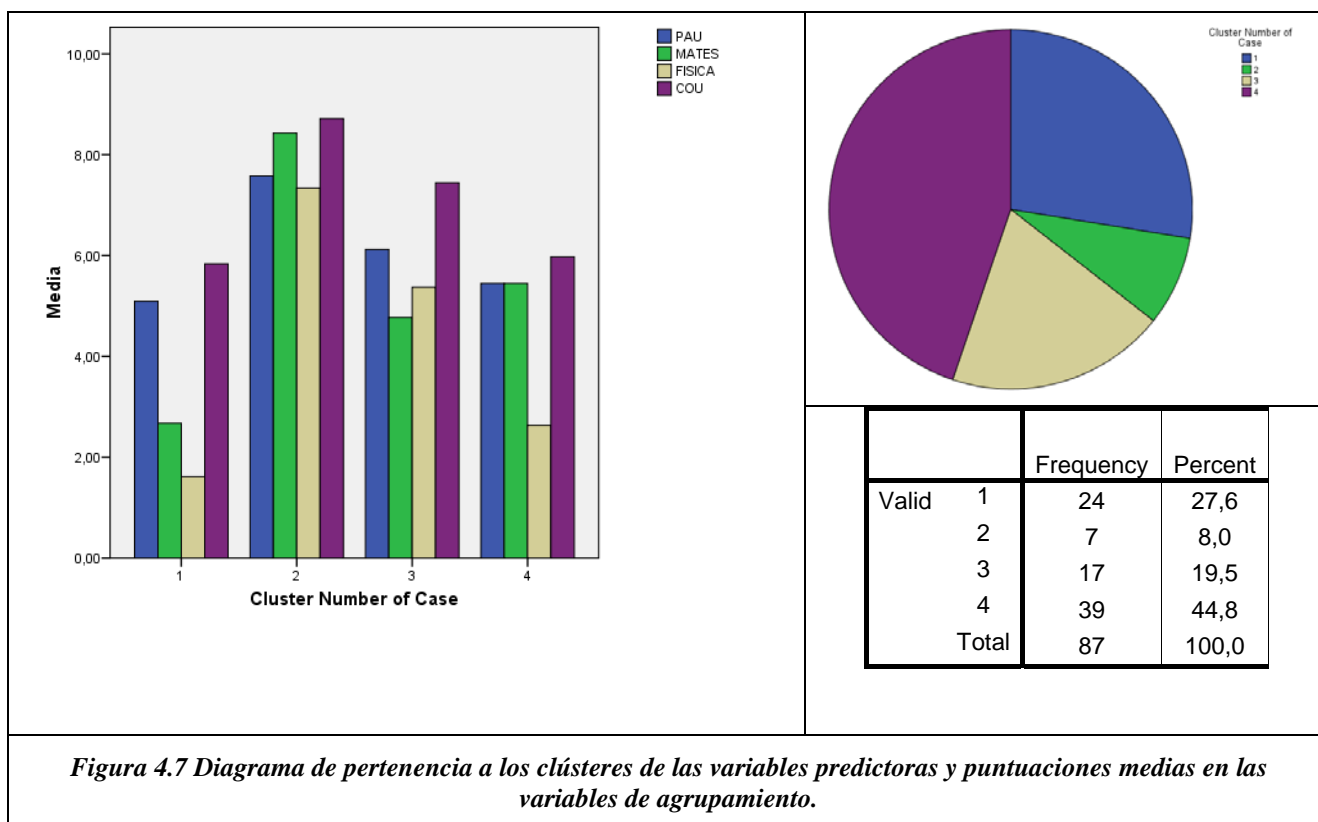
Por tanto, al igual que sucedía con los estudiantes de multimedia, podemos hablar de clústeres con puntuaciones significativamente diferentes para las variables de agrupación. La descripción de los mismos, se realiza a continuación:

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
PAU	Between Groups	38,827	3	12,942	22,636	,000
	Within Groups	47,455	83	,572		
	Total	86,282	86			
MATES	Between Groups	217,021	3	72,340	52,449	,000
	Within Groups	114,478	83	1,379		
	Total	331,499	86			
FISICA	Between Groups	272,743	3	90,914	63,976	,000
	Within Groups	117,949	83	1,421		
	Total	390,692	86			
COU	Between Groups	70,535	3	23,512	24,962	,000
	Within Groups	78,177	83	,942		
	Total	148,713	86			

La comparación de medias para los grupos obtenidos (ANOVA de un factor) resulta significativa para todas las variables predictoras.

En resumen el análisis de clústeres se plasma en la figura siguiente:



4.2.3 Descripción de los clústeres

A la vista del diagrama de barras de las variables de agrupación, observamos 4 tendencias diferentes por lo que se refiere a las mismas:

- Alumnos con puntuaciones bajas, con matemáticas y física muy por debajo del 5
 - En este caso alumnos que consiguieron aprobar COU con una nota alrededor del 6 y las PAU alrededor de la nota mínima. En la Salle no consiguieron aprobar ninguna de las 2 pruebas. Ambos controles se sitúan por debajo del 3. Este clúster es el de peor rendimiento.
- Alumnos con elevadas puntuaciones
 - Son alumnos que puntuaron por encima del 8 en COU y por encima del 7 en las PAU. En la Salle, observamos que tienen las puntuaciones más elevadas, estando en matemáticas por encima del 8 y en física rozando el 7. Son los alumnos de mayor rendimiento. Observamos que la nota de matemáticas es superior a la de física.
- Alumnos con puntuaciones medias (matemáticas < física)
 - En este caso son alumnos que en COU puntúan por encima del 7 y en las PAU por encima del 6. En la Salle, aprueban pero sin demasiado margen. Sin embargo, a diferencia de los alumnos de elevadas puntuaciones, en este caso la física es más alta que no las matemáticas. Este grupo corresponde a alumnos más pragmáticos.

- d) Alumnos con puntuaciones bajas, con matemáticas sobre el aprobado pero con física por debajo del 5.
- Observamos que en COU puntuaron sobre el 6 y en las PAU se sitúan en la franja de aprobado. Observamos que en la prueba de matemáticas del centro consiguieron aprobar situándose alrededor de la nota mínima, sin embargo, en física puntuaron por debajo del 3.

Cluster Number of Case

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	24	27,6	27,6	27,6
2	7	8,0	8,0	35,6
3	17	19,5	19,5	55,2
4	39	44,8	44,8	100,0
Total	87	100,0	100,0	

Cuantitativamente, es importante notar que el grueso de alumnos con una “buena predicción” supone un 27,5 % del total (8 + 19,5), y que los alumnos mayoritarios son los que tienen un buen expediente de COU y las PAU, pero que puntúan bajo en las pruebas de matemáticas y física de La Salle. En todo caso, esto no hace más que detectar deficiencias formativas de secundaria, como sucede con el estudio de Daniel Gil Pérez y Amparo Vilches¹:

“La enseñanza secundaria no proporciona una preparación adecuada, por lo que la ‘falta de base’ constituye la principal causa de fracaso de los estudiantes universitarios”

4.3 Clústeres resultantes en las variables de seguimiento

Las variables de seguimiento presentan irregularidades metodológicas debido a la inconstancia en la toma de medidas. En algunas pruebas se pierden notas importantes (variables que han definido clústeres en los parciales). Por ello deben tomarse a modo orientativo, y estos clústeres serán usados para obtener información complementaria y enriquecer las conclusiones finales.

Los clasificamos por pruebas, de manera que obtenemos las clasificaciones que siguen en los apartados siguientes. Las variables se han reordenado siguiendo los criterios anteriores de correlación para dar consistencia a las clasificaciones como conjunto.

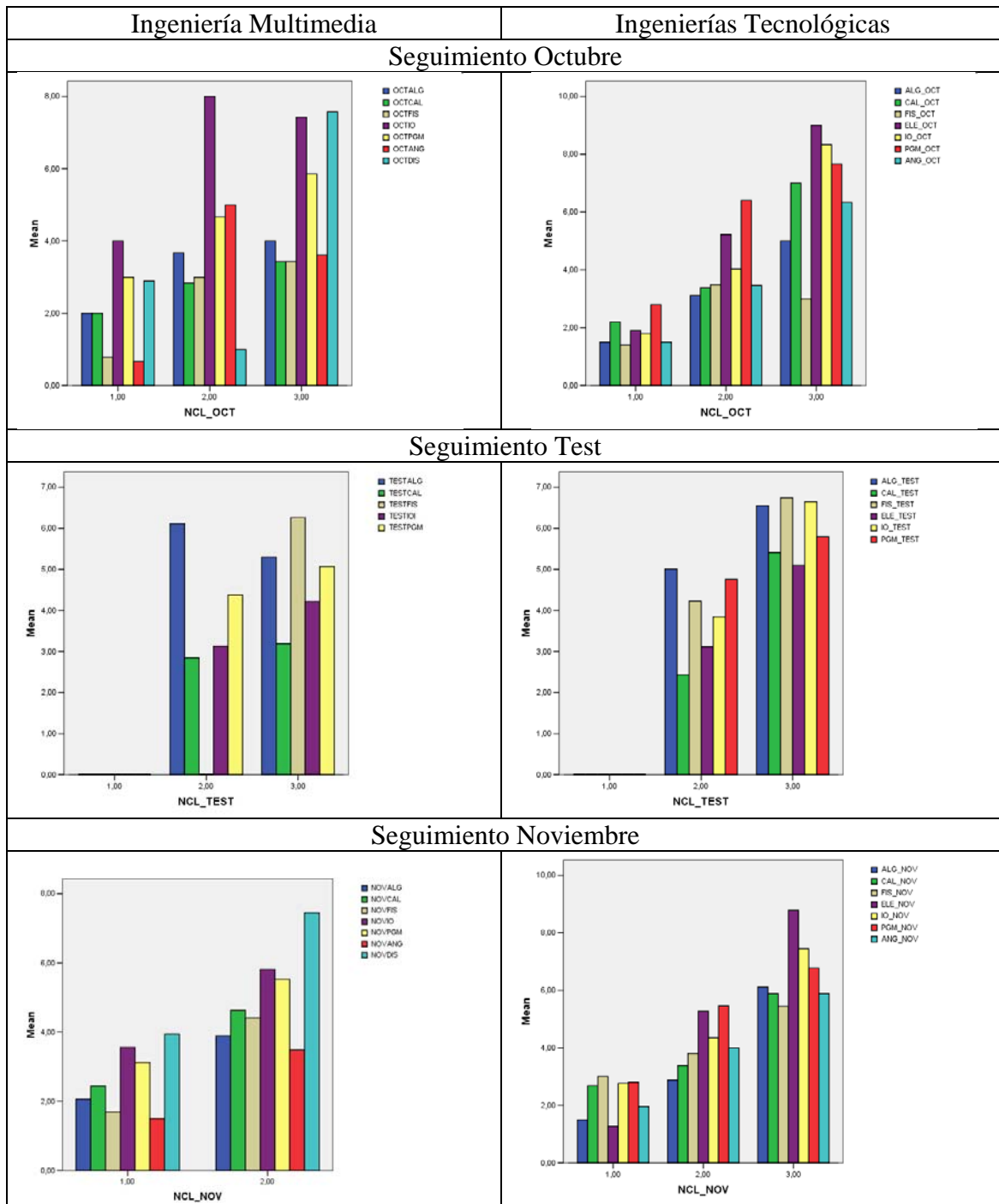
4.3.1 Aportación descriptiva de los clústeres de seguimiento

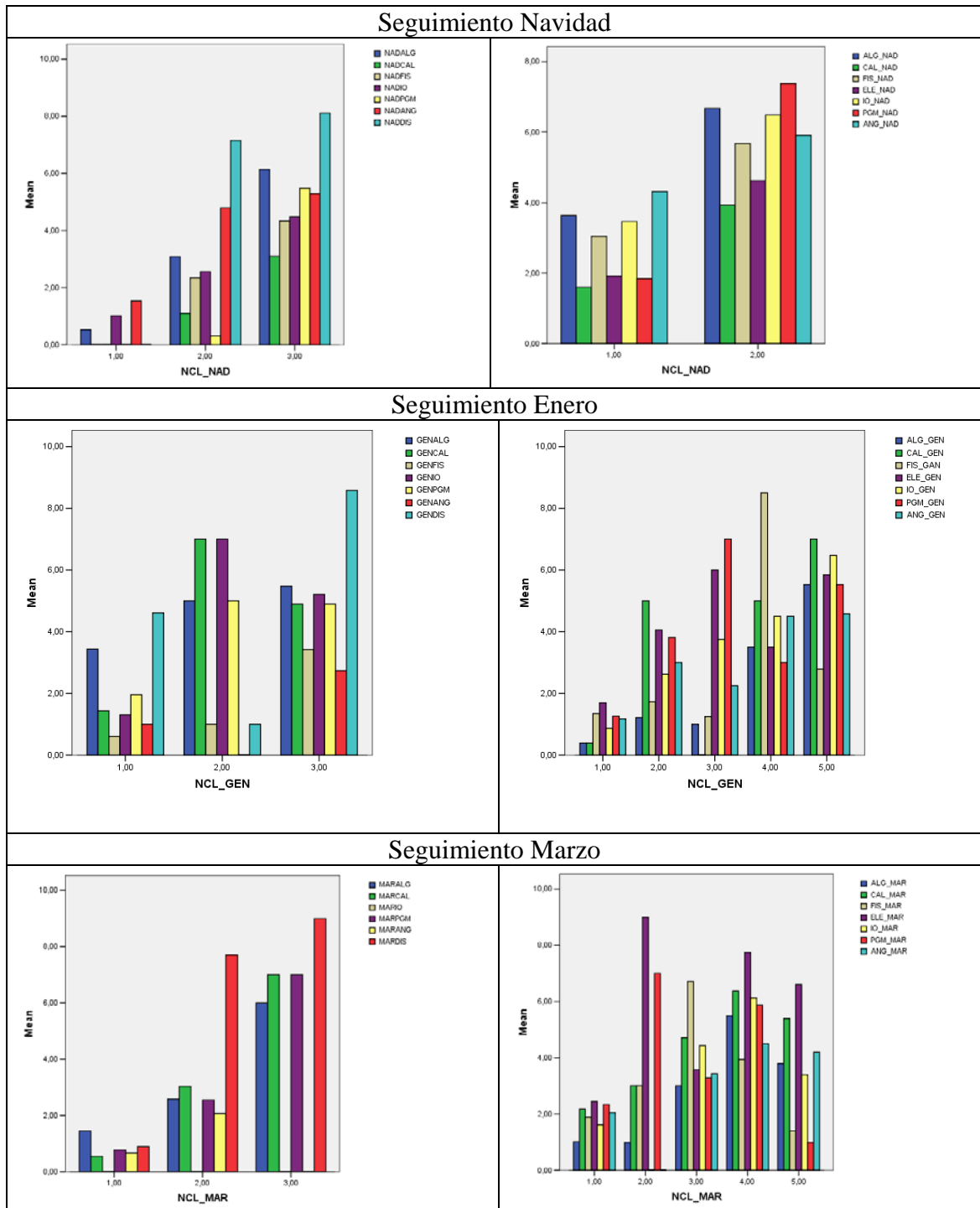
Es útil confeccionar una tabla comparativa de los grupos que se forman en multimedia en las otras ingenierías, porque esto no da idea de su diferente conducta académica. Tal

¹ EDWARDS, M. (2003). La atención a la situación del mundo en la educación científica. Presentada en el Departament de Didàctica de les Ciències de la Universitat de València. Dirigida por Daniel Gil y Amparo Vilches.

como se ha indicado, no entraremos a desarrollar el modelo por completo en estos clústeres, sino que realizaremos las pruebas para ver cómo aportan información orientadora para elaborar las conclusiones de la presente tesis.

A efectos descriptivos, por tanto, los clústeres de seguimiento nos ayudan a plasmar dos conductas académicas completamente diferentes, por lo que se refiere a la formación de grupos semejantes de rendimiento en las diferentes mediciones.





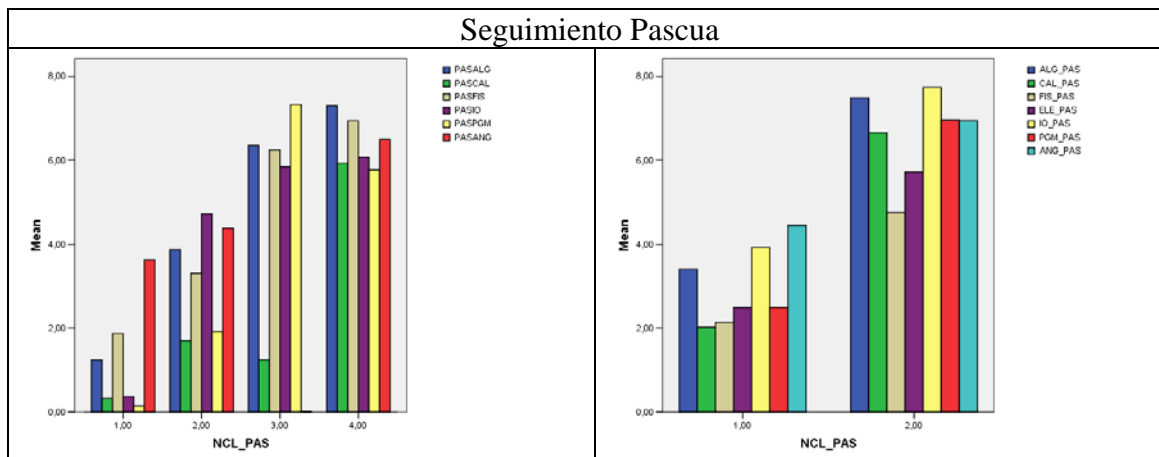


Figura 4.8 Perfiles de los clústeres de seguimiento.

4.3.2 Comentario metodológico a los clústeres de seguimiento

A pesar de la indudable aportación descriptiva, puede entenderse que los clústeres de seguimiento dibujan un panorama difícilmente tratable, por los siguientes motivos:

- No se puede construir un sistema consistente de clústeres porque no se dispone del mismo número de variables en cada prueba.
- Aún realizando un esfuerzo organizativo para recoger el mismo número de variables, resultarían sistemas consistentes no biunívocos. Como veremos en el capítulo 7, el tratamiento de los mismos puede ser muy problemático estadísticamente.
- Evidentemente, se podría realizar una investigación completa que entrara en estos clústeres, pero su extensión sería mucho mayor que la presente. En la presente tesis hemos optado, siendo la primera vez que se muestra esta metodología, aplicarla a dos sistemas de clústeres con notable dificultad (como veremos, el de multimedia es más tratable matemáticamente que el de tecnológicas).

Sin embargo, esta idea de clusterización del seguimiento nos induce a pensar en una posible aproximación al problema diferente a la de los sistemas consistentes de clústeres. Esta metodología los trata como entidades conceptuales (que define precisamente el sistema consistente), pero podemos preguntarnos qué pasaría si las tratásemos como entidades fijas.

Una muestra de los resultados que se obtienen con este enfoque (la evolución de los clústeres vistos como entidades fijas) la obtendremos a continuación usando las variables predictoras. No realizaremos el estudio completo para los clústeres de seguimiento debido a la extensión de esta investigación, pero sí adjuntaremos una muestra del camino alternativo con las notas de predicción como ejemplo.

4.4 Aproximación por evolución de clústeres predictores

Ésta es una técnica derivada de los sistemas consistentes de clústeres que no se centra en la evolución de los mismos con entidades a lo largo del tiempo (metáfora de las dunas a la que nos hemos referido en 2.1.2. Esta aproximación, en cambio, fija un número determinado de clústeres y estudia la evolución de su forma en el tiempo.

Tal como veremos en el capítulo 7, el método de los sistemas consistentes de clústeres tiene una parte heurística no exenta de incertidumbre. En los casos en que se llegue a un callejón sin salida, el enfoque de la evolución de los clústeres fijos puede ser un camino alternativo para obtener resultados en la investigación.

4.4.1 Una metáfora conceptual

Para hacer fácil la asimilación de este camino alternativo podemos pensar, conceptualmente, que la diferencia entre los enfoques es movimiento *versus* crecimiento.

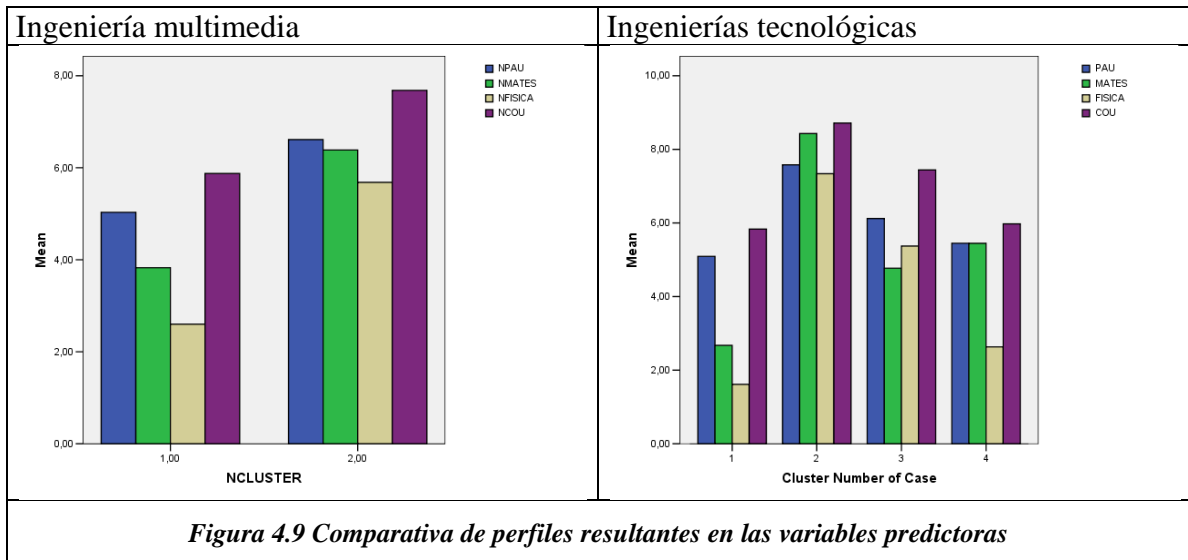
Con el vocablo *Movimiento* nos queremos referir a los sistemas consistentes de clústeres tal como los hemos introducido en el capítulo 2.1. Según este enfoque, los clústeres se miran como “objetos que se desplazan”, es decir, grupos de individuos a los que hemos dado “consistencia” a partir de unas características determinadas, pero que cambian su composición (los individuos que pertenecen a ellos) para dibujar nuevas formas estadísticas (metáfora de las dunas en el desierto).

En cambio, con el vocablo *Crecimiento* queremos indicar que el enfoque se centra en unas agrupaciones determinadas, y cual si fueran bonsáis, miramos qué formas toman a través del tiempo. Los estudios de crecimiento no son novedosos pues, de una u otra forma, se han tratado en diferentes investigaciones (si bien la aproximación matemática no es la misma que la que se realiza en esta tesis).

El estudio de la evolución de los clústeres como entidades fijas es, en definitiva, una visión “histórica” de los clústeres de predicción. Es decir, se trata de responder visualmente a la pregunta “¿Qué se hizo de los alumnos si nos atenemos a los grupos de predicción que se formaron inicialmente?” o, dicho de otro modo, “¿Qué futuro les esperaba a los grupos iniciales de rendimiento?”

4.4.2 Lectura evolutiva de las variables predictoras

Si agrupamos los resultados iniciales de multimedia y de ingenierías tecnológicas en una sola tabla, obtenemos:



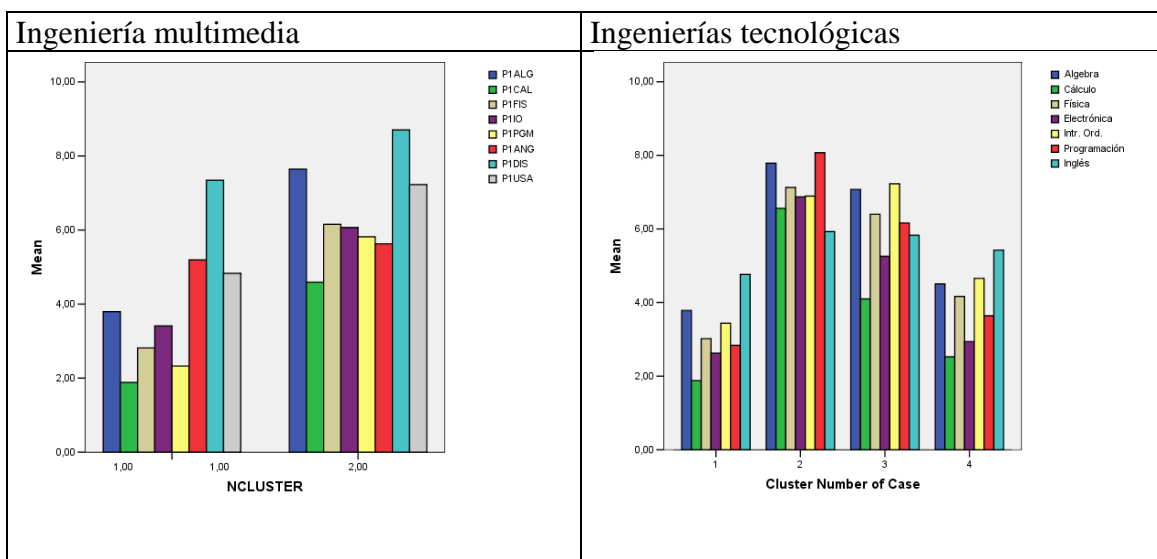
Para interpretar el procedimiento que desarrollaremos a continuación hay que recordar cuáles eran los clústeres iniciales de predicción. Para ingeniería multimedia (izquierda), teníamos:

- a) Clúster número 1: Suspendidos Salle, regulares en COU
- b) Clúster número 2: Aprobados Salle, altos en COU

Y para las ingenierías tecnológicas (derecha), resultaba:

- a) Alumnos con puntuaciones bajas, con matemáticas y física muy por debajo del 5
- b) Alumnos con elevadas puntuaciones
- c) Alumnos con puntuaciones medias (matemáticas < física)
- d) Alumnos con puntuaciones bajas, con matemáticas sobre el aprobado pero con física por debajo del 5

Con las consideraciones precedentes, la gráfica de evolución resultante es la siguiente:



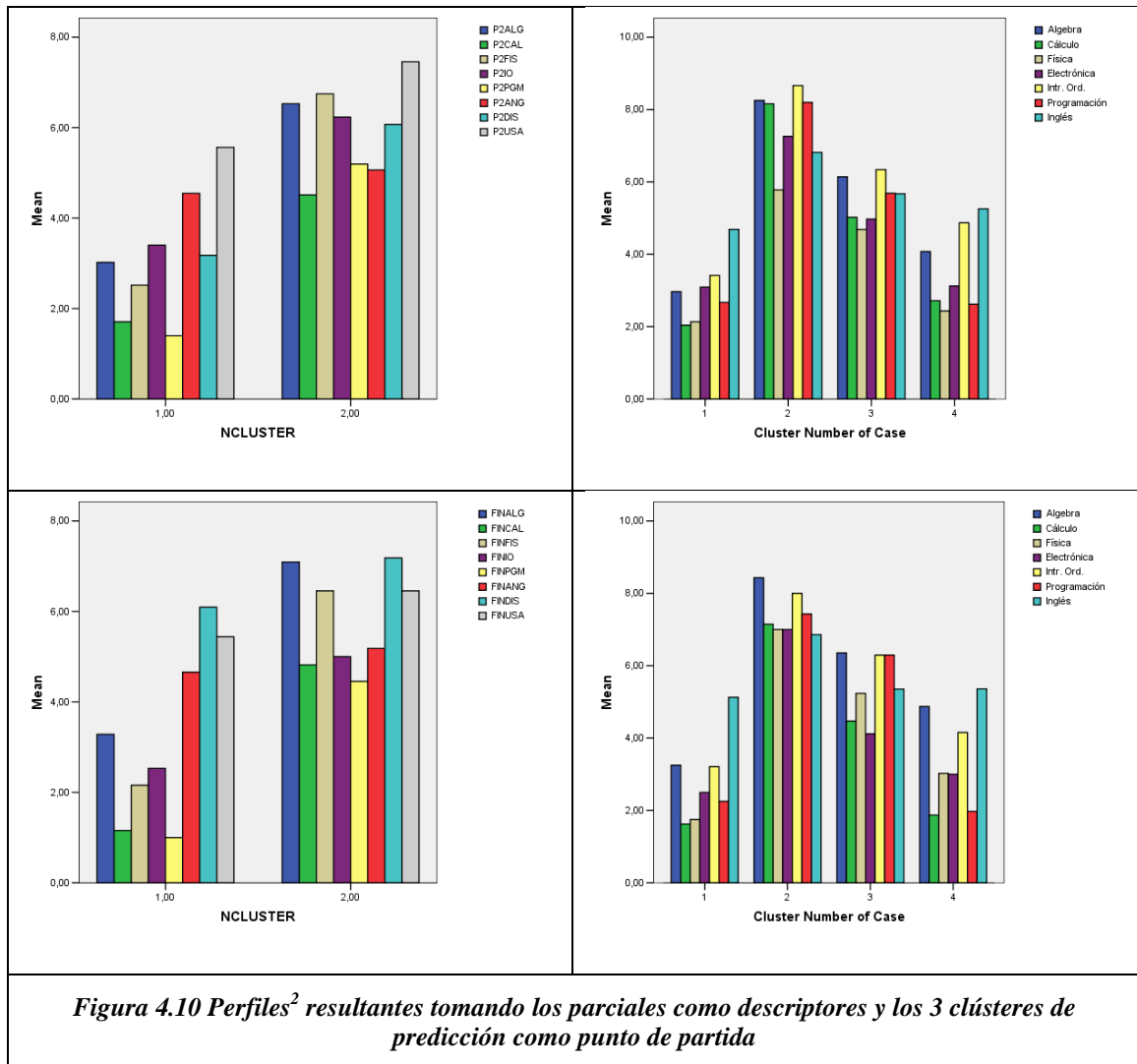


Figura 4.10 Perfiles² resultantes tomando los parciales como descriptores y los 3 clústeres de predicción como punto de partida

Lectura para ingeniería multimedia

Observamos que en ingeniería multimedia no se producen variaciones significativas en los grupos. Estos siguen conservando sus puntuaciones a lo largo del curso. Los dos clústeres se mantienen dando peso a las diferencias que existen en ambos grupos.

Específicamente en el grupo de rendimiento medio, todas las asignaturas tienen una evolución casi nula a excepción de Diseño que en el segundo parcial se ve afectado considerablemente y su media baja.

Lectura para ingenierías tecnológicas

En este caso, conforme avanza el curso vemos como las diferencias entre los clústeres se acentúan. El grupo de alto rendimiento mejora y los de bajo rendimiento empeoran.

El clúster de rendimiento bajo, que estaba por debajo del cinco tanto en matemáticas como en física y que habían aprobado COU y las PAU con una nota suficiente, sigue

² Para la mayor claridad del estudio hemos diferenciado las asignaturas de las tecnológicas de las de multimedia, dándoles un nombre diferente.

puntuando muy por debajo del 5, excepto en Inglés que la media se sitúa alrededor del aprobado.

El clúster de rendimiento medio-bajo, que aprobaban matemáticas pero suspendían física y su nota de COU y PAU se situaba sobre el aprobado, en el primer parcial superan el 4 en álgebra, física, introducción a los ordenadores e inglés, que son más de la mitad de las asignaturas. En el segundo parcial, de las anteriores solo se mantienen por encima del 4, álgebra, introducción a los ordenadores e inglés. Y, ya en el tercer parcial, física no se recupera, pero baja hasta rozar el cuatro, y álgebra sube hasta alcanzar una puntuación muy cercana al 5. Al final aprueban, alegra e inglés.

En el clúster de rendimiento alto, excepto en el primer parcial de inglés y el segundo parcial de física, todas las demás puntuaciones se sitúan por encima del 6. El mejor parcial es el segundo ya que álgebra, cálculo, introducción a los ordenadores y programación se sitúan alrededor, incluso superando, el 8.

El grupo de los alumnos de puntuaciones medias, se sitúa mayormente en el aprobado a pesar de que asignaturas como cálculo del primer parcial y tercer parcial, y electrónica del tercer parcial pueden llegar a bajar levemente del cinco.

Resultados estadísticos para consideraciones “en horizontal”

Con el método que estamos utilizando, para realizar las comparaciones entre clústeres o lecturas en horizontal, debemos tener en cuenta la significación de las diferentes variables.

Resultados del ANOVA para ingeniería multimedia

ANOVA PRIMER PARCIAL		Sum of Squares	df	Mean Square	F	Sig.
PIALG	Between Groups	113,480	2	56,740	9,741	,000
	Within Groups	232,986	40	5,825		
	Total	346,466	42			
PICAL	Between Groups	62,210	2	31,105	14,280	,000
	Within Groups	87,131	40	2,178		
	Total	149,341	42			
PIFIS	Between Groups	83,426	2	41,713	10,571	,000
	Within Groups	157,833	40	3,946		
	Total	241,259	42			
PIIO	Between Groups	67,007	2	33,504	8,038	,001
	Within Groups	166,723	40	4,168		
	Total	233,730	42			
PIPGM	Between Groups	151,267	2	75,633	9,676	,000
	Within Groups	312,664	40	7,817		
	Total	463,931	42			
PIANG	Between Groups	2,635	2	1,318	,220	,803
	Within Groups	239,284	40	5,982		
	Total	241,919	42			
PIDIS	Between Groups	14,425	2	7,212	1,220	,306
	Within Groups	236,555	40	5,914		
	Total	250,980	42			

4. Aplicaciones básicas de los sistemas consistentes de Clústeres. Predicción y seguimiento

PIUSA	Between Groups	41,683	2	20,841	3,291	,048
	Within Groups	253,343	40	6,334		
	Total	295,026	42			

ANOVA SEG. PARCIAL		Sum of Squares	df	Mean Square	F	Sig.
P2ALG	Between Groups	133,605	2	66,803	13,777	,000
	Within Groups	193,959	40	4,849		
	Total	327,564	42			
P2CAL	Between Groups	102,939	2	51,470	8,791	,001
	Within Groups	234,186	40	5,855		
	Total	337,125	42			
P2FIS	Between Groups	114,087	2	57,043	17,985	,000
	Within Groups	126,872	40	3,172		
	Total	240,959	42			
P2IO	Between Groups	81,183	2	40,592	5,916	,006
	Within Groups	274,461	40	6,862		
	Total	355,644	42			
P2PGM	Between Groups	146,370	2	73,185	15,693	,000
	Within Groups	186,547	40	4,664		
	Total	332,917	42			
P2ANG	Between Groups	4,411	2	2,206	,367	,695
	Within Groups	240,100	40	6,003		
	Total	244,511	42			
P2DIS	Between Groups	112,705	2	56,353	3,521	,039
	Within Groups	640,130	40	16,003		
	Total	752,835	42			
P2USA	Between Groups	30,746	2	15,373	4,516	,017
	Within Groups	136,161	40	3,404		
	Total	166,907	42			

ANOVA TERCER PARCIAL		Sum of Squares	df	Mean Square	F	Sig.
FINALG	Between Groups	140,453	2	70,226	15,457	,000
	Within Groups	181,733	40	4,543		
	Total	322,186	42			
FINCAL	Between Groups	123,652	2	61,826	19,324	,000
	Within Groups	127,976	40	3,199		
	Total	251,628	42			
FINFIS	Between Groups	150,515	2	75,257	16,390	,000
	Within Groups	183,671	40	4,592		
	Total	334,186	42			
FINIO	Between Groups	87,284	2	43,642	10,000	,000
	Within Groups	174,576	40	4,364		
	Total	261,860	42			
FINPGM	Between Groups	221,947	2	110,974	28,736	,000
	Within Groups	154,471	40	3,862		
	Total	376,419	42			
FINANG	Between Groups	5,397	2	2,699	,416	,663
	Within Groups	259,719	40	6,493		
	Total	265,116	42			

FINDIS	Between Groups	22,575	2	11,288	1,518	,232
	Within Groups	297,471	40	7,437		
	Total	320,047	42			
FINUSA	Between Groups	11,827	2	5,913	1,291	,286
	Within Groups	183,243	40	4,581		
	Total	195,070	42			

Podemos observar como tenemos una significación sustancial en casi todas las asignaturas excepto en inglés, diseño y usabilidad, que no resultan significativas. Esto es debido a que son asignaturas más fáciles que hacen que casi todos los alumnos aprueben. Por tanto nos fijaremos sobretodo en las demás asignaturas. Pero vemos que las diferencias están claramente definidas.

Capítulo 5

Construcción de los sistemas consistentes de clústeres en ingenierías multimedia y tecnológicas

5.1 Tratamiento de las diferentes variables implicadas en el estudio

Se han realizado pruebas de clústeres para los tres tipos de variables objeto de estudio: predictoras, de evaluación y de seguimiento. Hemos de tener en cuenta que los alumnos de multimedia tienen una asignatura más que los de las ingenierías tecnológicas (8 en vez de 7 asignaturas).

Hay que señalar que, en el caso de las ingenierías tecnológicas, al tener un volumen mayor de alumnos que en ingeniería multimedia y, por tanto, aparecer una casuística mucho más amplia de casos particulares de estudiantes, pueden aparecer algunos casos de clústeres espúreos.

Tal como se ha señalado en el marco teórico (capítulo 2) compete al investigador distinguir estos casos individuales (que normalmente forman grupos por sí solos debido a su diferencia con todos los demás grupos) de los casos normales que serán incluidos en el estudio de clústeres. Con todo, se consiguió finalmente eliminar a un solo alumno de la muestra, del cual se comprobó que, efectivamente, sus circunstancias particulares le convertían en una entidad académica de excepción. Por lo que se refiere a los demás casos de esta naturaleza, se consiguió incluirlos en el estudio de forma coherente dentro de diferentes clústeres, por lo que la mortalidad estadística de la investigación fue realmente baja.

El tratamiento de las variables predictoras es extremadamente complejo y, por añadidura, al tratarse de una investigación *expostfacto*, se padece la ausencia de recogida de datos en muchas de las pruebas realizadas. Es decir, si bien se dispone de la mayoría de variables estudiadas, es muy común que falten una o dos en diversas mediciones. Ello es debido a que dicho estudio nunca fue planificado de manera formal y que, como sucede en las investigaciones *expostfacto*, sufrió las alteraciones circunstanciales del quehacer cotidiano de los profesores. La carga de trabajo en muchos casos, fue la causa de que no se tomaran en determinados momentos las mediciones de interés.

Así las cosas, en el capítulo siguiente se realizará un estudio de las variables de seguimiento bajo el prisma de los sistemas consistentes de clústeres pero sin aplicar a fondo la metodología. En el capítulo presente, en cambio, se realizará la construcción del sistema para las variables de evaluación, siguiendo las directrices establecidas en el marco teórico.

El siguiente esquema plasma las pruebas que se realizarán en esta tesis y que constituyen el núcleo del tratamiento de datos sobre el cual serán desarrolladas las conclusiones:

- a) Gráficas de “rompimiento individual” de clústeres.
- b) Tablas de contingencia de la evolución entre los clústeres.
- c) Gráficas de evolución entre los clústeres que ilustrarán las tablas anteriores.
- d) Pruebas de intensidad de la evolución entre los clústeres.
- e) Pruebas estadísticas de poder explicativo del primer parcial sobre el tercero.
- f) Establecer los componentes de las cadenas de Markov.

- g) Calcular los estados esperados según el modelo de estas cadenas y medir su discrepancia con los estados reales.
- h) Realizar las pruebas estadísticas para evaluar la aportación académica del centro.

5.2 Detección de los clústeres de rendimiento en ingeniería multimedia

Los dendogramas de los clústeres resultantes de los 3 parciales para Ingeniería Multimedia son:

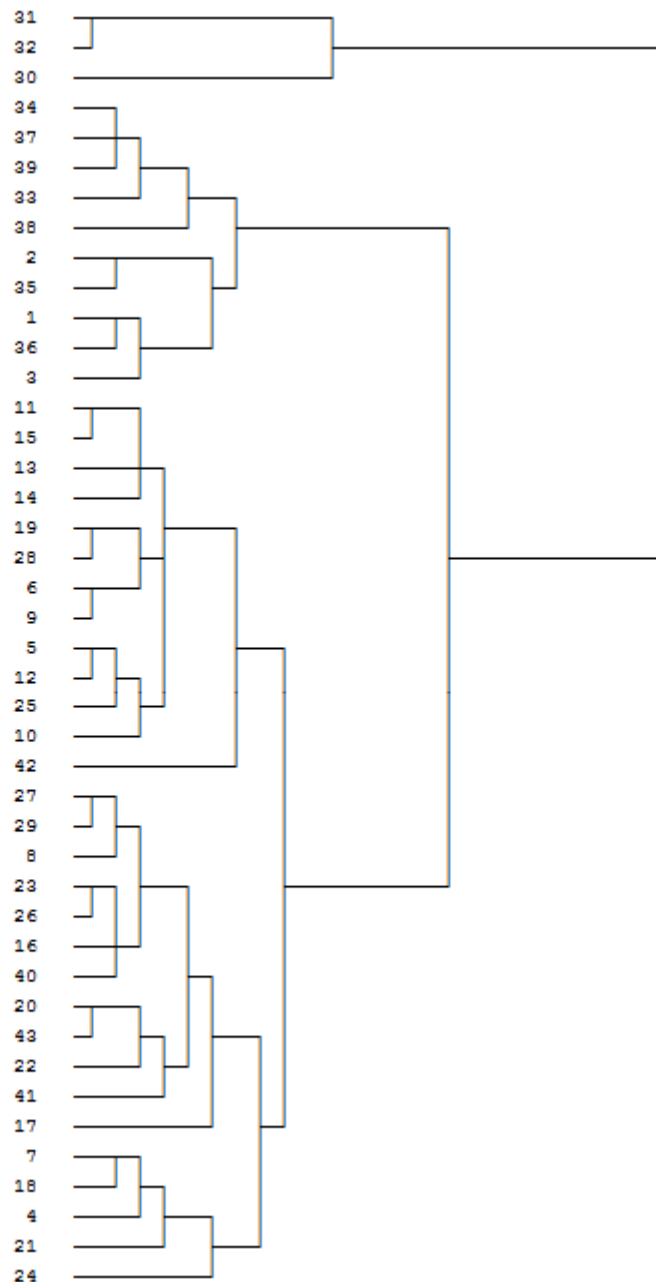


Figura 5.1 Dendrograma de clústeres de las variables de evaluación para el primer parcial y puntuaciones medias en las variables de agrupamiento de Multimedia.

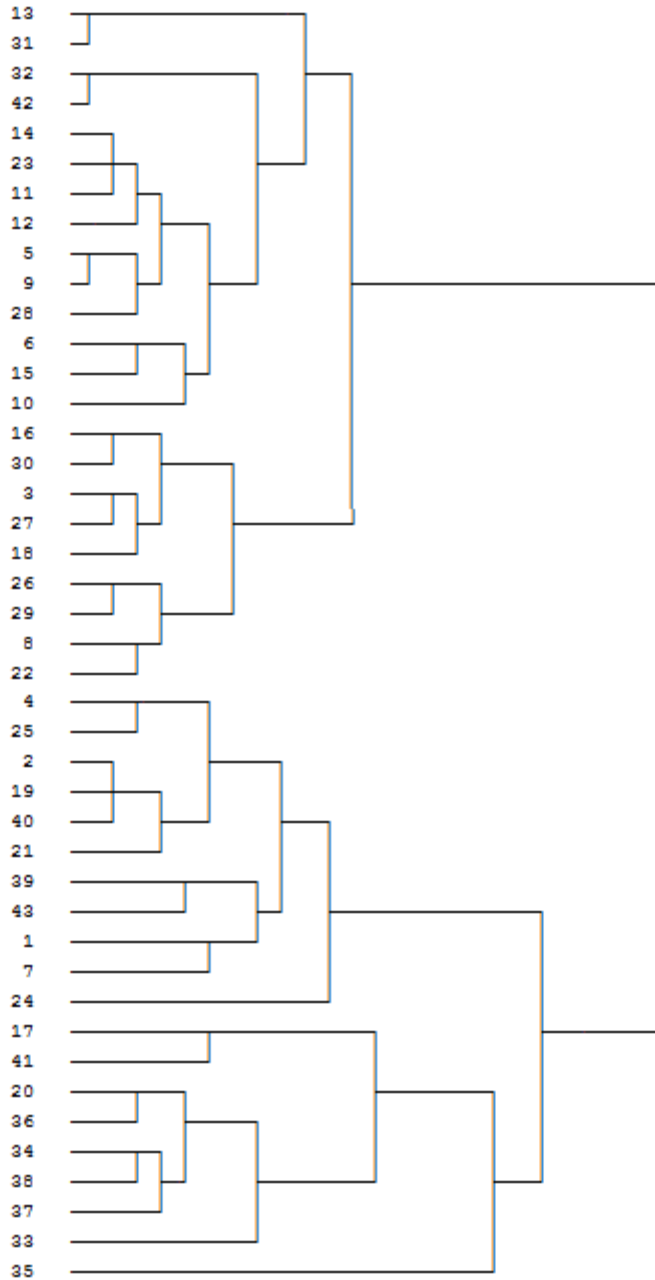


Figura 5.2 Dendrograma de clústeres de las variables de evaluación para el segundo parcial y puntuaciones medias en las variables de agrupamiento de Multimedia.

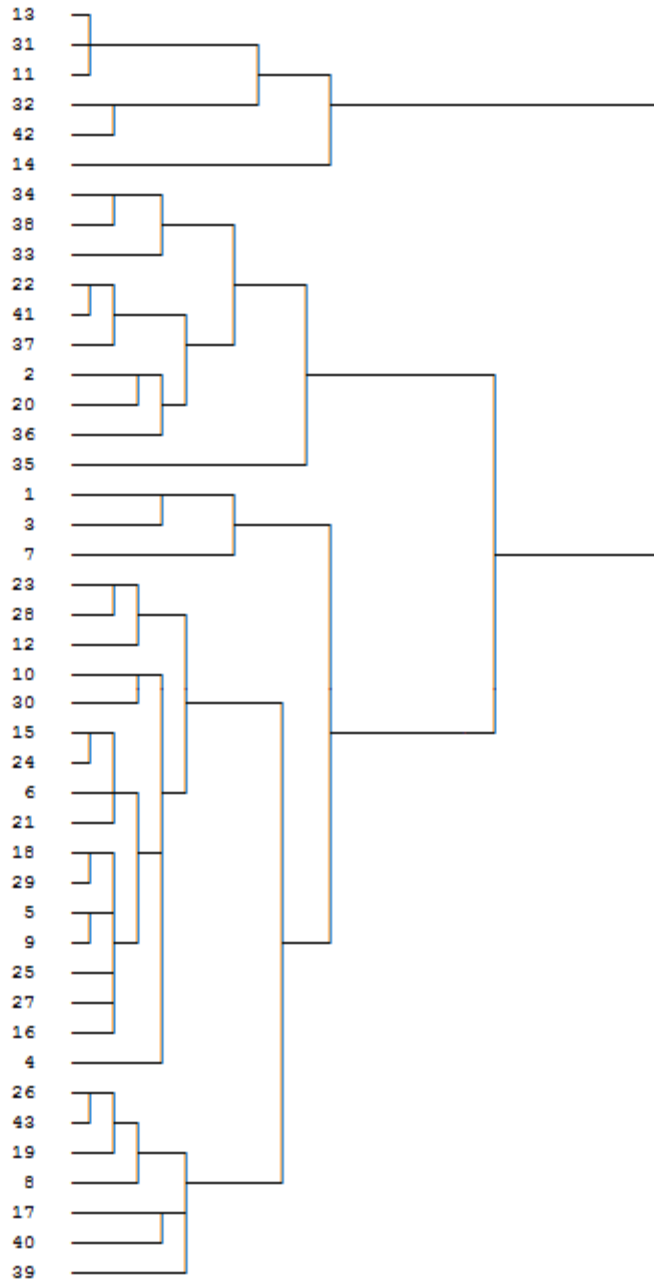


Figura 5.3 Dendrograma de clústeres de las variables de evaluación para el tercer parcial y puntuaciones medias en las variables de agrupamiento de Multimedia.

Se trata de las 3 pruebas parciales que se han efectuado sobre los alumnos de Multimedia. Las figuras que siguen resumen visualmente los resultados:

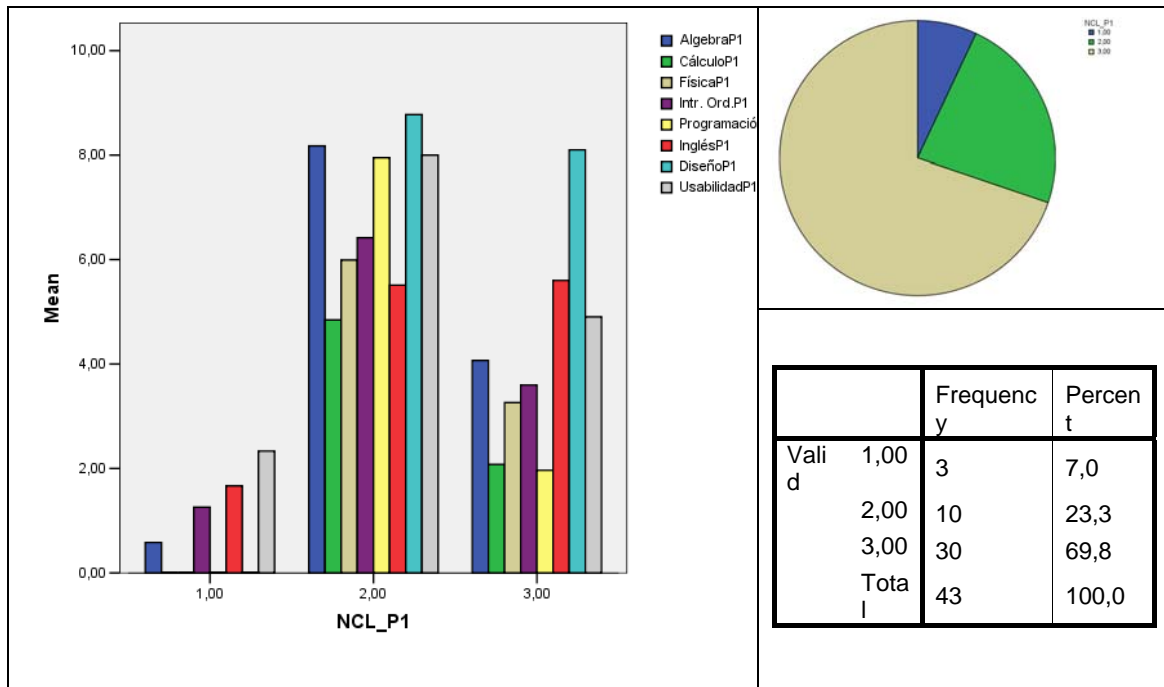


Figura 5.4 Diagrama de pertenencia a los clústeres de las variables de evaluación para el primer parcial y puntuaciones medias en las variables de agrupamiento.

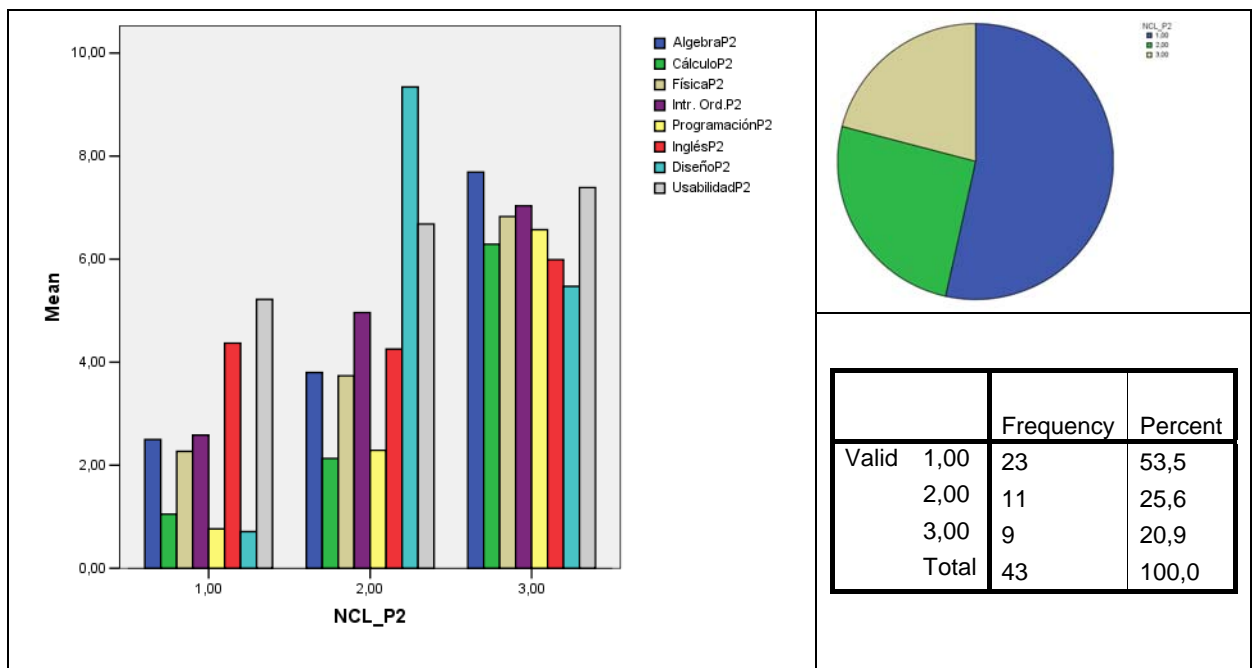


Figura 5.5 Diagrama de pertenencia a los clústeres de las variables de evaluación para el segundo parcial y puntuaciones medias en las variables de agrupamiento.

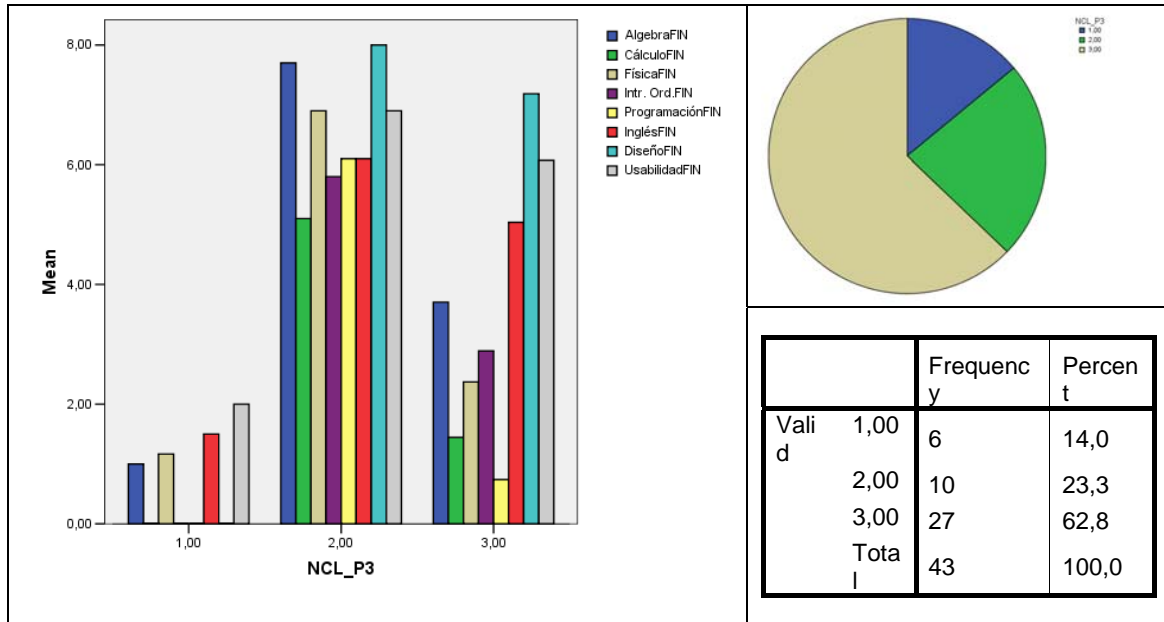


Figura 5.6 Diagrama de pertenencia a los clústeres de las variables de evaluación para el tercer parcial y puntuaciones medias en las variables de agrupamiento.

Las pruebas de significación (ANOVA) para cada una de estas clasificaciones son las siguientes:

ANOVA P1

		Sum of Squares	df	Mean Square	F	Sig.
AlgebraP1	Between Groups	183,290	2	91,645	22,465	,000
	Within Groups	163,175	40	4,079		
	Total	346,466	42			
CálculoP1	Between Groups	78,841	2	39,421	22,366	,000
	Within Groups	70,500	40	1,762		
	Total	149,341	42			
FísicaP1	Between Groups	99,271	2	49,635	13,983	,000
	Within Groups	141,988	40	3,550		
	Total	241,259	42			
Intr. Ord.P1	Between Groups	85,504	2	42,752	11,537	,000
	Within Groups	148,226	40	3,706		
	Total	233,730	42			
ProgramaciónP1	Between Groups	302,085	2	151,042	37,330	,000
	Within Groups	161,846	40	4,046		
	Total	463,931	42			
InglésP1	Between Groups	42,743	2	21,372	4,292	,020
	Within Groups	199,176	40	4,979		
	Total	241,919	42			
DiseñoP1	Between Groups	194,223	2	97,112	68,441	,000
	Within Groups	56,756	40	1,419		
	Total	250,980	42			
UsabilidadP1	Between Groups	103,130	2	51,565	10,748	,000
	Within Groups	191,896	40	4,797		
	Total	295,026	42			

Figura 5.7 ANOVA de los clústeres de las variables de rendimiento para el primer parcial para Multimedia.

ANOVA P2

		Sum of Squares	df	Mean Square	F	Sig.
AlgebraP2	Between Groups	174,518	2	87,259	22,806	,000
	Within Groups	153,046	40	3,826		
	Total	327,564	42			
CálculoP2	Between Groups	178,769	2	89,385	22,578	,000
	Within Groups	158,356	40	3,959		
	Total	337,125	42			
FísicaP2	Between Groups	134,661	2	67,330	25,336	,000
	Within Groups	106,298	40	2,657		
	Total	240,959	42			
Intr. Ord.P2	Between Groups	138,346	2	69,173	12,733	,000
	Within Groups	217,298	40	5,432		
	Total	355,644	42			
ProgramaciónP2	Between Groups	218,297	2	109,148	38,091	,000
	Within Groups	114,620	40	2,865		
	Total	332,917	42			
InglésP2	Between Groups	19,626	2	9,813	1,745	,188
	Within Groups	224,885	40	5,622		
	Total	244,511	42			
DiseñoP2	Between Groups	581,638	2	290,819	67,950	,000
	Within Groups	171,197	40	4,280		
	Total	752,835	42			
UsabilidadP2	Between Groups	36,469	2	18,234	5,592	,007
	Within Groups	130,438	40	3,261		
	Total	166,907	42			

Figura 5.8 ANOVA de los clústeres de las variables de rendimiento para el segundo parcial para Multimedia.

ANOVA P3

		Sum of Squares	df	Mean Square	F	Sig.
AlgebraFIN	Between Groups	190,456	2	95,228	28,916	,000
	Within Groups	131,730	40	3,293		
	Total	322,186	42			
CálculoFIN	Between Groups	128,061	2	64,031	20,727	,000
	Within Groups	123,567	40	3,089		
	Total	251,628	42			
FísicaFIN	Between Groups	180,156	2	90,078	23,392	,000
	Within Groups	154,030	40	3,851		
	Total	334,186	42			
Intr. Ord.FIN	Between Groups	131,594	2	65,797	20,204	,000
	Within Groups	130,267	40	3,257		
	Total	261,860	42			
ProgramaciónFIN	Between Groups	234,333	2	117,167	32,985	,000
	Within Groups	142,085	40	3,552		
	Total	376,419	42			
InglésFIN	Between Groups	83,753	2	41,877	9,236	,001
	Within Groups	181,363	40	4,534		
	Total	265,116	42			
DiseñoFIN	Between Groups	287,972	2	143,986	179,567	,000
	Within Groups	32,074	40	,802		
	Total	320,047	42			
UsabilidadFIN	Between Groups	100,318	2	50,159	21,175	,000
	Within Groups	94,752	40	2,369		
	Total	195,070	42			

Figura 5.9 ANOVA de los clústeres de las variables de rendimiento para el tercer parcial para Multimedia.

La prueba de análisis de la varianza resultó significativa para todas las variables en los 3 parciales, por lo que los clústeres obtenidos son entidades validadas por el tratamiento estadístico.

5.2.1 Descripción de los clústeres

5.2.1.1 Para el primer parcial

A la vista del diagrama de barras de la figura 5.4 observamos tres grupos claros de clasificación:

- a) Alumnos de rendimiento notable (clúster 2):

Llama la atención que las notas más altas sean la de álgebra y la de diseño, seguidas de cerca por la de usabilidad y la de programación. Esto nos hace pensar en un alumno que se siente a gusto en la titulación y nos recuerda un poco al perfil inicial de “matemáticos, pero no físicos”, en el sentido de alumnos cautivados por la belleza formal.

- b) Alumnos de rendimiento pobre pero que salvan las notas propias de la titulación (clúster 3):

No es una sorpresa que alumnos a los que les cuesta aprobar aseguren el criterio de permanencia (aprobar al menos tres asignaturas) y se dediquen al diseño, la usabilidad o el inglés, por ser más asequibles. Llama la atención, sin embargo, que la nota de álgebra sea un poco más elevada en comparación con las demás.

- c) Alumnos de rendimiento deficiente en todas las asignaturas y de abandono (clúster 1):

Para este grupo podría decirse que se tiene un “perfil de intentos” de los alumnos a los que les cuesta más seguir las asignaturas, dado que aparecen 4 asignaturas que, en orden creciente de medias, son: Álgebra, Introducción a los ordenadores, Inglés y Diseño. Al resto, no se presentaron.

5.2.1.2 Para el segundo parcial

Se producen cambios llamativos sobre los grupos resultantes, en comparación con los del parcial anterior. Siguen siendo 3 grupos, pero ahora los perfiles han variado:

- a) Alumnos de rendimiento notable (clúster 3):

Las notas de estos alumnos son ahora más altas en su parte tecnológica que en su parte propia del multimedia. Es decir, quitado de la usabilidad, inglés y diseño son las notas más bajas del grupo. Nos sugiere un alumno que se atreve con las materias difíciles de la carrera.

- b) Alumnos de rendimiento pobre (clúster 2):

Éstos obtienen un perfil muy similar al del primer parcial, casi idéntico.

c) Alumnos de rendimiento deficiente (clúster 1):

Se nota que hacen un esfuerzo por aprobar y no perder la conexión con la carrera. Esto hay que atribuirlo a la labor de tutoría, porque es llamativo que no se hayan descolgado después el primer parcial. Apuestan por inglés y diseño, con una clara voluntad de permanencia en el centro.

5.2.1.3 Para el tercer parcial

En el tercer parcial los tres grupos se definen con un carácter más marcado, resultando:

a) Alumnos de rendimiento notable (clúster 2):

Llama la atención que han ido mejorando en física respecto sobretodo a introducción a los ordenadores y programación.

b) Alumnos de rendimiento pobre (clúster 3):

Han asegurado aprobar las asequibles para poder, como mínimo, repetir curso y aprovechar el examen de septiembre para conseguir superar alguna asignatura más.

c) Alumnos de rendimiento deficiente (clúster 1):

Finalmente han abandonado muchas asignaturas para apostar por una segura. Sin embargo, los resultados no son buenos, aunque el grupo se ha reducido mucho.

5.2.2 Reordenación de los clústeres de rendimiento.

Dado que nuestra metodología busca crear un sistema de clústeres, donde podamos establecer correspondencias conceptuales y estadísticas entre ellos (es decir, entre los clústeres obtenidos en diferentes pruebas), se impone reordenar los clústeres obtenidos a efectos de no perder la coherencia.

Importante: hay que tener en cuenta que la asignación de un número de clúster por parte del paquete estadístico SPSS se hace a la hora de ejecutar el algoritmo de agrupamiento y obedece a criterios de rápida formación de los grupos, pero no de puntuaciones altas o bajas.

Esta ordenación puede ser difícil de detectar en ciertos casos pero en el presente está muy clara. Visualmente y numéricamente apreciamos en cada prueba de clústeres realizada tres grupos de rendimiento. Nos favorecerá la construcción del sistema, además, que en las variables predictoras hayan aparecido también tres grupos similares.

Por tanto, siguiendo el criterio de numerar de mejor a peor los clústeres (tal como hemos establecido en las variables predictoras) la tabla de reordenación es la siguiente:

Primer parcial			Segundo parcial			Tercer parcial		
Inicial		Final	Inicial		Final	Inicial		Final
1	Bajo	3	1	Bajo	3	1	Bajo	3
2	Alto	1	2	Medio	2	2	Alto	1
3	Medio	2	3	Alto	1	3	Medio	2

Con esta nueva recodificación, los gráficos resultantes son:

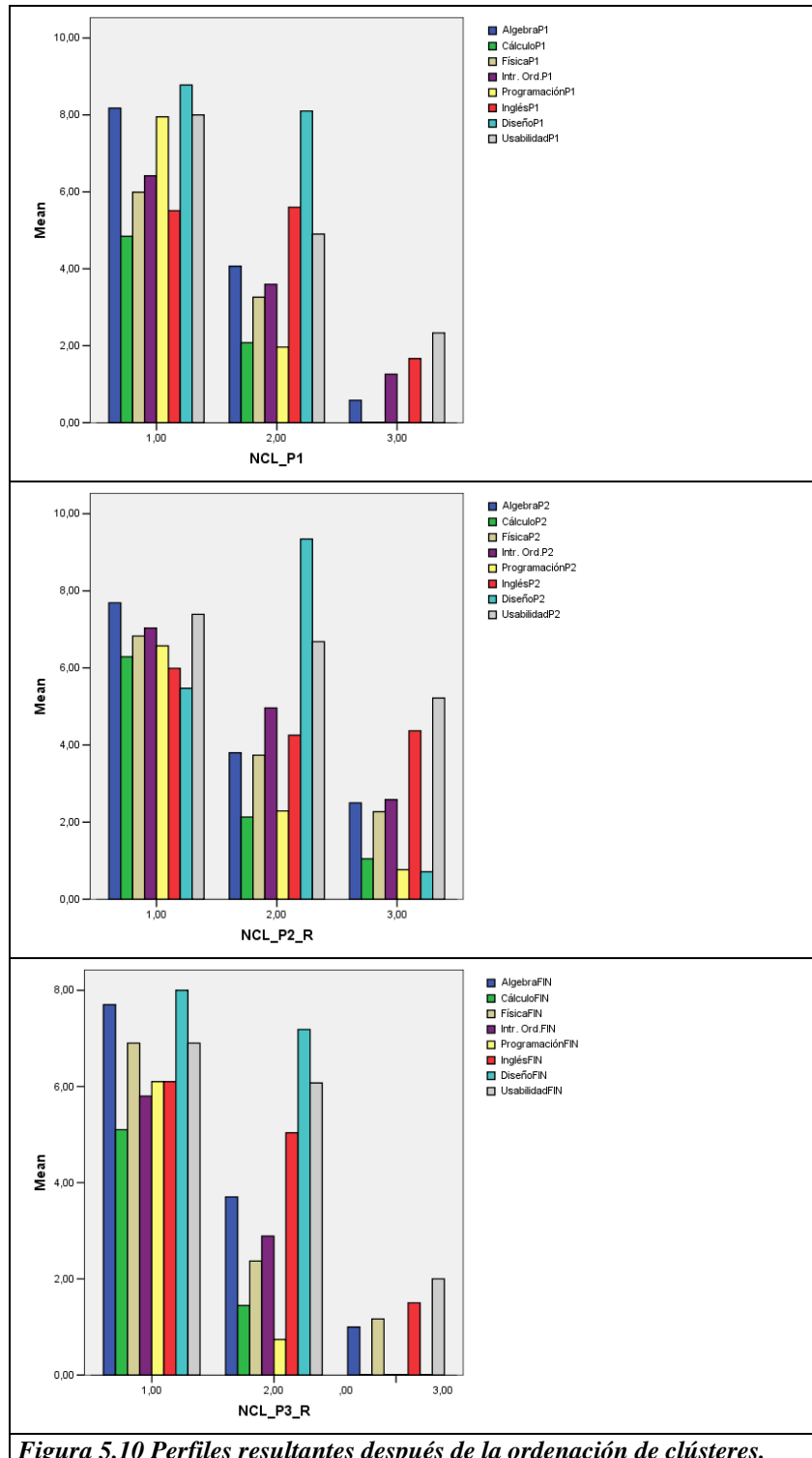


Figura 5.10 Perfiles resultantes después de la ordenación de clústeres.

5.3 Detección de los clústeres de rendimiento en ingenierías tecnológicas

Los dendogramas de los clústeres resultantes para los 3 parciales de ingenierías tecnológicas son:

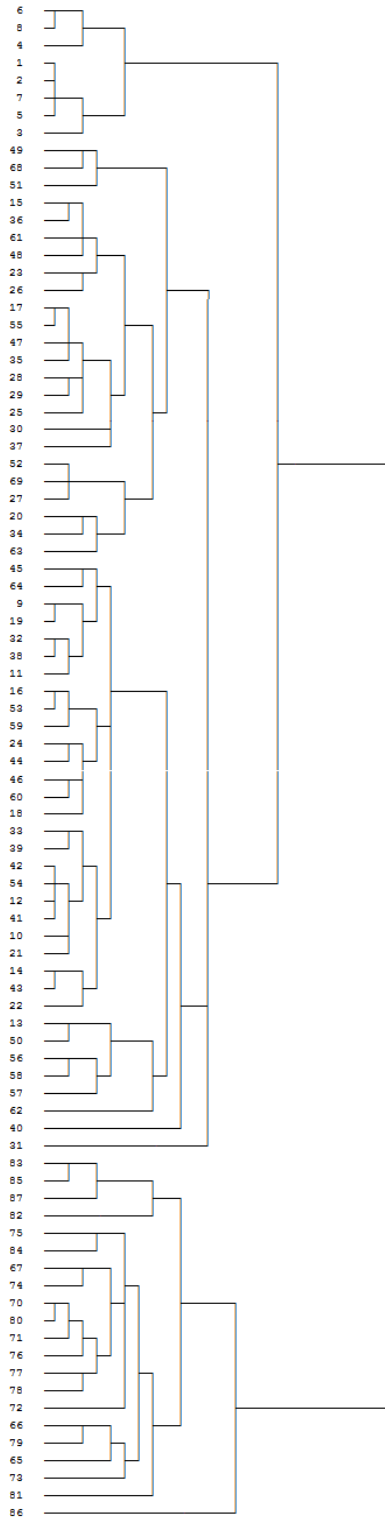


Figura 5.11 Dendograma de clústeres de las variables de evaluación para el primer parcial y puntuaciones medias en las variables de agrupamiento de las Ingenierías Tecnológicas.

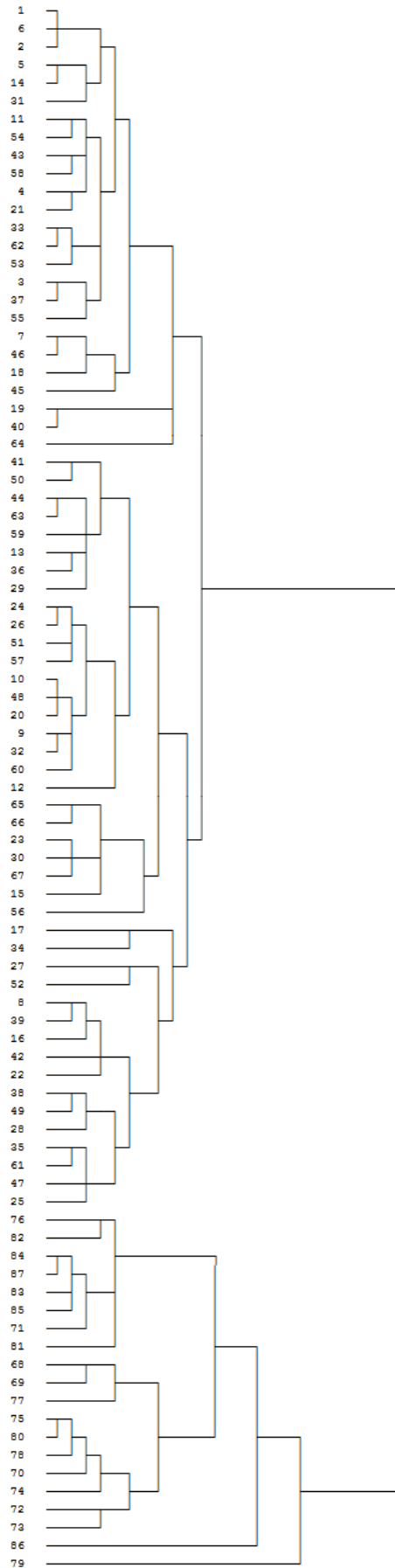


Figura 5.12 Dendrograma de clústeres de las variables de evaluación para el segundo parcial y puntuaciones medias en las variables de agrupamiento de las Ingenierías Tecnológicas.

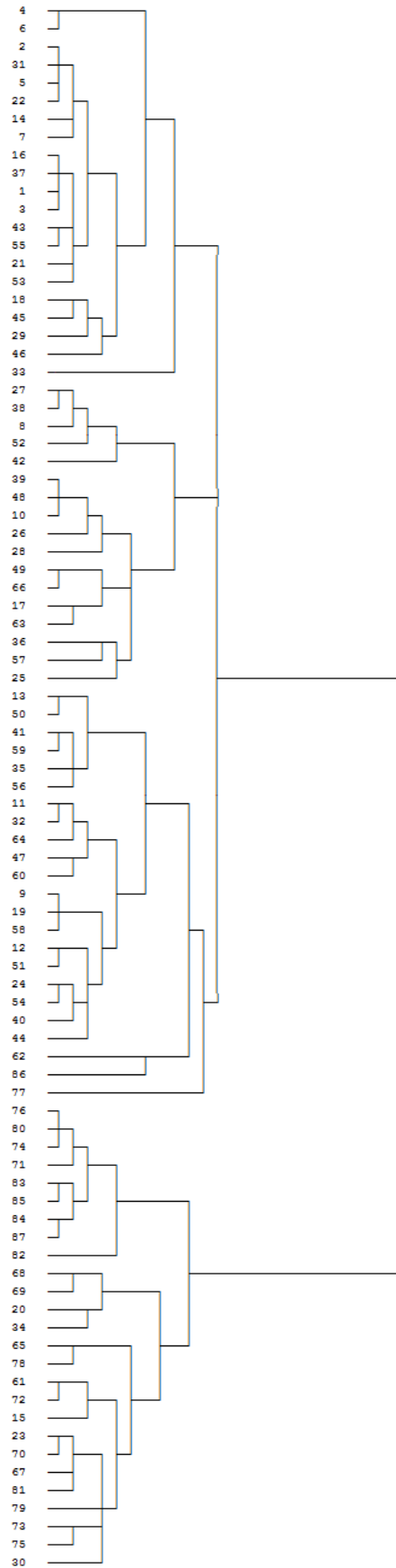


Figura 5.13 Dendrograma de clústeres de las variables de evaluación para el tercer parcial y puntuaciones medias en las variables de agrupamiento de las Ingenierías Tecnológicas.

Se trata de las 3 pruebas parciales que se han efectuado sobre los alumnos de ingenierías tecnológicas. Las figuras que siguen resumen visualmente los resultados:

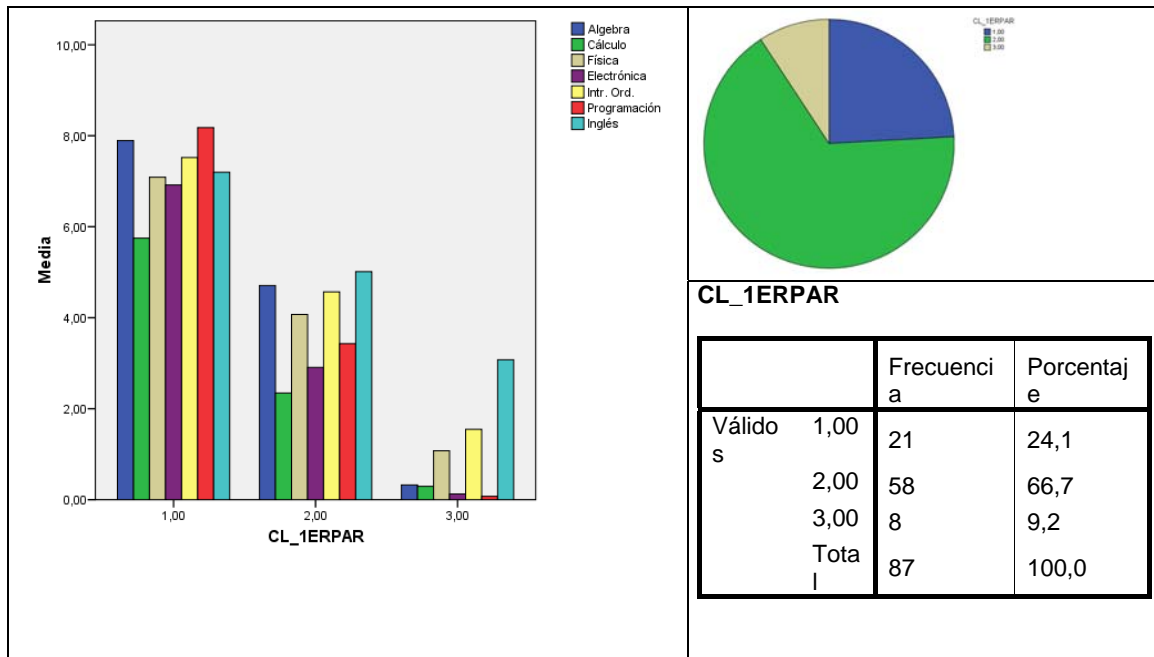


Figura 5.14 Diagrama de pertenencia a los clústeres de las variables de evaluación para el primer parcial y puntuaciones medias en las variables de agrupamiento.

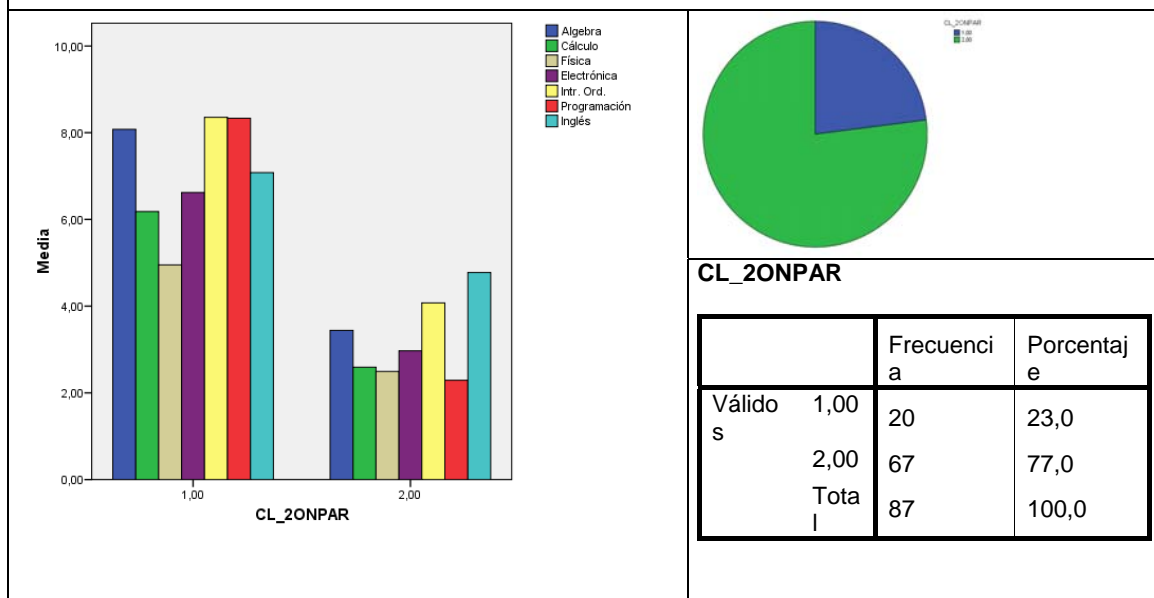


Figura 5.15 Diagrama de pertenencia a los clústeres de las variables de evaluación para el segundo parcial y puntuaciones medias en las variables de agrupamiento.

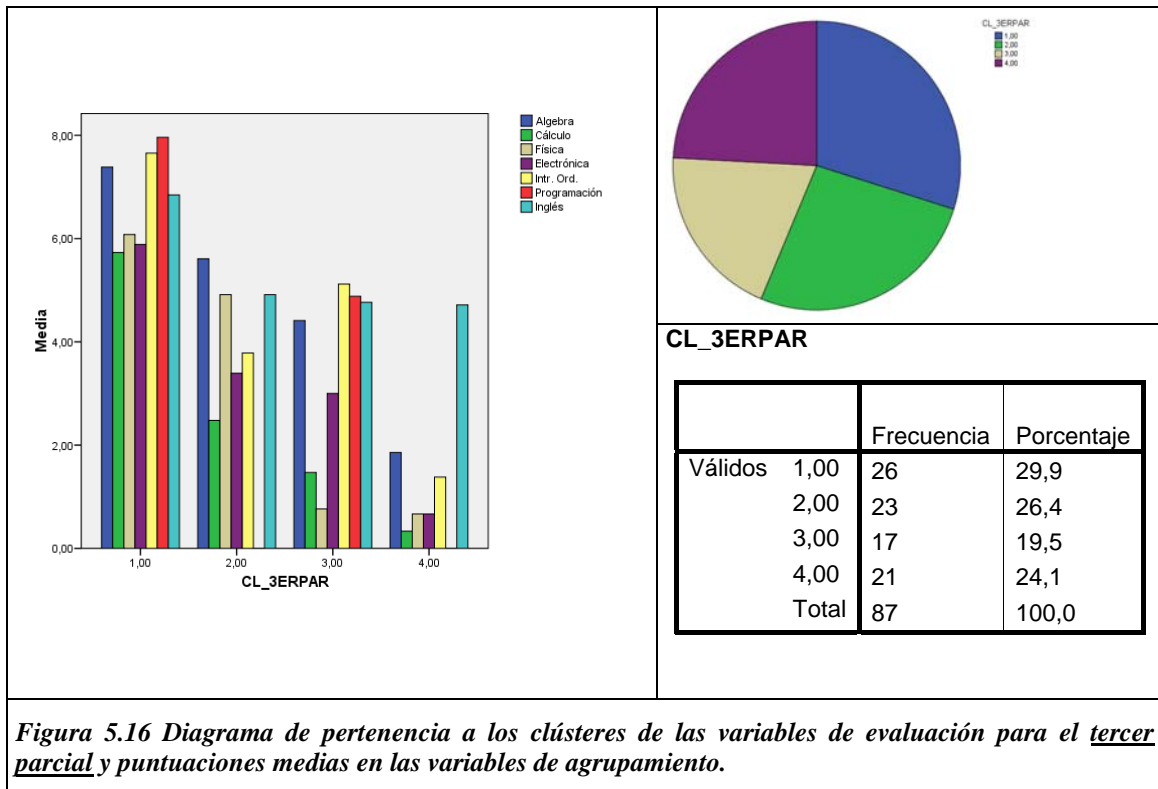


Figura 5.16 Diagrama de pertenencia a los clústeres de las variables de evaluación para el tercer parcial y puntuaciones medias en las variables de agrupamiento.

Las pruebas ANOVA pueden ser consultadas a continuación. Cabe indicar que las pruebas de significación han salido válidas para todas las variables de los tres parciales.

ANOVA P1

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Algebra	Inter-grupos	355,105	2	177,553	56,897	,000
	Intra-grupos	262,129	84	3,121		
	Total	617,234	86			
Cálculo	Inter-grupos	242,157	2	121,078	50,490	,000
	Intra-grupos	201,436	84	2,398		
	Total	443,593	86			
Física	Inter-grupos	245,014	2	122,507	39,352	,000
	Intra-grupos	261,501	84	3,113		
	Total	506,514	86			
Electrónica	Inter-grupos	355,487	2	177,744	88,988	,000
	Intra-grupos	167,781	84	1,997		
	Total	523,269	86			
Intr. Ord.	Inter-grupos	239,890	2	119,945	33,826	,000
	Intra-grupos	297,858	84	3,546		
	Total	537,748	86			
Programación	Inter-grupos	502,424	2	251,212	37,105	,000
	Intra-grupos	568,707	84	6,770		
	Total	1071,131	86			
Inglés	Inter-grupos	119,888	2	59,944	15,005	,000
	Intra-grupos	335,577	84	3,995		
	Total	455,465	86			

Figura 5.17 ANOVA de los clústeres de las variables de rendimiento para el primer parcial de las Ingenierías Tecnológicas.

ANOVA P2

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Álgebra	Inter-grupos	331,456	1	331,456	92,356	,000
	Intra-grupos	305,057	85	3,589		
	Total	636,513	86			
Cálculo	Inter-grupos	198,375	1	198,375	32,841	,000
	Intra-grupos	513,437	85	6,040		
	Total	711,812	86			
Física	Inter-grupos	92,881	1	92,881	28,410	,000
	Intra-grupos	277,892	85	3,269		
	Total	370,773	86			
Electrónica	Inter-grupos	205,348	1	205,348	43,753	,000
	Intra-grupos	398,936	85	4,693		
	Total	604,284	86			
Intr. Ord.	Inter-grupos	282,118	1	282,118	57,882	,000
	Intra-grupos	414,289	85	4,874		
	Total	696,407	86			
Programación	Inter-grupos	562,246	1	562,246	98,209	,000
	Intra-grupos	486,627	85	5,725		
	Total	1048,873	86			
Inglés	Inter-grupos	81,753	1	81,753	20,964	,000
	Intra-grupos	331,474	85	3,900		
	Total	413,227	86			

Figura 5.18 ANOVA de los clústeres de las variables de rendimiento para el segundo parcial de las Ingenierías Tecnológicas.

ANOVA P3

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Álgebra	Inter-grupos	369,679	3	123,226	48,171	,000
	Intra-grupos	212,321	83	2,558		
	Total	582,000	86			
Cálculo	Inter-grupos	383,163	3	127,721	30,309	,000
	Intra-grupos	349,756	83	4,214		
	Total	732,920	86			
Física	Inter-grupos	513,867	3	171,289	60,913	,000
	Intra-grupos	233,398	83	2,812		
	Total	747,264	86			
Electrónica	Inter-grupos	320,121	3	106,707	29,057	,000
	Intra-grupos	304,799	83	3,672		
	Total	624,920	86			
Intr. Ord.	Inter-grupos	479,968	3	159,989	57,111	,000
	Intra-grupos	232,515	83	2,801		
	Total	712,483	86			
Programación	Inter-grupos	1086,607	3	362,202	276,500	,000
	Intra-grupos	108,726	83	1,310		
	Total	1195,333	86			
Inglés	Inter-grupos	76,548	3	25,516	6,219	,001
	Intra-grupos	340,555	83	4,103		
	Total	417,103	86			

Figura 5.19. ANOVA de los clústeres de las variables de rendimiento para el tercer parcial de las Ingenierías Tecnológicas.

5.3.1 Descripción de los clústeres

5.3.1.1 Para el primer parcial

Si observamos el diagrama de la figura 5.14, podemos definir los siguientes clústeres:

- a) Alumnos de rendimiento notable (clúster 1):

En todas las asignaturas puntúan entre el 7 y el 8, excepto en cálculo que vemos como la nota es inferior a las demás.

- b) Alumnos de rendimiento pobre (clúster 2):

Destaca que solo están sobre el aprobado algebra, introducción a los ordenadores e inglés. Mientras que cálculo, física, electrónica y programación no llegan al aprobado.

- c) Alumnos de rendimiento deficiente (clúster 3):

Es un clúster en que no llegan al aprobado en ninguna asignatura. La media más alta es la de inglés aunque no llega a rebasar el 4. Este clúster es el menos numeroso, teniendo en cuenta que es un 9,2% de los 87 estudiantes.

5.3.1.2 Para el segundo parcial

Como podemos observar en el diagrama de la figura 5.15, en este caso tenemos 2 clústeres:

- a) Alumnos de rendimiento notable (clúster 1):

En el primer parcial, la física tenía unos resultados bastante buenos, sin embargo en este parcial la nota roza el aprobado.

- b) Alumnos de rendimiento deficiente (clúster 2):

Se repite el mismo patrón que en el primer parcial para los alumnos de rendimiento deficiente puesto que todas las asignaturas están suspendidas, excepto inglés que tiene la media sobre el aprobado.

5.3.1.3 Para el tercer parcial

El caso del tercer parcial merece un comentario aparte porque la asignatura de introducción a los ordenadores no sigue la tendencia de las demás asignaturas de peso, como álgebra, cálculo, física o electrónica. Sin embargo, ello puede entenderse como que los alumnos se preparan la introducción a los ordenadores para asegurar un aprobado o, como mínimo, sacar nota en una asignatura que ven asequible.

- a) Alumnos de rendimiento notable (clúster 1):

Son alumnos con todas las asignaturas aprobadas que han puesto un especial interés en programación, introducción a los ordenadores y álgebra.

b) Alumnos de rendimiento pobre (clúster 2):

En este clúster las asignaturas que están sobre el aprobado son álgebra, física e inglés. El resto de asignaturas están suspendidas por encima del 2, menos programación que la puntuación es muy deficiente.

c) Alumnos de rendimiento débil (clúster 3):

Las materias sobre el aprobado son introducción a los ordenadores, programación e inglés. Del resto la única asignatura que puntúa por encima del 4 es álgebra.

d) Alumnos de rendimiento deficiente (clúster 4):

La única asignatura que tiene una puntuación cercana al aprobado es inglés. Ninguna de las demás asignaturas tiene la media superior al 2. Este clúster pertenece a aquellos que prácticamente abandonan la titulación.

5.3.2 Reordenación de los clústeres de rendimiento

En el caso de los alumnos de ingenierías tecnológicas, queda claro el diferente nivel de rendimiento de los grupos que, además, sigue el mismo orden en todas las agrupaciones (de mayor a menor rendimiento).

Sin embargo, esta ordenación no es coherente con las variables predictoras, puesto que en ellas la asignación de niveles de rendimiento a clústeres seguía un camino diferente.

Importante: seguimos como criterio la puntuación en las pruebas de acceso de matemáticas y física del centro. Hay que señalar que la prueba de PAU es más cercana a la de La Salle que no la nota del expediente (Véase figura 4.6). En este caso, no hace falta reordenar 3 parciales sino, de manera más simple, directamente los clústeres predictores, tal como se muestra en la siguiente tabla:

Clústeres Variables predictoras		
Inicial		Final
1	Muy Bajo	4
2	Alto	1
3	Medio	2
4	Bajo	3

5.4 Sistema de clústeres resultante

Tal como se ha indicado en el marco teórico (aspectos metodológicos, capítulo 2), para la construcción de un *sistema consistente de clústeres* deben considerarse, al menos, las siguientes condiciones:

- Los clústeres espurios deben eliminarse del sistema consistente.
- Los clústeres de aparición única deben etiquetarse y explicarse.
- Los clústeres admitidos deben superar las pruebas habituales de significación.

- d) Los clústeres deben reordenarse de manera que la interpretación sea cómoda y lo más directa posible.
- e) El número de variables de cada prueba, que produce los clústeres, debe permanecer constante.

Con los criterios del apartado anterior, tenemos por tanto, un diseño de la investigación basado en los *sistemas consistentes de clústeres* que tomará los clústeres de rendimiento como los integrantes principales del sistema. Tal como se ha indicado, en este contexto tendremos dos fuentes más de información:

- a) Los clústeres de las variables predictoras y diversas pruebas estadísticas que se puedan hacer sobre ellas. Esta fuente nos dibuja el panorama inicial del colectivo que cursa una carrera.
- b) La información suministrada por las variables de seguimiento así como otras que se puedan considerar, que será tratada como factores que influyen en la evolución de los clústeres de rendimiento.

El estudio del *sistema consistente de clústeres* se hará para los alumnos de ingeniería multimedia y de ingenierías tecnológicas, con el objetivo de observar semejanzas y diferencias.

La hipótesis subyacente es, pues, que las variables predictoras retratan un escenario inicial, y que la acción del centro (reflejada en datos sobre tutoría y en las variables de seguimiento) actúa influyendo en la evolución de los clústeres.

Así las cosas, un primer retrato comparativo de los sistemas consistentes de clústeres con los que trabajaremos en profundidad, para ingeniería multimedia e ingenierías tecnológicas, es el que se plasma en la figura 4.9.

Capítulo 6

Estudio descriptivo de los sistemas consistentes de clústeres

6.1 Objetivo del estudio

Hay que considerar, a raíz de lo indicado en capítulos anteriores, que la metodología de los sistemas consistentes de clústeres es doblemente heurística:

- a) Por una parte, las investigaciones basadas en análisis de clústeres lo son *per se*, ya que ofrecen normalmente múltiples clasificaciones al investigador, el cual tiene la responsabilidad de elegir una que encaje con los resultados del fenómeno estudiado y lo explique.
- b) Por otra parte, la asignación de los clústeres obtenidos en una prueba a clústeres obtenidos en otra prueba posterior, también tiene esta componente de *trial and error* y también deja un espacio de maniobra al investigador, haciéndole responsable de la eficacia de sus decisiones a la hora de arrojar resultados.

Por tales razones siempre será aconsejable que el investigador disponga, antes de aplicar la metodología de los sistemas consistentes de clústeres, del suficiente apoyo descriptivo que le permita acercarse a los datos para observarlos con detenimiento. Así pues, en la misma metodología, sin ser la parte principal, consideramos que deben realizarse estudios descriptivos como los que exponemos en este capítulo.

El esquema de pruebas estadísticas que proponemos en nuestro método es el siguiente:

- a) Gráficas de “rompimiento individual” de clústeres: reflejan las “deserciones” de los individuos de unos clústeres a otros. Obtendremos una buena visualización de estas gráficas combinando las funciones LINE del paquete estadístico SPSS.
- b) Tablas de contingencia de la evolución entre los clústeres: tablas descriptivas de recuento de individuos que cambian o se mantienen en los clústeres. Se realizará con un CROSSTABS descriptivo del paquete estadístico SPSS.
- c) Gráficas de evolución entre los clústeres que ilustrarán las tablas anteriores. Se realizará con la función BAR del paquete estadístico SPSS.
- d) Pruebas de intensidad de la evolución entre los clústeres: en los casos que se pueda (que se cumplan las condiciones de número mínimo de individuos por celda) se aplicará el Coeficiente de Contingencia C en el marco de una prueba Chi Cuadrado.
- e) Pruebas estadísticas de poder explicativo del primer parcial sobre el tercero: son complementarias a las del apartado d) y nos permiten observar si la relación de contingencia entre los clústeres 1-3 es mayor o menor que la de los clústeres 1-2 y 2-3.

6.2 Evolución de los sistemas consistentes de clústeres de ingeniería multimedia

6.2.1 Estudio preliminar: clústeres predictores y clústeres de rendimiento

Este capítulo lo vamos a tratar de manera puramente descriptiva, a partir de los resultados obtenidos en las diferentes partes del estudio. Para el primer parcial, segundo y tercero, sobre los predictores, las gráficas de evolución son las siguientes:

6.2.2 Análisis de evolución con Crosstabs

Primer parcial

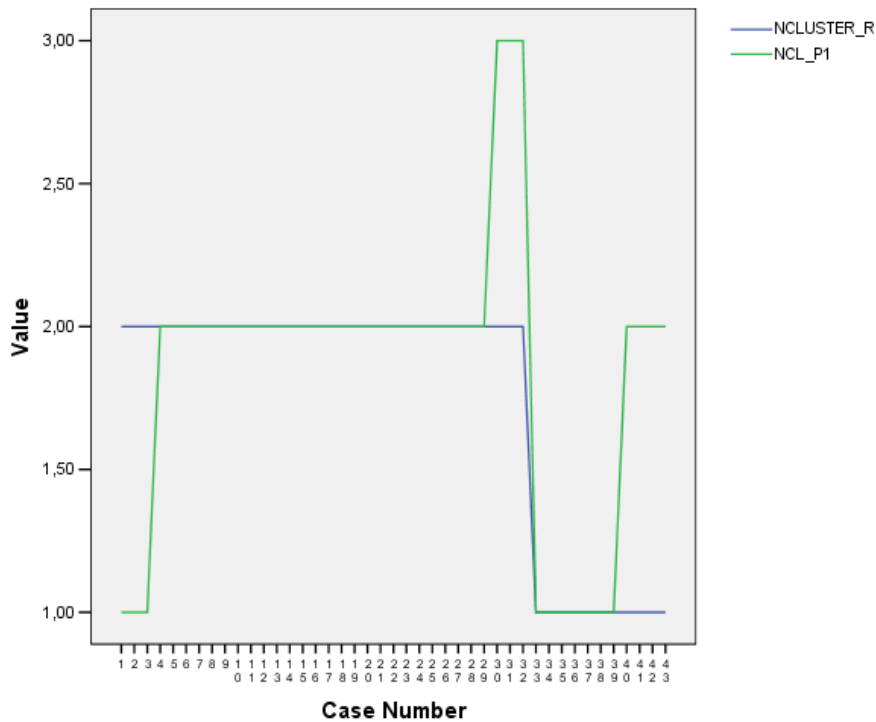


Figura 6.1 Gráfico de rompimiento individual de los predictores al primer parcial.

En este gráfico de rompimiento individual del clúster de predictores al primer parcial podemos ver una leve mejora y pocos alumnos van al clúster 3 (el de peor rendimiento). Esto es debido a que el primer parcial es de repaso.

NCLUSTER_R * NCL_P1 Crosstabulation

		NCL_P1			Total	
		1,00	2,00	3,00		
NCLUSTER_R	1,00	Count	7	4	0	11
		% within NCLUSTER_R	63,6%	36,4%	,0%	100,0%
	2,00	Count	3	26	3	32
		% within NCLUSTER_R	9,4%	81,3%	9,4%	100,0%
Total		Count	10	30	3	43
		% within NCLUSTER_R	23,3%	69,8%	7,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13,759(a)	2	,001
Likelihood Ratio	13,125	2	,001
Linear-by-Linear Association	11,739	1	,001
N of Valid Cases	43		

a 3 cells (50,0%) have expected count less than 5. The minimum expected count is ,77.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,492	,001
N of Valid Cases		43	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

El coeficiente de contingencia nos indica la relación que existe entre un clúster y otro. En este caso podemos ver que tenemos una buena relación. Más adelante la compararemos con los otros parciales.

Bar Chart

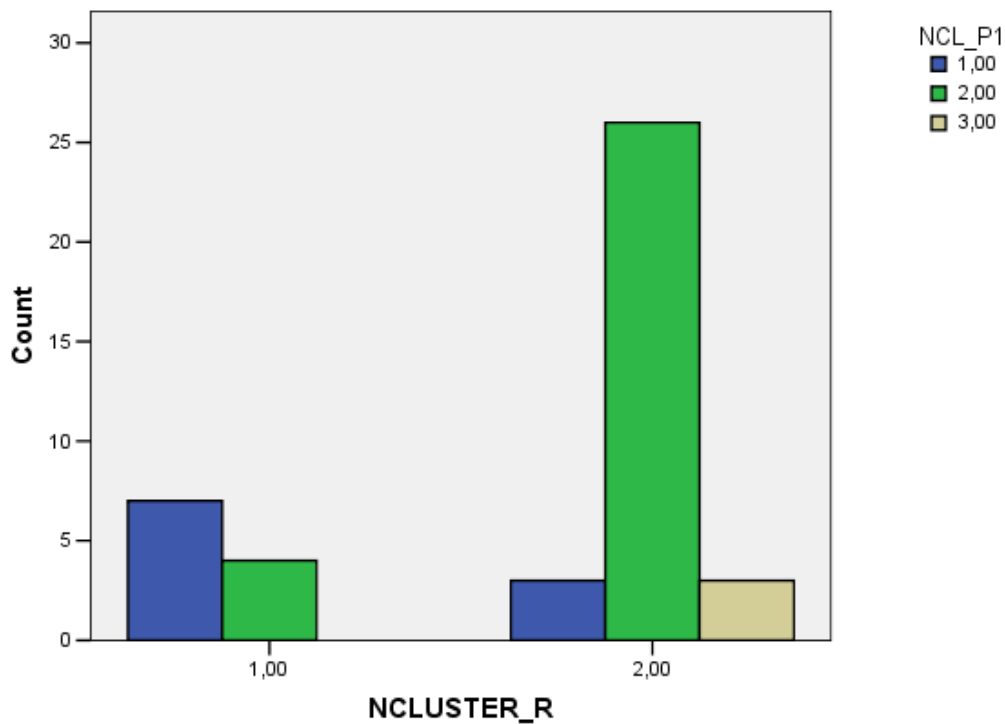


Figura 6.2 Gráfico de evolución de los individuos entre los clústeres predictores hacia los clústeres del primer parcial.

En este gráfico podemos observar como el clúster de los alumnos que tenían un buen rendimiento en el clúster de predictores alimentan a los dos clústeres mejores del primer parcial, siendo casi todos los del clúster de alto rendimiento del primer parcial.

Segundo parcial

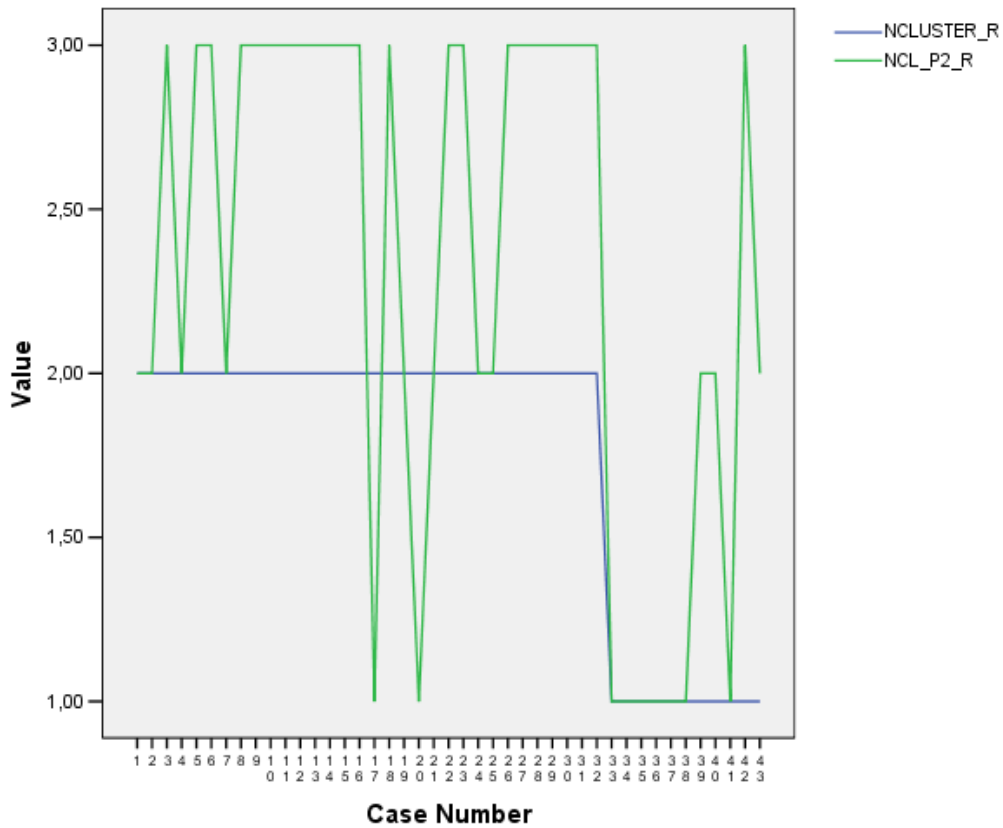


Figura 6.3 Gráfico de rompimiento individual de los predictores al segundo parcial.

En este gráfico observamos una mayor dispersión que el del primer parcial. Han caído muchos alumnos al clúster de peligro.

NCLUSTER_R * NCL_P2_R Crosstabulation

		NCL_P2_R			Total	
		1,00	2,00	3,00		
NCLUSTER_R	1,00	Count	7	3	1	11
		% within NCLUSTER_R	63,6%	27,3%	9,1%	100,0%
	2,00	Count	2	8	22	32
		% within NCLUSTER_R	6,3%	25,0%	68,8%	100,0%
Total		Count	9	11	23	43
		% within NCLUSTER_R	20,9%	25,6%	53,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18,344(a)	2	,000
Likelihood Ratio	18,250	2	,000
Linear-by-Linear Association	17,164	1	,000
N of Valid Cases	43		

a 2 cells (33,3%) have expected count less than 5. The minimum expected count is 2,30.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,547	,000
N of Valid Cases		43	

- a Not assuming the null hypothesis.
 b Using the asymptotic standard error assuming the null hypothesis.

Si nos fijamos en el coeficiente de contingencia vemos que la relación es un poco más fuerte que las del primer parcial.

Bar Chart

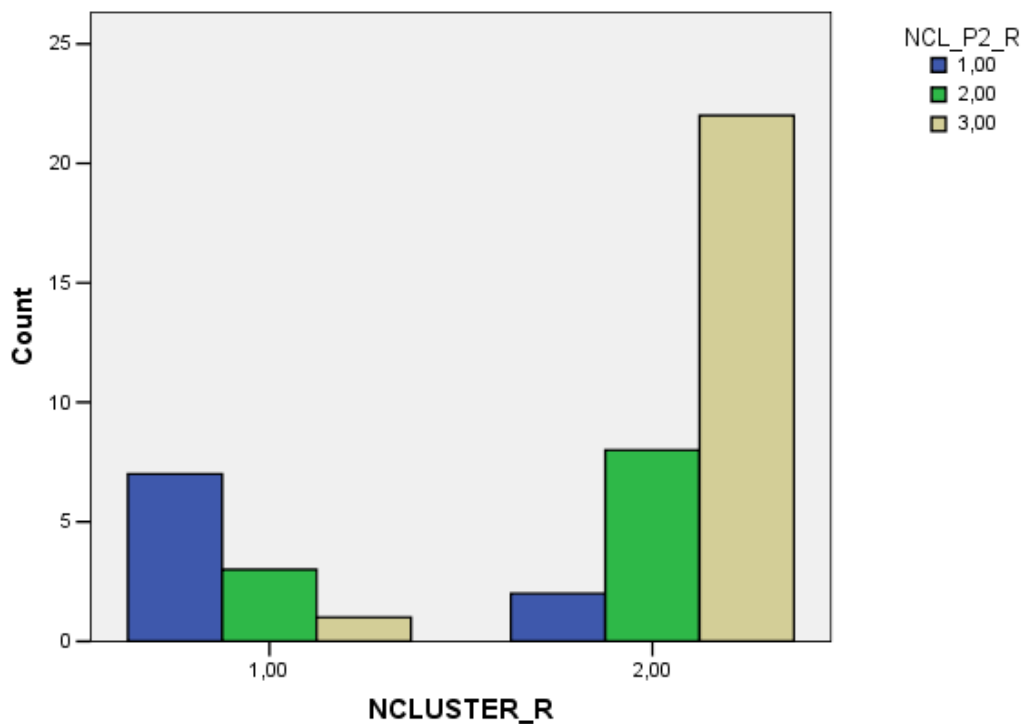


Figura 6.4 Gráfico de evolución de los individuos entre los clústeres predictores hacia los clústeres del segundo parcial.

En el segundo parcial vemos que hay más alumnos que están en el clúster de peligro, debido a que el nivel ya no es de repaso sino propiamente el temario de las asignaturas.

Tercer parcial

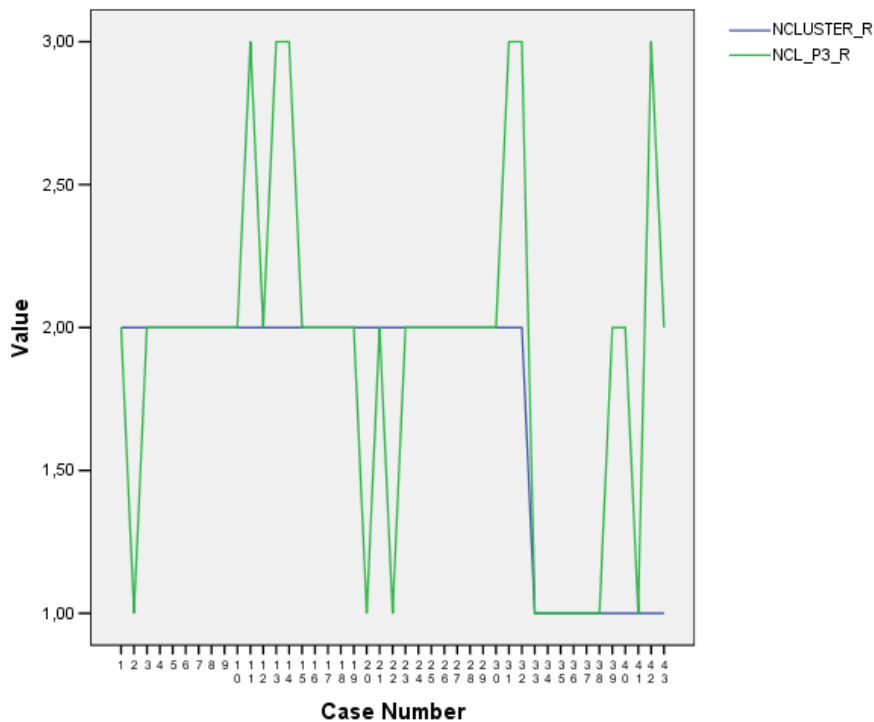


Figura 6.5 Gráfico de rompimiento individual de los predictores al tercer parcial.

En este caso podemos ver como hay una mejora sustancial respecto a los predictores. Aquí es donde se ve reflejado el trabajo de la Universidad para mejorar el rendimiento de los estudiantes.

NCLUSTER_R * NCL_P3_R Crosstabulation

			NCL_P3_R			Total
			1,00	2,00	3,00	
NCLUSTER_R	1,00	Count	7	3	1	11
		% within NCLUSTER_R	63,6%	27,3%	9,1%	100,0%
	2,00	Count	3	24	5	32
		% within NCLUSTER_R	9,4%	75,0%	15,6%	100,0%
Total		Count	10	27	6	43
		% within NCLUSTER_R	23,3%	62,8%	14,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13,584(a)	2	,001
Likelihood Ratio	12,441	2	,002
Linear-by-Linear Association	8,131	1	,004
N of Valid Cases	43		

a 3 cells (50,0%) have expected count less than 5. The minimum expected count is 1,53.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,490	,001
N of Valid Cases		43	

- a Not assuming the null hypothesis.
 b Using the asymptotic standard error assuming the null hypothesis.

El coeficiente de contingencia continua manteniéndose sobre los mismos valores, el significado de estos valores nos indica que si un alumno ha sacado buena nota en las pruebas de predictores es probable que también obtenga buena nota en los parciales. Así vemos como las pruebas preparadas por La Salle tienen una relación bastante directa con el rendimiento de los alumnos en los posteriores parciales.

Bar Chart

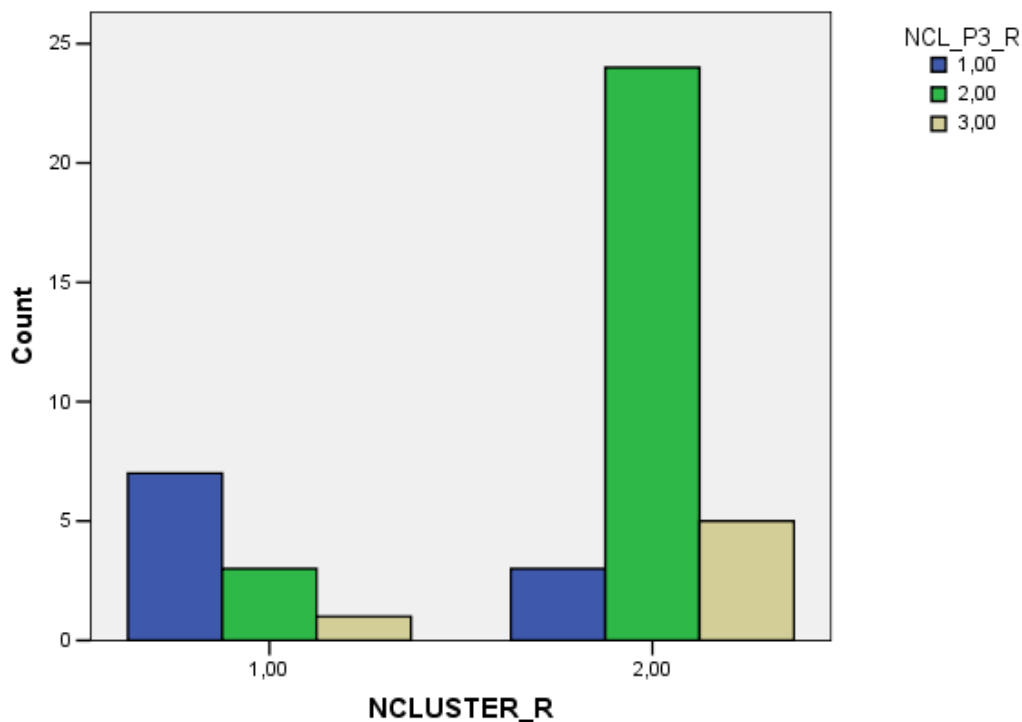


Figura 6.6 Gráfico de evolución de los individuos entre los clústeres predictores hacia los clústeres del tercer parcial.

En este tercer parcial vemos como se recuperan bastantes alumnos que habían caído al peor clúster.

6.2.3 Gráficas de la evolución de los clústeres de rendimiento

En este apartado vamos a ver cómo evolucionan los clústeres de evolución a través de los parciales.

6.2.3.1 Análisis de evolución con Crosstabs

Del primer parcial al segundo

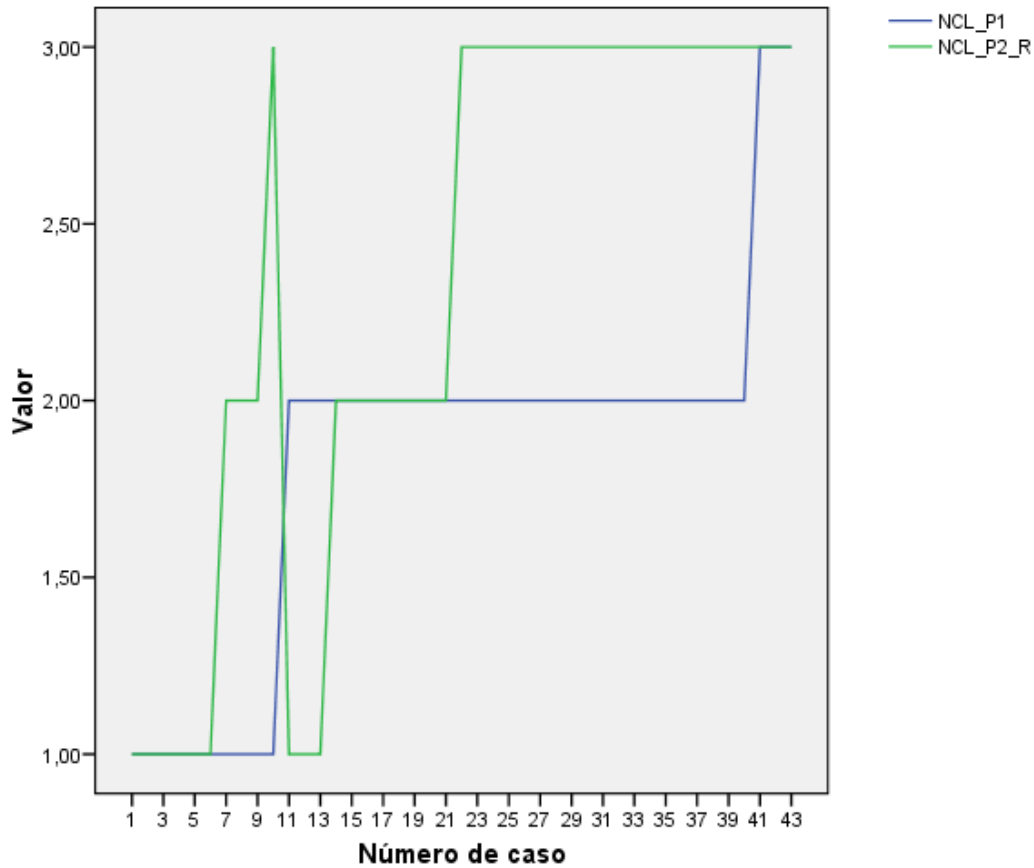


Figura 6.7 Gráfico de rompimiento individual del primer al segundo parcial.

CL_NOT1 * CL_NOT2 Crosstabulation

			CL_NOT2			Total
			1,00	2,00	3,00	
CL_NOT1	1,00	Count	6	3	1	10
		% of Total	14,0%	7,0%	2,3%	23,3%
	2,00	Count	3	8	19	30
		% of Total	7,0%	18,6%	44,2%	69,8%
	3,00	Count	0	0	3	3
		% of Total	,0%	,0%	7,0%	7,0%
Total		Count	9	11	23	43
		% of Total	20,9%	25,6%	53,5%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15,784(a)	4	,003
Likelihood Ratio	16,647	4	,002
Linear-by-Linear Association	13,635	1	,000
N of Valid Cases	43		

a 5 cells (55,6%) have expected count less than 5. The minimum expected count is ,63.

Symmetric Measures

Nominal by Nominal	Value	Approx. Sig.
Contingency Coefficient	,518	,003
N of Valid Cases	43	

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Bar Chart

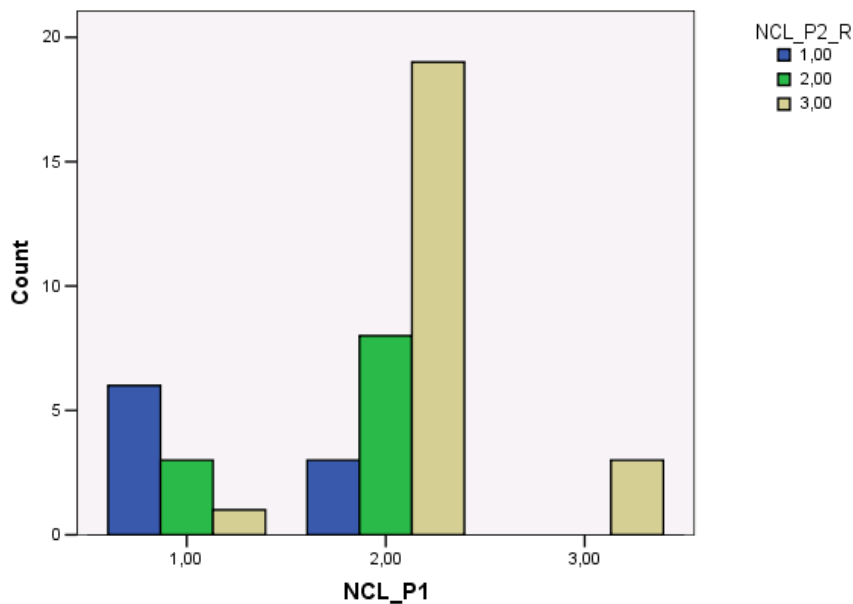


Figura 6.8 Gráfico de evolución de los individuos entre los clústeres del primer parcial hacia los clústeres del segundo parcial.

Comparando el primer parcial con el segundo vemos la decaída de los tres clústeres. El clúster de los alumnos de rendimiento medio es el que más alimenta al clúster de los de bajo rendimiento pero observamos también como algún alumno que estaba en rendimiento alto baja hasta el último clúster. Y el clúster de los de peor rendimiento no alimenta a ningún otro clúster.

Del segundo parcial al tercero

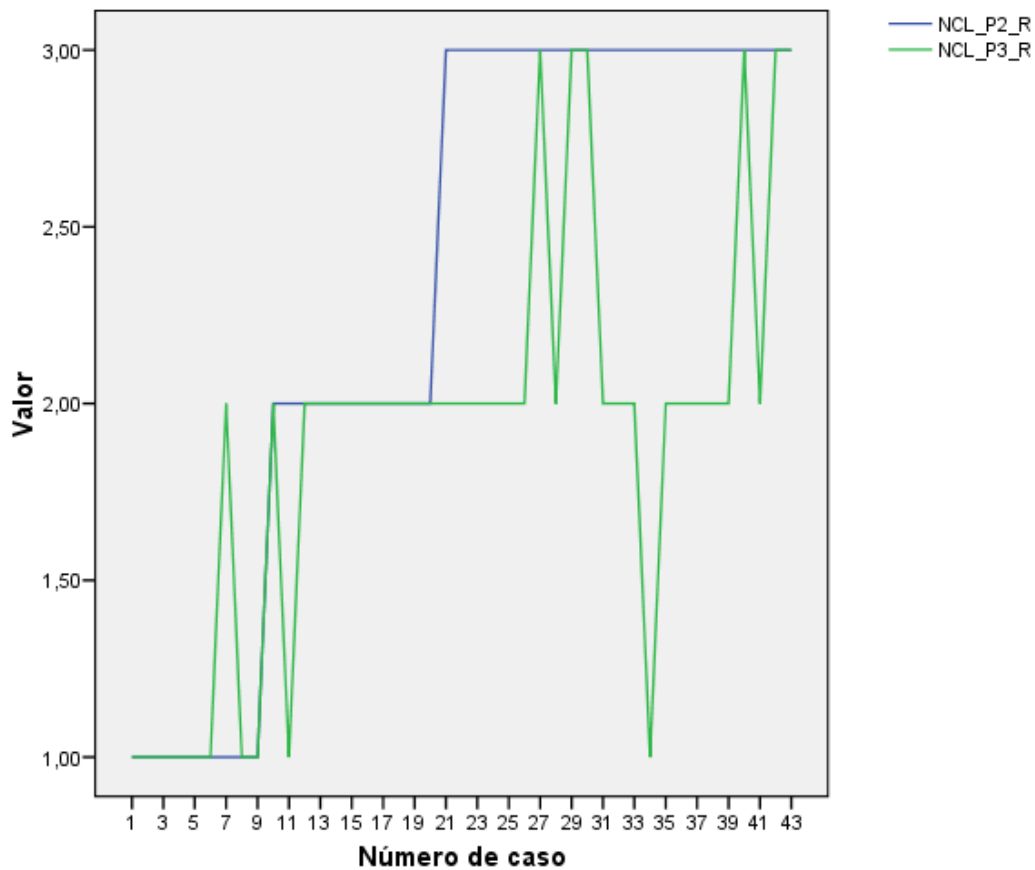


Figura 6.9 Gráfico de rompimiento individual del segundo al tercer parcial.

CL_NOT2 * CL_NOT3 Crosstabulation

		CL_NOT3			Total	
		1,00	2,00	3,00		
CL_NOT2	1,00	Count	8	1	0	9
		% of Total	18,6%	2,3%	,0%	20,9%
	2,00	Count	1	10	0	11
		% of Total	2,3%	23,3%	,0%	25,6%
	3,00	Count	1	16	6	23
		% of Total	2,3%	37,2%	14,0%	53,5%
Total		Count	10	27	6	43
		% of Total	23,3%	62,8%	14,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	31,754(a)	4	,000
Likelihood Ratio	30,945	4	,000
Linear-by-Linear Association	20,033	1	,000
N of Valid Cases	43		

a 5 cells (55,6%) have expected count less than 5. The minimum expected count is 1,26.

Symmetric Measures

Nominal by Nominal	Value	Approx. Sig.
Contingency Coefficient	,652	,000
N of Valid Cases	43	

- a Not assuming the null hypothesis.
 b Using the asymptotic standard error assuming the null hypothesis.

Bar Chart

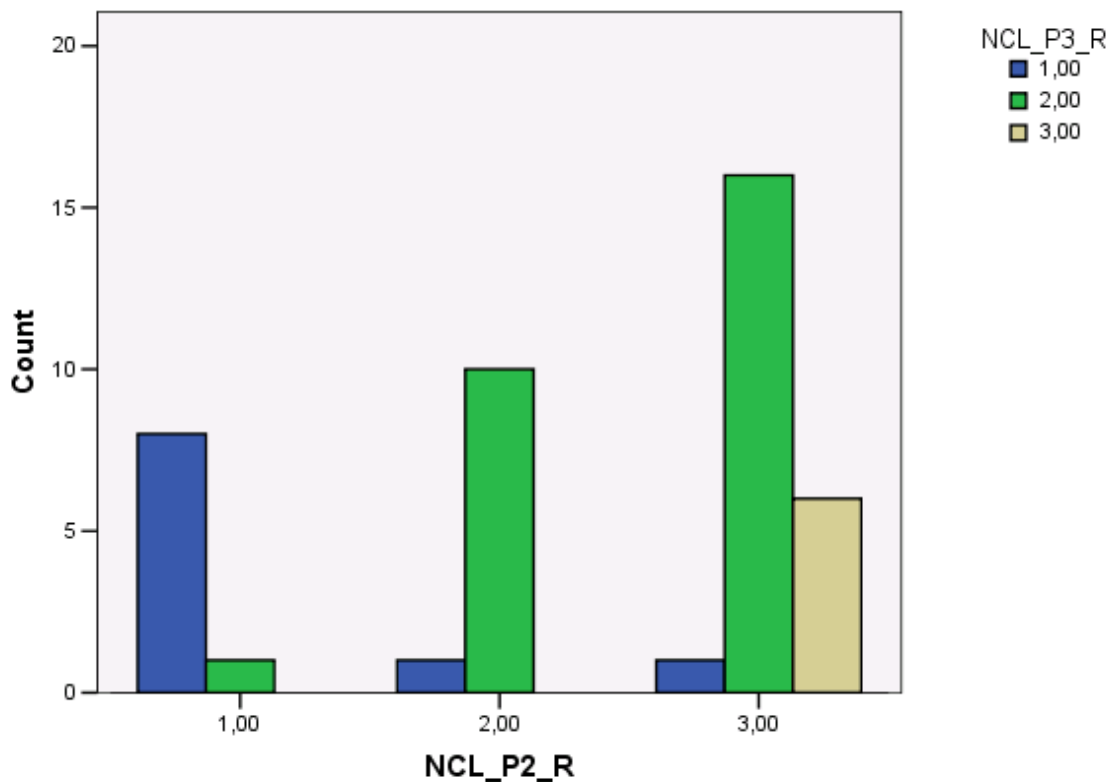


Figura 6.10 Gráfico de evolución de los individuos entre los clústeres del segundo parcial hacia los clústeres del tercer parcial.

En esta evolución del segundo al tercer parcial podemos observar una consistente mejora general ya que sobretodo se recuperan alumnos del clúster de bajo rendimiento pasando al clúster de rendimiento medio.

6.2.4 Análisis de correlación directa sobre las notas de matemáticas

Hemos querido hacer un análisis directamente a la relación que existe entre los resultados de la prueba de predictores de matemáticas y la asignatura de álgebra durante el curso. Con esto tendremos una visión de que realmente existe una relación significativa entre las diferentes pruebas.

Correlations

		NMATES	P1ALG
NMATES	Pearson Correlation	1	,510(**)
	Sig. (2-tailed)		,000
	N	43	43
P1ALG	Pearson Correlation	,510(**)	1
	Sig. (2-tailed)	,000	
	N	43	43

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

		NMATES	P2ALG
NMATES	Pearson Correlation	1	,549(**)
	Sig. (2-tailed)		,000
	N	43	43
P2ALG	Pearson Correlation	,549(**)	1
	Sig. (2-tailed)	,000	
	N	43	43

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

		NMATES	FINALG
NMATES	Pearson Correlation	1	,524(**)
	Sig. (2-tailed)		,000
	N	43	43
FINALG	Pearson Correlation	,524(**)	1
	Sig. (2-tailed)	,000	
	N	43	43

** Correlation is significant at the 0.01 level (2-tailed).

Observamos que en todo momento la prueba de la correlación de Pearson nos da un valor entre el 0.5 y el 0.6. Este resultado indica que hay una tendencia de que si un alumno ya obtuvo buenos resultados en las pruebas predictoras es probable que también obtenga buenos resultados en los demás parciales.

6.3 Evolución de los sistemas consistentes de clústeres de ingenierías tecnológicas.

6.3.1 Estudio preliminar: clústeres predictores y clústeres de rendimiento

Así como hemos hecho en multimedia, vamos a tratar de manera descriptiva los resultados obtenidos durante las diferentes fases del estudio. Las gráficas para los parciales respecto de los predictores son las siguientes:

6.3.2 Análisis de evolución con Crosstabs

Primer Parcial

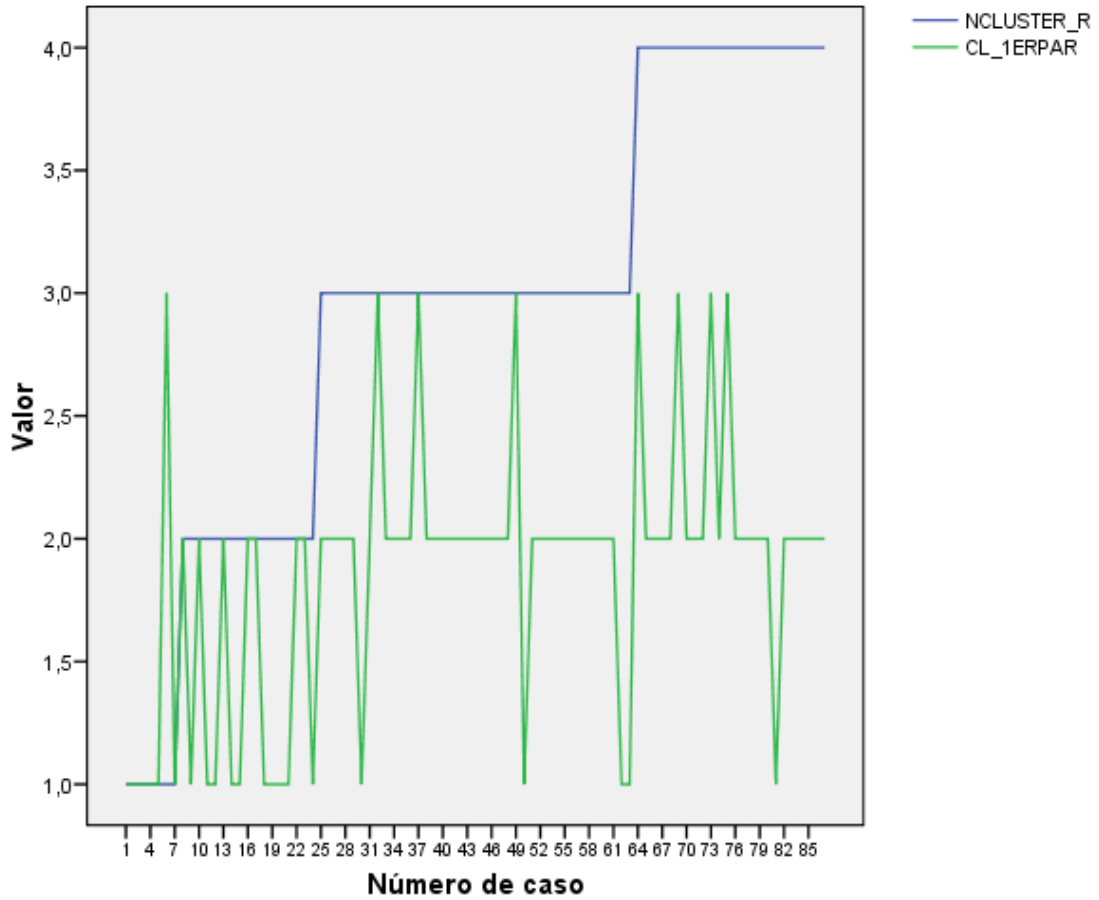


Figura 6.11 Gráfico de rompimiento individual de los predictores al primer parcial.

NCLUSTER_R * CL_1ERPAR Crosstabulation

			CL_1ERPAR			Total
			1,00	2,00	3,00	
NCLUSTER_R	1	Count	6	0	1	7
		% within NCLUSTER_R	85,7%	,0%	14,3%	100,0%
	2	Count	10	7	0	17
		% within NCLUSTER_R	58,8%	41,2%	,0%	100,0%
	3	Count	4	32	3	39
		% within NCLUSTER_R	10,3%	82,1%	7,7%	100,0%
	4	Count	1	19	4	24
		% within NCLUSTER_R	4,2%	79,2%	16,7%	100,0%
Total		Count	21	58	8	87
		% within NCLUSTER_R	24,1%	66,7%	9,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38,132 ^a	6	,000
Likelihood Ratio	40,303	6	,000
Linear-by-Linear Association	22,385	1	,000
N of Valid Cases	87		

a. 7 cells (58,3%) have expected count less than 5. The minimum expected count is ,64.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,552	,000
N of Valid Cases		87	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Y las tablas de evolución de los clústeres son:

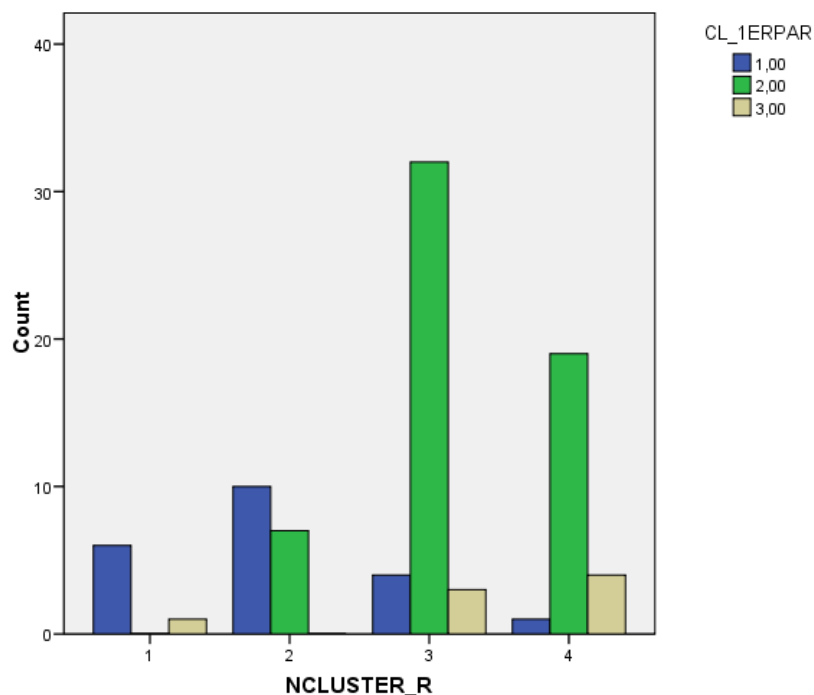


Figura 6.12 Desplazamiento de los clústeres de las variables predictoras hacia las del 1er parcial en valores absolutos.

En las ingenierías tecnológicas partimos de cuatro clústeres predictores hacia tres clústeres de rendimiento del primer parcial. Recordamos que el azul son los alumnos de alto rendimiento, el verde los de rendimiento medio y el amarillo los alumnos de peor rendimiento.

Segundo parcial

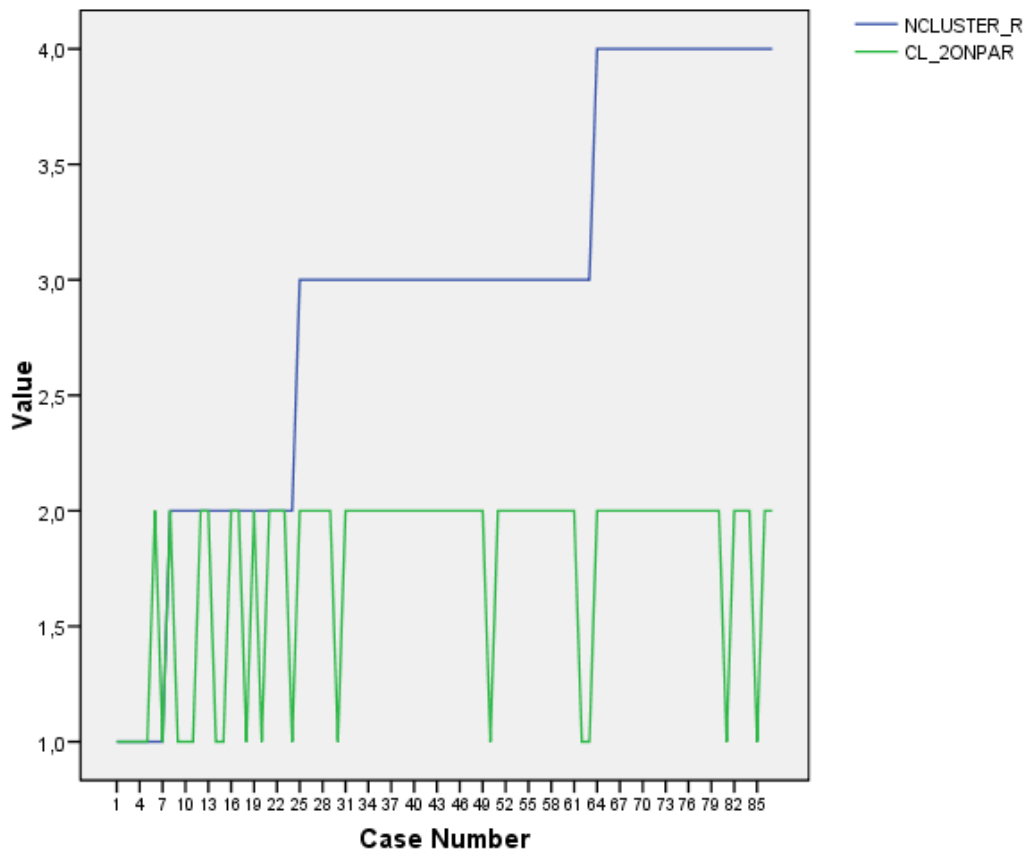


Figura 6.13 Gráfico de rompimiento individual de los predictores al segundo parcial.

Vemos como se pierde el tráfico. Para más detalle observemos las tablas siguientes:

NCLUSTER_R * CL_2ONPAR Crosstabulation

			CL_2ONPAR		Total
			1,00	2,00	
NCLUSTER_R 1	Count	6	1	7	
	% within NCLUSTER_R	85,7%	14,3%	100,0%	
2	Count	8	9	17	
	% within NCLUSTER_R	47,1%	52,9%	100,0%	
3	Count	4	35	39	
	% within NCLUSTER_R	10,3%	89,7%	100,0%	
4	Count	2	22	24	
	% within NCLUSTER_R	8,3%	91,7%	100,0%	
Total	Count	20	67	87	
	% within NCLUSTER_R	23,0%	77,0%	100,0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27,603 ^a	3	,000
Likelihood Ratio	24,999	3	,000
Linear-by-Linear Association	21,919	1	,000
N of Valid Cases	87		

a. 2 cells (25,0%) have expected count less than 5. The minimum expected count is 1,61.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,491	,000
N of Valid Cases		87	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

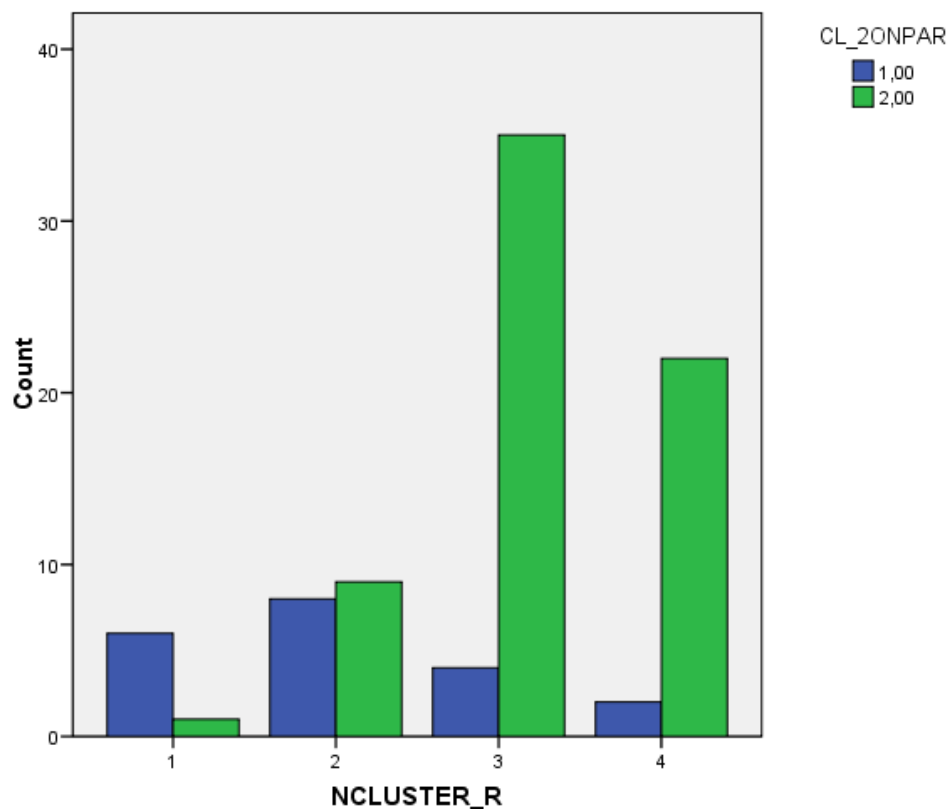


Figura 6.14 Desplazamiento de los clústeres de las variables predictoras hacia las del segundo parcial en valores absolutos.

En el segundo parcial respecto al predictor se pasa a tener dos clústeres.

Tercer Parcial

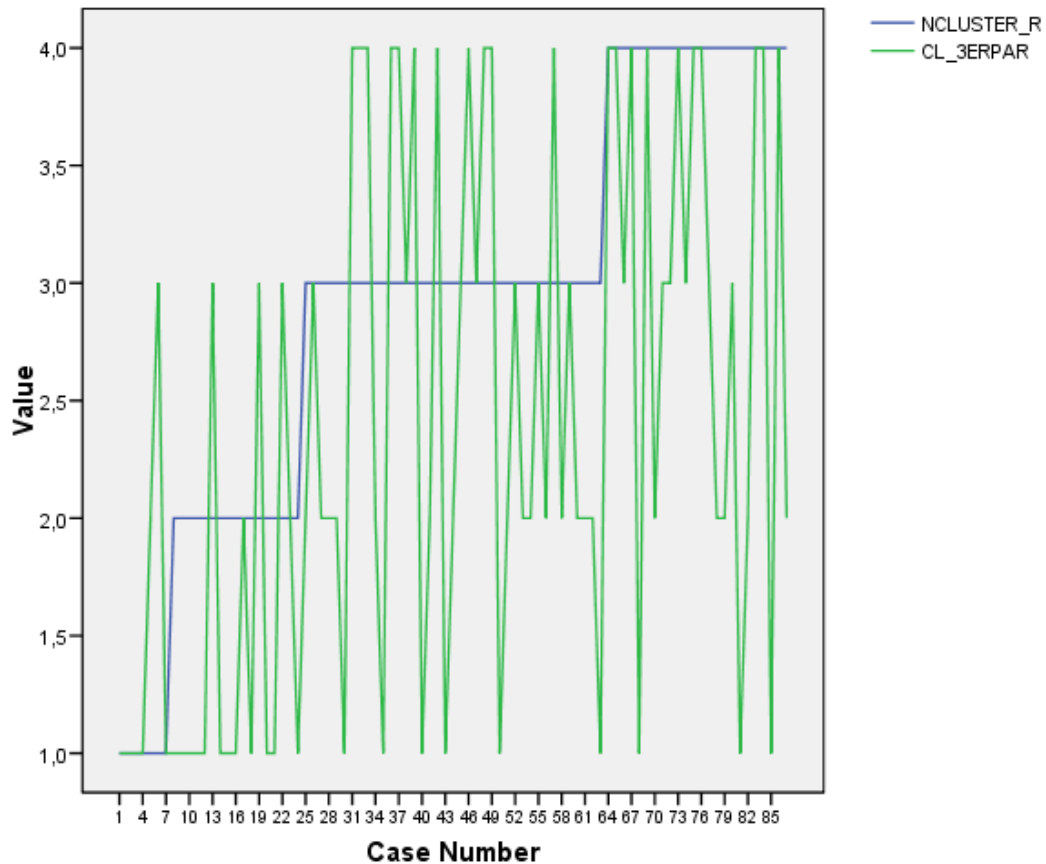


Figura 6.15 Gráfico de rompimiento individual de los predictores al tercer parcial.

El rompimiento individual es muy disperso. No tenemos tráfico definido.

NCLUSTER_R * CL_3ERPAR Crosstabulation

			CL_3ERPAR				Total
			1,00	2,00	3,00	4,00	
NCLUSTER_R	1	Count	5	1	1	0	7
		% within NCLUSTER_R	71,4%	14,3%	14,3%	,0%	100,0%
	2	Count	12	2	3	0	17
		% within NCLUSTER_R	70,6%	11,8%	17,6%	,0%	100,0%
	3	Count	6	15	7	11	39
		% within NCLUSTER_R	15,4%	38,5%	17,9%	28,2%	100,0%
	4	Count	3	5	6	10	24
		% within NCLUSTER_R	12,5%	20,8%	25,0%	41,7%	100,0%
Total		Count	26	23	17	21	87
		% within NCLUSTER_R	29,9%	26,4%	19,5%	24,1%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	32,494 ^a	9	,000
Likelihood Ratio	35,413	9	,000
Linear-by-Linear Association	19,972	1	,000
N of Valid Cases	87		

a. 8 cells (50,0%) have expected count less than 5. The minimum expected count is 1,37.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,521	,000
N of Valid Cases		87	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

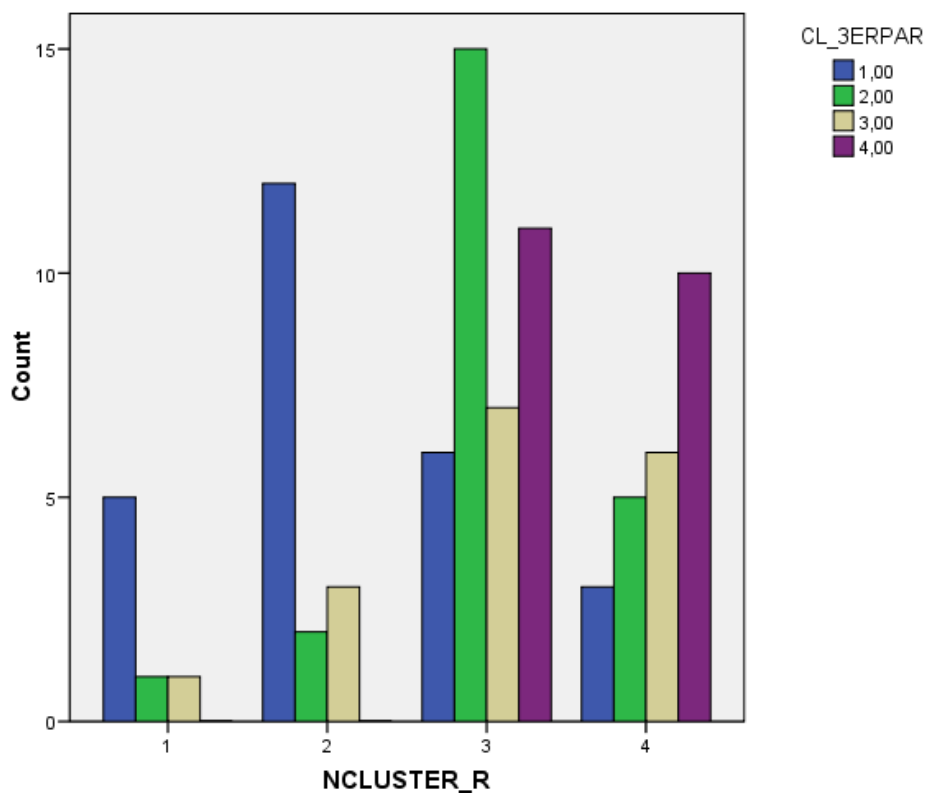


Figura 6.16 Desplazamiento de los clústeres de las variables predictoras hacia las del tercer parcial en valores absolutos.

Lo cual nos lleva a pensar en hacer el procedimiento normal de la evolución del sistema consistente de clústeres.

6.3.3 Gráficas de la evolución de los clústeres de rendimiento

6.3.3.1 Análisis de evolución con crosstabs

Del primer parcial al segundo

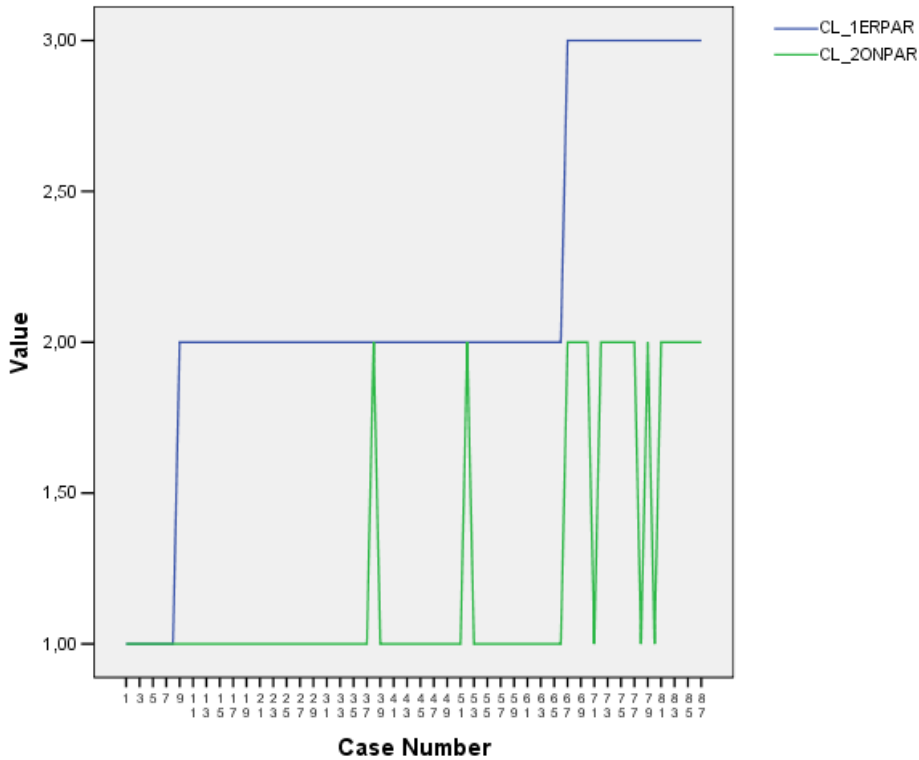


Figura 6.17 Evolución del primer parcial al segundo de los individuos en el clúster.

CL_1ERPAR * CL_2ONPAR Crosstabulation

			CL_2ONPAR		Total
			1,00	2,00	
CL_1ERP AR	1,00	Count	8	0	8
		% of Total	9,2%	,0%	9,2%
	2,00	Count	56	2	58
		% of Total	64,4%	2,3%	66,7%
	3,00	Count	3	18	21
		% of Total	3,4%	20,7%	24,1%
Total		Count	67	20	87
		% of Total	77,0%	23,0%	100,0%

Symmetric Measures

Nominal by Nominal	Value	Approx. Sig.
Contingency Coefficient	,644	,000
N of Valid Cases	87	

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.

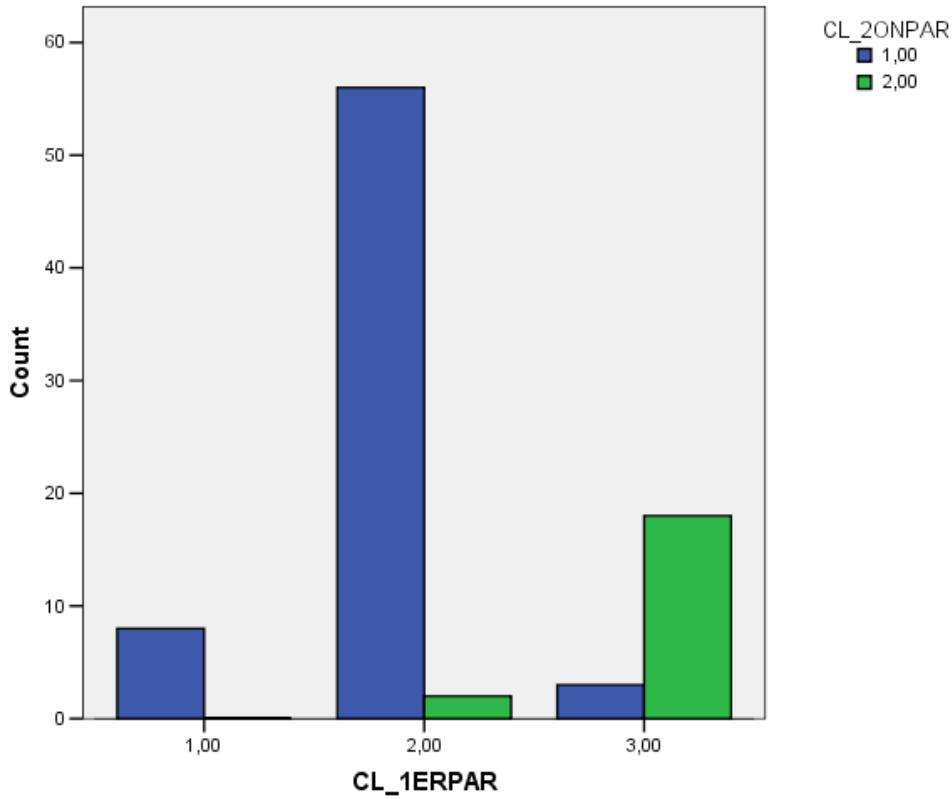


Figura 6.18 Evolución de los clústeres del primer parcial al segundo.

Observamos el mismo fenómeno que con los alumnos de multimedia, se denota un empeoramiento en los dos clústeres, ya que el primer parcial es de repaso.

Del segundo al tercero

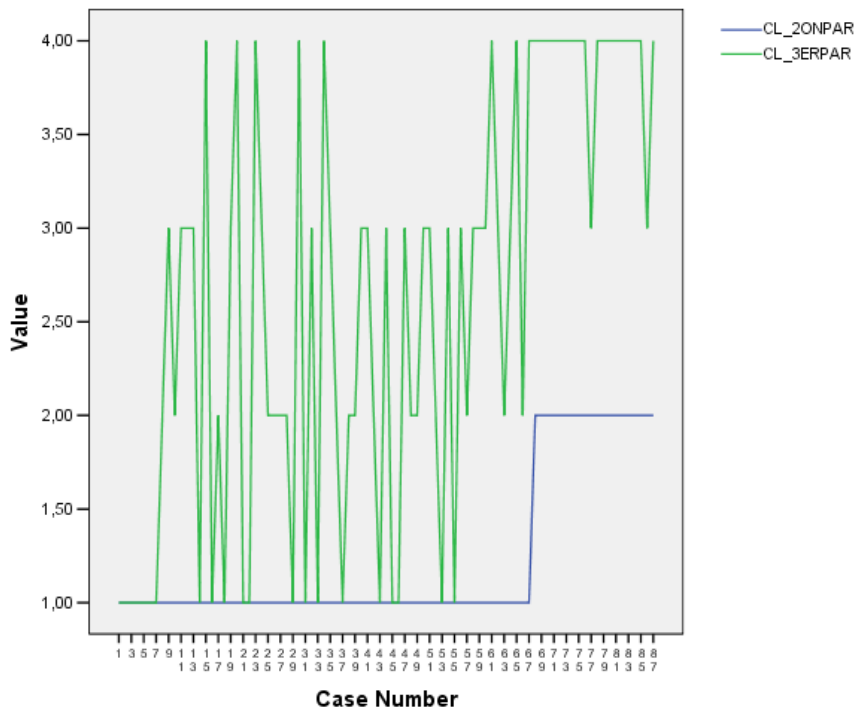


Figura 6.19 “Rompimiento individual” entre el segundo y el tercer parcial.

CL_2ONPAR * CL_3ERPAR Crosstabulation

			CL_3ERPAR				Total
			1,00	2,00	3,00	4,00	
CL_2ONPAR	1,00	Count	21	17	21	8	67
		% of Total	24,1%	19,5%	24,1%	9,2%	77,0%
	2,00	Count	0	0	2	18	20
		% of Total	,0%	,0%	2,3%	20,7%	23,0%
Total		Count	21	17	23	26	87
		% of Total	24,1%	19,5%	26,4%	29,9%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	45,401(a)	3	,000
Likelihood Ratio	48,123	3	,000
Linear-by-Linear Association	31,929	1	,000
N of Valid Cases	87		

a 2 cells (25,0%) have expected count less than 5. The minimum expected count is 3,91.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	,586	,000
N of Valid Cases		87	

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.

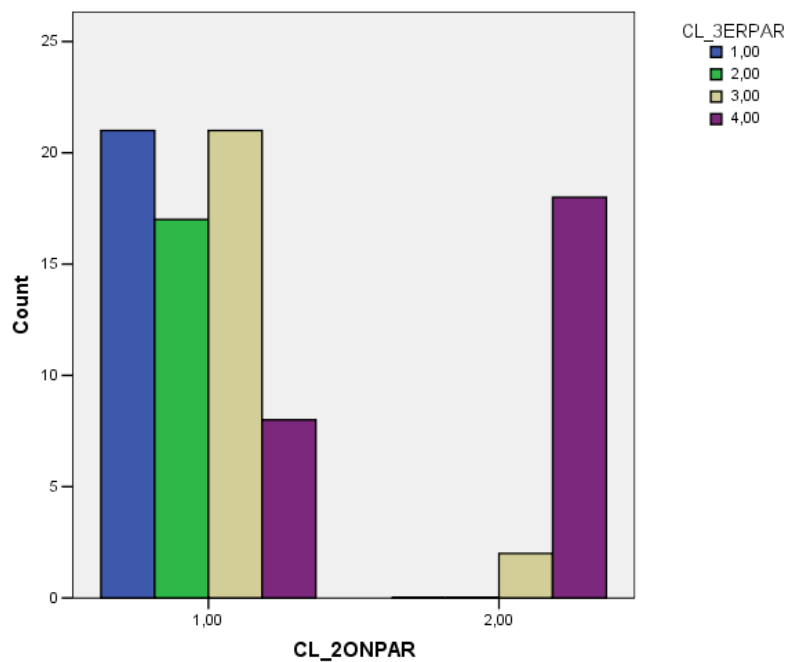


Figura 6.20 Evolución del segundo al tercer parcial.

En esta evolución hay una alimentación de 4 clústeres sobre 2.

Una primera conclusión que podríamos extraer entre los alumnos de Multimedia y los de las Ingenierías Tecnológicas, son las diferencias que se observan a través de la evolución de los clústeres, que denotan asimismo las diferencias en las personalidades de los alumnos, ya que los multimedia son más homogéneos que los tecnológicos.

Capítulo 7

Análisis bajo el enfoque de las cadenas de Markov

7.1 Recapitulación

En los capítulos precedentes, 5 y 6, hemos construido los sistemas de clústeres que explican el rendimiento académico de los alumnos en ingeniería multimedia e ingenierías tecnológicas. La labor realizada hasta el momento ha sido, metodológicamente, descriptiva y de fundamentación. Es decir:

- a) Descriptiva: Se ha elaborado una “imagen” del fenómeno estudiado (descripción de los clústeres implicados, etiquetación de los mismos, pruebas descriptivas de evolución, etc.).
- b) De fundamentación: Se han establecido los sistemas consistentes de clústeres (biunívocos o no) que mejor se adaptan a las puntuaciones obtenidas por los estudiantes y que, a partir de ahora, serán los elementos que utilizará la presente tesis para llegar a sus conclusiones más importantes.

Llegados a este punto, el hecho que tengamos sistemas consistentes de clústeres (por tanto, una identificación conceptual clara entre los grupos resultantes en cada clasificación) nos permite tratar dichos clústeres como las mismas entidades vistas en estados diferentes. En tal situación, por tanto, es aplicable el modelo de análisis de cadenas de Markov.

Siguiendo lo establecido en la metodología que esta tesis desarrolla, tanto para el grupo de estudiantes multimedia como para los tecnológicos, se va a realizar el siguiente procedimiento:

- a) Establecer los componentes de las cadenas de Markov (gráficos de evoluciones de estados y matrices de transición) de manera que se ajusten a la realidad estudiada.
- b) Calcular los estados esperados según el modelo de estas cadenas y medir su discrepancia con los estados reales.
- c) Realizar las pruebas estadísticas para evaluar la aportación académica del centro.

Como se verá, este procedimiento no está exento de cierta heurística, dado que pueden encontrarse modelos de clústeres parcialmente explicativos pero que, al final, no nos sirvan para medir la aportación académica del centro. En el caso de los estudiantes de ingenierías tecnológicas veremos una situación estadísticamente complicada que nos permite desarrollar nuestra metodología con todos sus recursos, aunque se abandone la línea más purista de su aplicación.

7.2 Evolución de los clústeres en ingeniería multimedia

El sistema consistente de clústeres obtenido en el capítulo 6 ha resultado relativamente sencillo y altamente ilustrador. Al tratarse de tres clústeres en un sistema biunívoco todo parece indicar que no tendremos las complicaciones de las ingenierías tecnológicas, en las que la no biunivocidad del sistema puede complicar los cálculos.

Ahora bien, el sistema biunívoco no será útil si no nos permite descubrir la intervención educativa del centro. Podría suceder que encajara bien el modelo pero que fuera inútil a la hora de producir resultados estadísticamente significativos que nos permitieran validar la metodología docente. Sin más dilación, pues, iniciemos el procedimiento para los estudiantes de multimedia.

Los pasos que seguiremos son:

- Elaborar los Grafos y matrices de transiciones entre el primer y segundo parcial
- Elaborar los Grafos y matrices de transiciones entre el segundo y tercer parcial
- Calcular la evolución esperada para el tercer parcial a partir de la matriz primero-segundo
- Elaborar los grafos y matrices de transiciones entre el primer y el tercer parcial
- Realizar la prueba de comparación de discrepancias con los resultados finales (chi cuadrado)

7.2.1 Transiciones entre el primer y segundo parcial: Grafos y matrices de transición

Tal como hemos visto en el capítulo anterior, la distribución de los alumnos en el segundo parcial, considerando su pertenencia a los clústeres del primero, se ajusta a la siguiente gráfica:

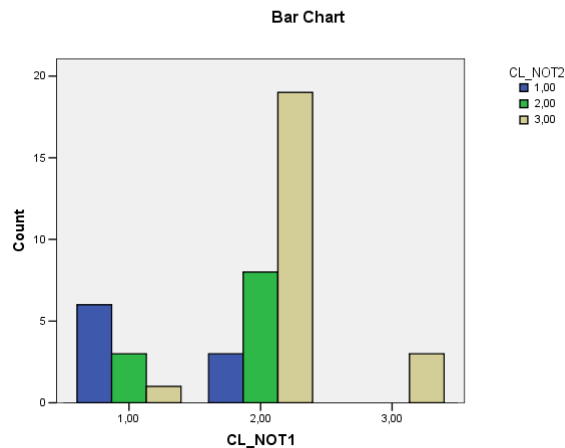


Figura 7.1 Multimedia. Diagrama de barras entre el primer y segundo parcial.

Para obtener, a partir de la misma, las probabilidades de transición, realizamos una prueba CROSSTABS con el paquete estadístico SPSS/PC obteniendo la siguiente tabla:

CL_NOT1		CL_NOT2			Total
		1,00	2,00	3,00	
1,00	Count	6	3	1	10
	% within CL_NOT1	60,0%	30,0%	10,0%	100,0%
2,00	Count	3	8	19	30
	% within CL_NOT1	10,0%	26,7%	63,3%	100,0%
3,00	Count	0	0	3	3
	% within CL_NOT1	,0%	,0%	100,0%	100,0%
Total	Count	9	11	23	43
	% within CL_NOT1	20,9%	25,6%	53,5%	100,0%

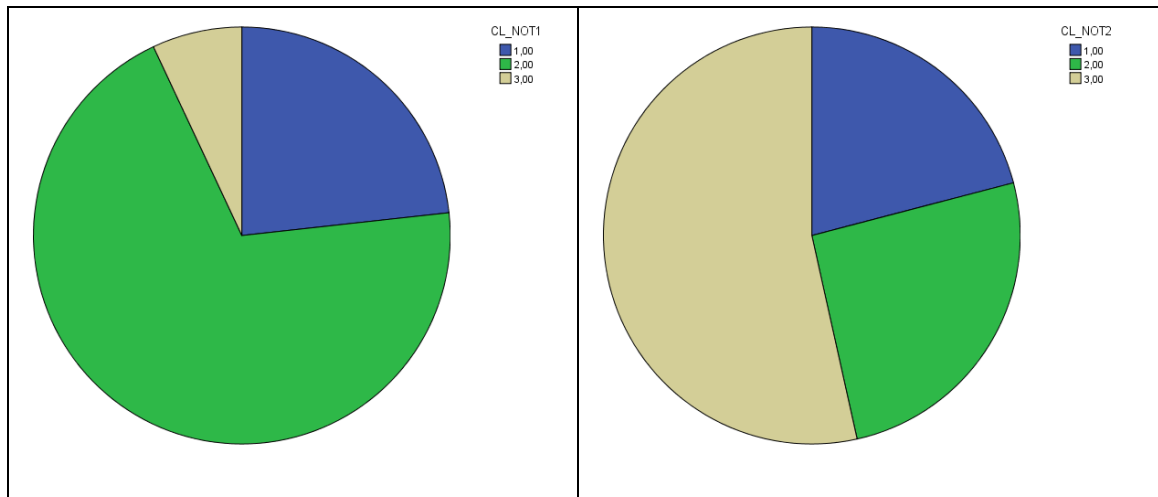


Figura 7.2 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del primer y segundo parcial.

Como puede observarse, el tercer clúster es *absorbente* (es decir, si se entra en él, no se sale de él) y ello nos produce una fila de ceros (en todas las casillas menos en una) en la matriz de probabilidades. Una matriz con demasiados ceros (muchos estados absorbentes o muchos estados incomunicados) puede hacer que nuestro método sea estéril ya que, al aplicarla sucesivamente, nos remita a un estado tan degenerado que la realidad siempre discrepe significativamente de él. Sin embargo, son habituales en este tipo de investigaciones los estados *absorbentes* puesto que, por desgracia, siempre hay grupos de abandono en las carreras de ingeniería.

En el caso que nos ocupa, que de momento no parece metodológicamente patológico, se puede obtener el siguiente Grafo de transiciones:

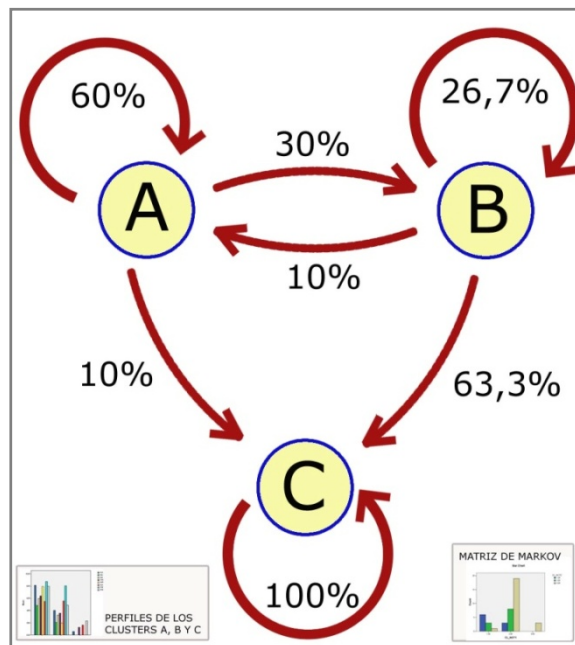


Figura 7.3 Grafo de transiciones entre los clústeres del primer parcial y los del segundo

El hecho que la transición del estado A al estado C (absorbente) sea tan solo de un 10% de probabilidad, unido al poco volumen inicial de C, nos anima a pensar que no entraremos en un modelo degenerado. La fuerte probabilidad de transición de B a C, en

cambio (63,3%), es más preocupante metodológicamente, porque al multiplicarse por sí misma podría producir resultados próximos a cero.

7.2.2 Transiciones entre el segundo y tercer parcial: Grafos y matrices de transición

Si realizamos las mismas pruebas para el paso del segundo al tercer parcial, obtenemos el siguiente diagrama de barras:

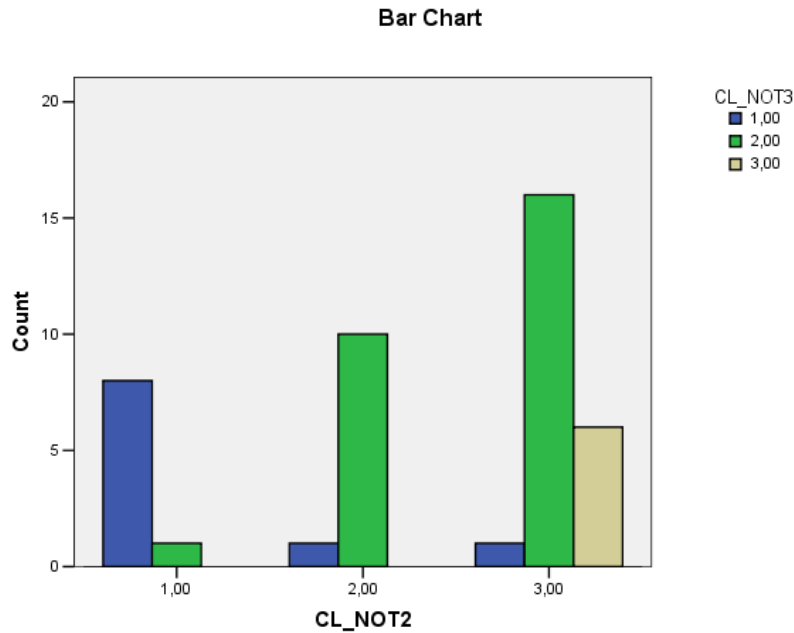


Figura 7.4 Multimedia. Diagrama de barras entre el segundo y tercer parcial

Y, nuevamente, con una prueba CROSSTABS obtenemos las probabilidades de transición:

CL_NOT2 * CL_NOT3 Crosstabulation

			CL_NOT3			Total
			1,00	2,00	3,00	
CL_NOT2	1,00	Count	8	1	0	9
		% within CL_NOT2	88,9%	11,1%	,0%	100,0%
	2,00	Count	1	10	0	11
		% within CL_NOT2	9,1%	90,9%	,0%	100,0%
	3,00	Count	1	16	6	23
		% within CL_NOT2	4,3%	69,6%	26,1%	100,0%
Total		Count	10	27	6	43
		% within CL_NOT2	23,3%	62,8%	14,0%	100,0%

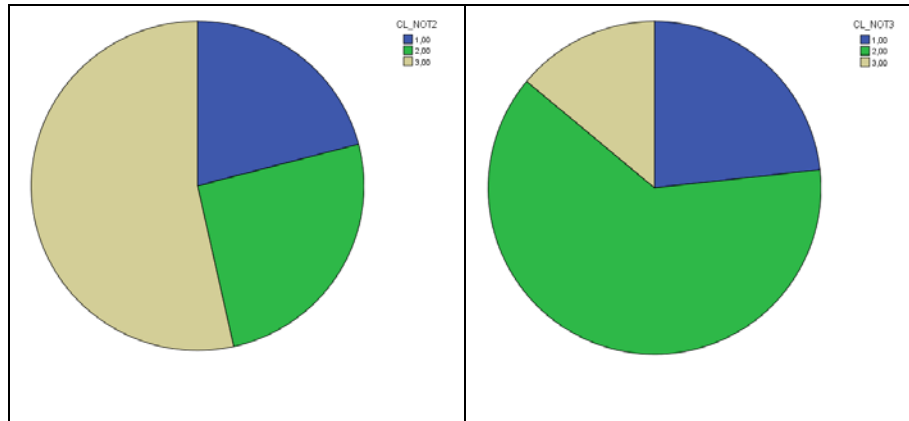


Figura 7.5 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del segundo y tercer parcial.

Como puede observarse, la situación ha cambiado porque hay una mayor “resistencia” de los estados A y B a perder alumnos, lo cual es positivo bajo el punto de vista académico.

El Grafo de transiciones para los 3 estados, queda ahora de la siguiente manera:

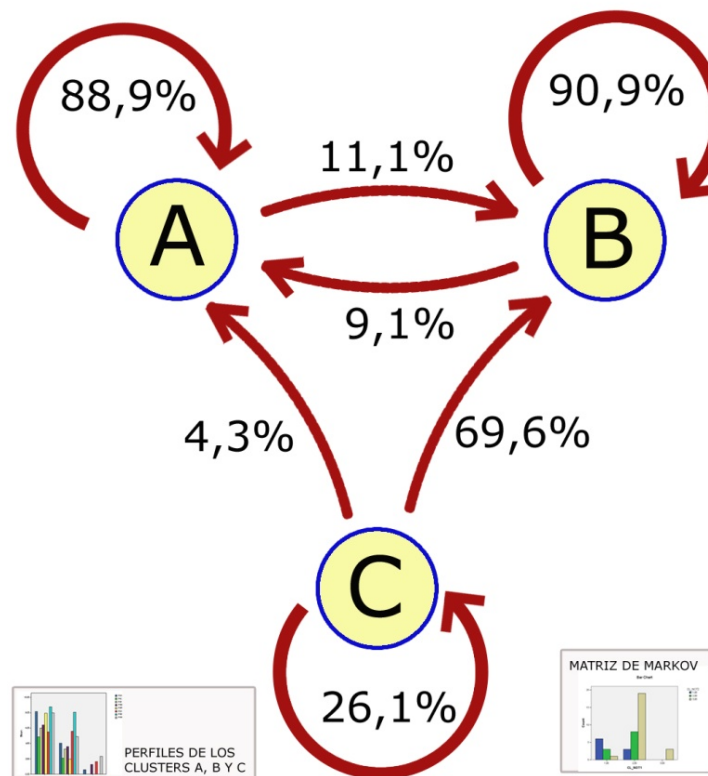


Figura 7.6 Grafo de transiciones entre los clústeres del segundo parcial y los del tercero

¿Es significativa la intervención académica del centro? La evolución de este Grafo respecto del anterior es un buen augurio. Sin embargo, para medir la intervención hace falta aplicar la hipótesis que rige nuestro modelo, que se resume en este esquema:

Sistema consistente de conglomerados

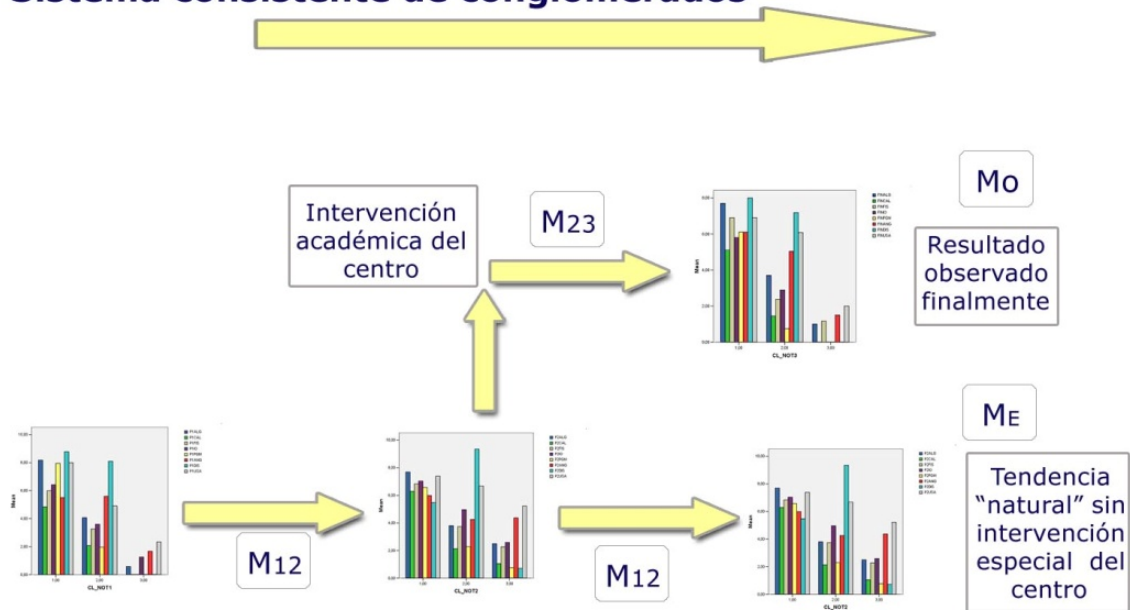


Figura 7.7 Esquema de las pruebas estadísticas para medir la intervención del centro

Es decir, suponemos que la matriz de transición entre el primer y segundo parcial, M_{12} , marca la tendencia natural del grupo. Por tanto, bajo la hipótesis de no intervención del centro, el resultado final esperado es la aplicación de dos veces M_{12} . Multiplicándola, pues, por sí misma nos llevará a un vector V_E (estado esperado) para el tercer parcial. Ahora bien, el resultado final observado, vector V_O , debido a la intervención del centro, discrepa del resultado pronosticado por la aplicación dos veces consecutivas de M_{12} .

Con este esquema, en definitiva, la intervención académica del centro queda constatada si la diferencia entre V_E y V_O es significativa, cosa que se puede comprobar en nuestro caso con una prueba Chi cuadrado.

7.2.3 Cálculo de la Matriz Esperada M_E a partir de la Matriz primero-segundo

Así pues, siguiendo el esquema de la figura 7.7, debemos calcular la matriz resultante de aplicar dos veces la primera matriz de transiciones M_{12} . El resultado, que denominamos M_E , matriz esperada, es el siguiente:

$$\begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.267 & 0.633 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.267 & 0.633 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.39 & 0.2601 & 0.34099 \\ 0.0867 & 0.101289 & 0.812011 \\ 0 & 0 & 1 \end{pmatrix}$$

Así pues, como se puede apreciar en la figura 7.8, M_E se caracteriza por tener un estado absorbente, el estado C, y una fuerte tendencia por parte de los dos otros estados, A y B, hacia este estado C (probabilidades del 34,099% y 81,2011%). Ésta es la evolución esperada de nuestro sistema consistente de clústeres si no se produjera una variación de la vida académica entre el segundo y tercer parcial.

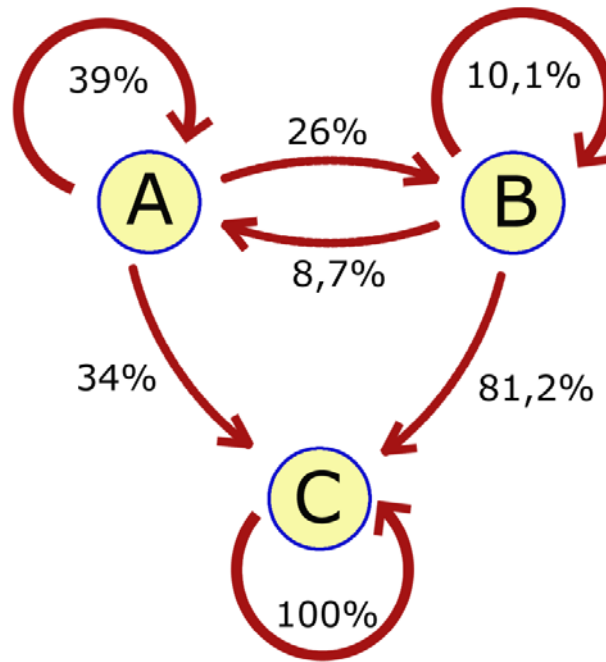


Figura 7.8 Grafo de transiciones de la Matriz esperada

7.2.4 Transiciones entre el primer y el tercer parcial: Grafos y matrices de transición

Siguiendo con nuestro método, vamos a considerar la evolución del estado inicial (primer parcial) al estado final (tercer parcial) mediante una nueva matriz de transiciones de Markov.

Con los datos que disponemos, la evolución obtenida desde el primer parcial directamente al tercero se plasma en el siguiente diagrama de barras:

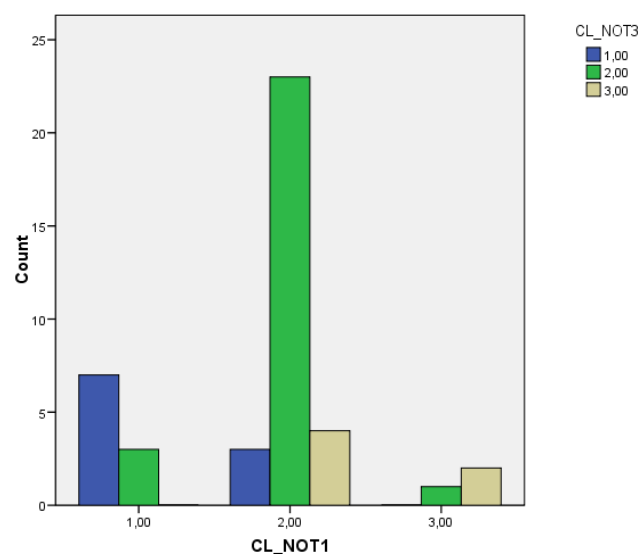


Figura 7.9 Multimedia. Diagrama de barras entre el primero y tercer parcial

Para obtener las probabilidades de transición realizamos una prueba CROSSTABS y obtenemos:

CL_NOT1 * CL_NOT3 Crosstabulation

			CL_NOT3			Total
			1,00	2,00	3,00	
CL_NOT1	1,00	Count	7	3	0	10
		% within CL_NOT1	70,0%	30,0%	,0%	100,0%
	2,00	Count	3	23	4	30
		% within CL_NOT1	10,0%	76,7%	13,3%	100,0%
	3,00	Count	0	1	2	3
		% within CL_NOT1	,0%	33,3%	66,7%	100,0%
Total		Count	10	27	6	43
		% within CL_NOT1	23,3%	62,8%	14,0%	100,0%

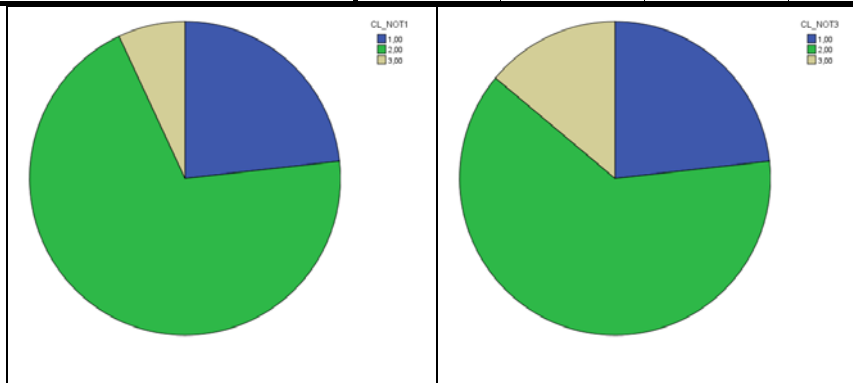


Figura 7.10 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del primer y tercer parcial.

Por lo que obtenemos una matriz de transiciones de Markov, que denominamos M_O , matriz obtenida, que nos marca la transición del estado inicial al final a la luz de los datos del sistema consistente de clústeres.

$$M_{13} = \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.1 & 0.767 & 0.133 \\ 0 & 0.333 & 0.667 \end{pmatrix}$$

Hay que notar que esta matriz nos mide la tendencia académica del centro, tomando como unidad el año académico y partiendo de un esquema consistente de clústeres. Ello nos abre las puertas a otros tipos de trabajo, en los que quisiéramos comparar centros diferentes, cursos diferentes, o construir cadenas infinitas de Markov. Sin embargo, como veremos a continuación, en el caso que nos ocupa la usamos para plantear la medición de la intervención académica del centro.

7.2.5 Intervención del Centro: Planteamiento estadístico del problema

7.2.5.1 Abandono del primer enfoque de comparación de matrices

A efectos experimentales, nuestro método termina con la comparación estadística de dos matrices, M_E y M_O , es decir, matriz esperada siguiendo la tendencia “natural” de los alumnos y matriz observada debido a que se produce una intervención del centro.

$$M_E \begin{pmatrix} 0.39 & 0.2601 & 0.34099 \\ 0.0867 & 0.101289 & 0.812011 \\ 0 & 0 & 1 \end{pmatrix} \quad M_O \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.10 & 0.767 & 0.133 \\ 0 & 0.333 & 0.667 \end{pmatrix}$$

La hipótesis nula en este punto de nuestra investigación es

$$H_0: M_E = M_O$$

lo cual nos sirve para medir la diferencia entre la “producción académica” sin intervención del centro o con intervención del mismo. En ciertas investigaciones va a interesar esta diferencia entre “producciones”, para lo cual se deberá echar mano de pruebas estadísticas de comparación de matrices bajo distribuciones libres o conocidas.

Sin embargo, en nuestra investigación nos interesa más la discrepancia en sí que no las matrices de evolución, por lo que la prueba de comparación puede reducirse a una prueba alternativa sobre los resultados finales esperados y obtenidos.

Dicho cambio de hipótesis nula (a partir de ahora tendremos la hipótesis estándar de prueba Chi cuadrado) nos evita el caso de patologías en aquellos tratamientos estadísticos donde se requieran frecuencias superiores al 5% en todas las casillas. En las matrices de transiciones de Markov son frecuentes los ceros, puesto que aparecen en la realidad muchos estados “incomunicados”; en la metodología de sistemas de clústeres también se dan, puesto que son inevitables los grupos de abandono (exponente más frecuente de estados absorbentes). Así pues, el cambio de hipótesis nos permite, al disponer de casillas con frecuencias finales no nulas, un tratamiento estadístico más cómodo en las diferentes investigaciones.

El mismo programa estadístico SPSS, aplicado a la comparación de matrices, nos informa que éste es un caso patológico de Chi cuadrado, porque hay dos celdas con frecuencias inferiores al 5%. Por tanto, seguiremos nuestro método usando el cambio de enfoque estadístico. Sustituiremos la hipótesis nula de comparación de matrices por una prueba de comparación de estados finales.

7.2.5.2 Adopción del nuevo enfoque de comparación de estados finales

Para medir la intervención del centro en el caso de los sistemas consistentes de clústeres y aplicando cadenas de Markov, nuestro método pasa, a la luz de lo que hemos visto, por realizar una prueba de bondad de ajuste con el test Chi cuadrado (X^2). La hipótesis nula, por tanto, es:

$$H_0: f_E = f_O$$

Donde se ha seguido la notación habitual de frecuencias observadas y esperadas:

$$X^2 = \sum_{i=1}^n \frac{(f_e - f_o)^2}{f_e}$$

$f_e = \text{frecuencia esperada}$
 $f_o = \text{frecuencia observada}$

La prueba Chi cuadrado es una prueba no paramétrica que mide la discrepancia entre una distribución observada y otra teórica (bondad de ajuste), indicando en qué medida las diferencias existentes entre ambas, de haberlas, se deben al azar en el contraste de hipótesis. También se utiliza para probar la independencia de dos variables entre sí, mediante la presentación de los datos en tablas de contingencia.

En nuestro caso comparamos dos vectores de 3 componentes, V_E y V_O , correspondientes al número de alumnos que haya en cada uno de los tres clústeres del sistema consistente, según el estado final pronosticado por las matrices de transición y el estado final observado.

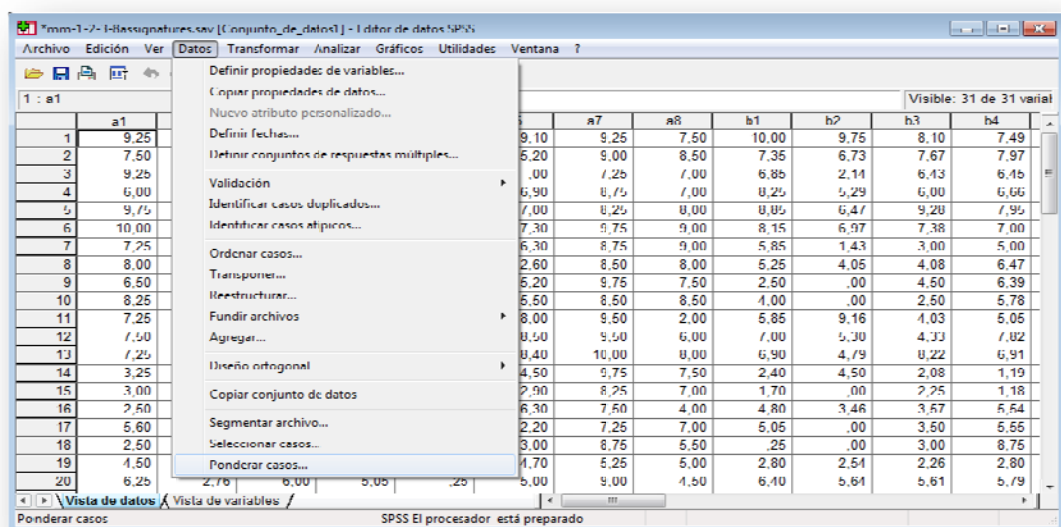
Para realizar la prueba Chi cuadrado, debemos señalar dos apuntes metodológicos:

a) Patologías en las frecuencias de aparición de casos

En la metodología de los sistemas consistentes de clústeres pueden aparecer ceros en las matrices de transición puesto que se dan en la realidad muchos estados “incomunicados” (como es nuestro estado C del primer parcial, véase figura 7.8). En el caso de aplicar esta técnica a estudios superiores, hay que considerar los inevitables grupos de abandono, que son el exponente más frecuente de estos estados absorbentes. Todo ello nos puede llevar a la aparición de vectores de estado con alguna casilla de frecuencia inferior al 5%, lo cual exige realizar correcciones a la prueba Chi cuadrado. Sin embargo, en nuestro caso hemos tenido la fortuna de obtener ambos vectores de estado, V_E y V_O , con frecuencias mayores del 5% en todas las casillas, por lo que el tratamiento estadístico de reduce en este punto a una prueba Chi cuadrado ordinaria.

b) Diferente inserción de datos en el programa informático de tratamiento estadístico

Como segundo apunte metodológico sólo nos cabe señalar aquí que, para realizar pruebas Chi cuadrado con el paquete SPSS, se tiene que usar la opción WEIGHT CASES (Ponderar casos) para poder tratar la columna de frecuencias como tal (es decir, como datos relativos a frecuencias y no a un individuo). Se adjuntan unas capturas de pantalla explicativas a continuación:



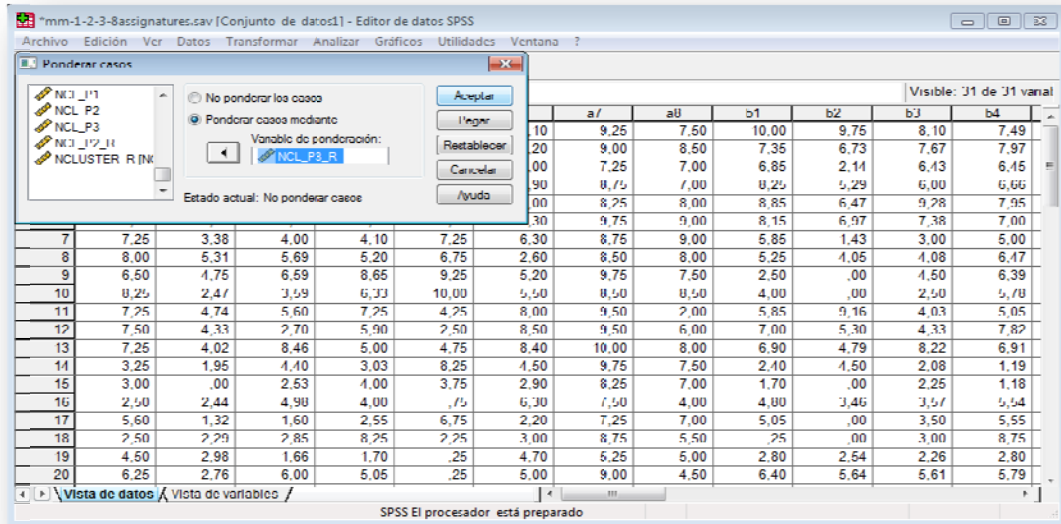


Figura 7.11 Capturas de pantalla de la opción *WEIGHT CASES* del paquete SPSS

7.2.6 Resultados de la prueba de comparación

7.2.6.1 Tratamiento estadístico

Hemos optado por un enfoque estadístico que se centra en el *Estado final pronosticado*. El sentido conceptual es importante, ya que resume la totalidad del experimento, es decir, la desviación producida por la intervención del centro sobre un resultado académico esperado.

Bajo tal punto de vista, tenemos que el estado inicial es:

Clúster	Individuos
A	10
B	30
C	3

En el apartado 7.2.3 hemos obtenido la matriz esperada M_E resultante de aplicar dos veces la primera matriz de transiciones M_{12} :

Así pues, M_E se caracteriza por tener un estado absorbente, el estado C, y una fuerte tendencia por parte de los dos otros estados, A y B, hacia este estado C (probabilidades del 34,099% y 81,2011%). Ésta es la evolución esperada de nuestro sistema consistente de clústeres si no se produjera una variación de la vida académica entre el segundo y tercer parcial.

Y, por tanto, para obtener las frecuencias esperadas, debemos aplicar la matriz M_E al vector unidimensional del estado inicial, es decir:

$$(10 \ 30 \ 3) \times \begin{pmatrix} 0.39 & 0.2601 & 0.34099 \\ 0.0867 & 0.101289 & 0.812011 \\ 0 & 0 & 1 \end{pmatrix}$$

Con lo que obtenemos el vector esperado V_E que nos pronostica el estado final de nuestro sistema consistente de clústeres:

Clúster	Individuos
A	6,501
B	5,63967
C	30,77023

Esta tabla será usada para contrastar con los resultados obtenidos en el experimento (es decir, debidos a la intervención del centro), que son los correspondientes al estado final observado V_O :

Clúster	Individuos
A	10
B	27
C	6

Así las cosas, ya podemos aplicar la prueba de Chi cuadrado para medir la discrepancia y pronunciarnos sobre su significación. Realizada la prueba en SPSS, obtenemos:

CLÚSTERDESTI			
	Observed N	Expected N	Residual
1,00	10	6,5	3,5
2,00	27	5,7	21,3
3,00	6	30,8	-24,8
Total	43		

Test Statistics	
	CLÚSTERDESTI
Chi-Square(a)	102,505
df	2
Asymp. Sig.	,000

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 5,7.

Figura 7.12 Resultado obtenido con SPSS de la prueba de Chi cuadrado

7.2.6.2 Conclusión

Obtenemos, por tanto, un resultado significativo de la prueba, sin patologías estadísticas debidas a celdas con frecuencias pequeñas.

Por tanto, por lo que se refiere a los alumnos de ingeniería multimedia, podemos decir que hemos llegado a un momento importante, como resultado del método propuesto en esta tesis, ya que:

- a) Hemos conseguido elaborar un sistema consistente de clústeres que explica la conducta académica, mientras que las pruebas correlacionales no echaban demasiada luz sobre el fenómeno estudiado.
- b) El modelo de encaje para la ingeniería multimedia es de un sistema consistente biunívoco de 3 clústeres, con las características que se han descrito en los capítulos anteriores, y que sirve de guía para la acción tutorial y, en definitiva, para mejorar cualquier tipo de intervención del centro.
- c) El modelo nos descubre, además, que tal como se están haciendo ahora las cosas se consigue una aportación del centro que provoca una desviación positiva y significativa del rendimiento académico.
- d) Las diferencias entre el modelo de encaje de los alumnos de multimedia y el que se obtenga en los alumnos de ingenierías tecnológicas serán un resultado también importante de esta investigación, porque no sólo informan sobre rendimientos diferentes en ciertas asignaturas sino también, y aquí está lo fundamental, entran a mostrarnos y a profundizar en la diferente “naturaleza académica” de los dos tipos de estudiantes.

Ahora sólo nos queda, para concluir este capítulo central de la investigación, aplicar el mismo método a los estudiantes de ingenierías tecnológicas. Veremos, sin embargo, como nos encontramos con un terreno mucho más irregular que no puede ser ajustado de entrada a un sistema de 3 clústeres biunívoco. Las dificultades que aparecerán a continuación, sin embargo, van a ser altamente ilustrativas tanto de la potencia del método como de su belleza heurística.

7.3 Evolución de los clústeres en ingenierías tecnológicas. Primer intento de ajuste.

Tal como hemos anticipado, el sistema de clústeres para ingenierías tecnológicas no va a resultar de manera tan inmediata como en el caso de las ingenierías multimedia. Como se trata de un caso más complicado de ajuste, vamos a reproducir los pasos anteriores pero con un mayor detalle del procedimiento, a fin de ilustrar los recursos heurísticos del método de investigación que se expone en esta tesis.

7.3.1 Transiciones entre el primer y segundo parcial: Grafos y matrices de transición

De los datos suministrados en los capítulos 5 y 6, observamos de entrada que no tenemos un sistema biunívoco. Ello se debe a que los clústeres del segundo parcial están

sometidos a una presión general de radicalización (estudiantes que se sostienen vs estudiantes que fracasan).

La información que debemos tener en mente para seguir el método es la que se plasma en esta figura:

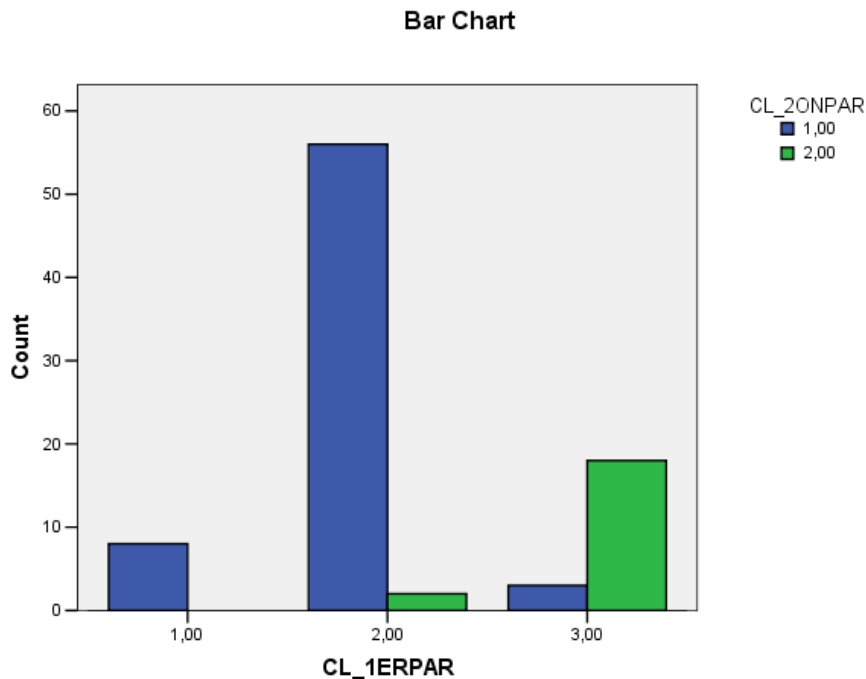


Figura 7.13 Diagrama de barras de transiciones entre los 3 clústeres del primer parcial y los 2 del segundo

Una ojeada a los volúmenes de los clústeres (columna derecha) nos muestra esta presión a la que nos referíamos, puesto que el clúster central del primer parcial (color verde) desaparece en el segundo parcial.

Del mismo modo que hacíamos con los alumnos de multimedia, hemos obtenido las probabilidades de transición realizando una prueba CROSSTABS:

CL_1ERPAR * CL_2ONPAR Crosstabulation

			CL_2ONPAR		Total
			1,00	2,00	
CL_1ERPAR	1,00	Count	8	0	8
		% within CL_1ERPAR	100,0%	,0%	100,0%
	2,00	Count	56	2	58
		% within CL_1ERPAR	96,6%	3,4%	100,0%
	3,00	Count	3	18	21
		% within CL_1ERPAR	14,3%	85,7%	100,0%
Total		Count	67	20	87
		% within CL_1ERPAR	77,0%	23,0%	100,0%

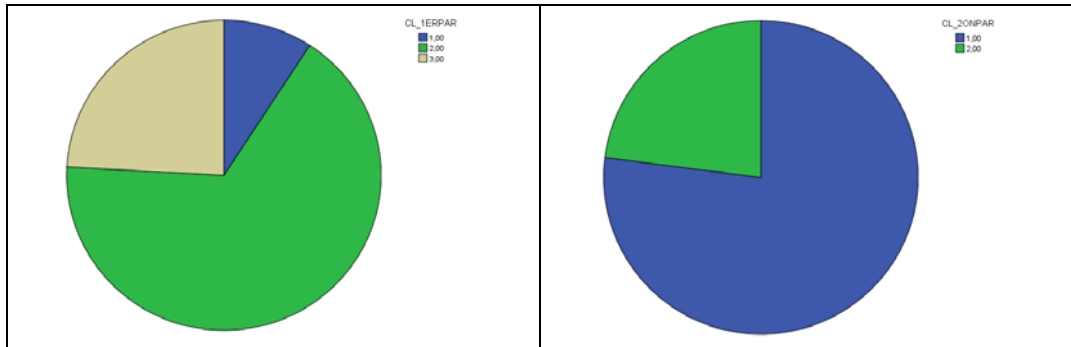


Figura 7.14 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del primer y segundo parcial.

Resulta, por tanto, un Grafo de transiciones entre 3 estados distinguibles iniciales y sólo 2 finales:

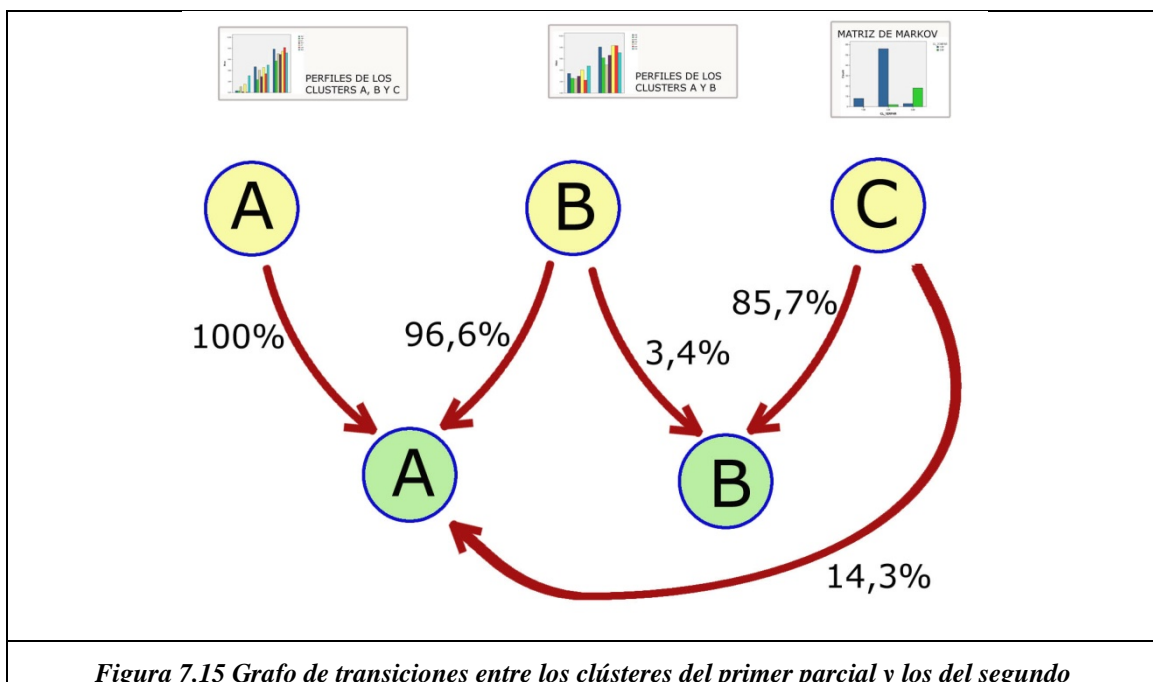


Figura 7.15 Grafo de transiciones entre los clústeres del primer parcial y los del segundo

Y disponemos ya de la matriz M_E de transiciones entre el primer y segundo parcial. Se trata de una matriz 2 x 3 porque en el segundo parcial ha desaparecido un clúster:

$$\begin{pmatrix} 1 & 0 \\ 0.966 & 0.034 \\ 0.143 & 0.857 \end{pmatrix}$$

Llama la atención que en este caso el estado A (alumnos de rendimiento bajo) sea absorbente, al igual que en multimedia lo era en estado C (alumnos de poco o escaso rendimiento, en peligro de abandono).

El hecho de aplicar el método con matrices no cuadradas (número variable de clústeres debido a que el sistema no es biunívoco) no es metodológicamente ningún escollo, por lo que procedemos a seguir con los cálculos previstos en busca de la discrepancia debida a la intervención del centro en el estado final.

7.3.2 Transiciones entre el segundo y tercer parcial: Grafos y matrices de transición

Si observamos los diagramas resultantes, vemos que se vuelve a una situación no radicalizada en dos clústeres, tal como lo plasma la siguiente figura:

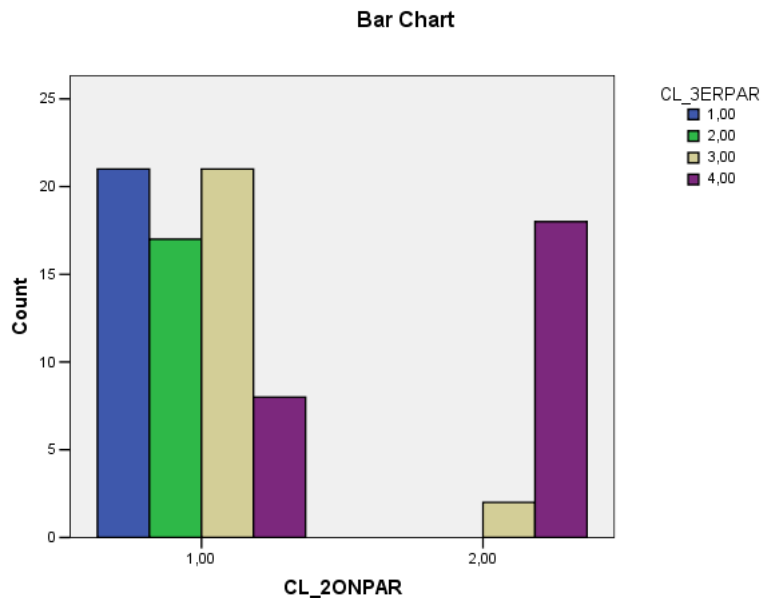


Figura 7.16 Diagrama de barras de transiciones entre los 2 clústeres del segundo y los 4 del tercero

Es obvio, por tanto, que se ha producido una variación en la tendencia natural (debida a la intervención del centro). Al no tener entre manos un sistema consistente biunívoco se hace ahora más difícil visualizar la aportación del centro. El hecho que aparezcan 4 clústeres, además, no simplifica en ningún modo las cosas.

Sin embargo, tal como se ha discutido en el apartado relativo a los estudiantes multimedia, nuestro método compara los estados finales, lo que nos permite continuar con el procedimiento estadístico en busca de la discrepancia entre lo esperado y lo obtenido.

A efectos descriptivos, podemos realizar la prueba CROSSTABS y así obtener las probabilidades de transición:

CL_2ONPAR * CL_3ERPAR Crosstabulation

			CL_3ERPAR				Total
			1,00	2,00	3,00	4,00	
CL_2ONPAR	1,00	Count	21	17	21	8	67
		% within CL_2ONPAR	31,3%	25,4%	31,3%	11,9%	100,0%
	2,00	Count	0	0	2	18	20
		% within CL_2ONPAR	,0%	,0%	10,0%	90,0%	100,0%
Total		Count	21	17	23	26	87
		% within CL_2ONPAR	24,1%	19,5%	26,4%	29,9%	100,0%

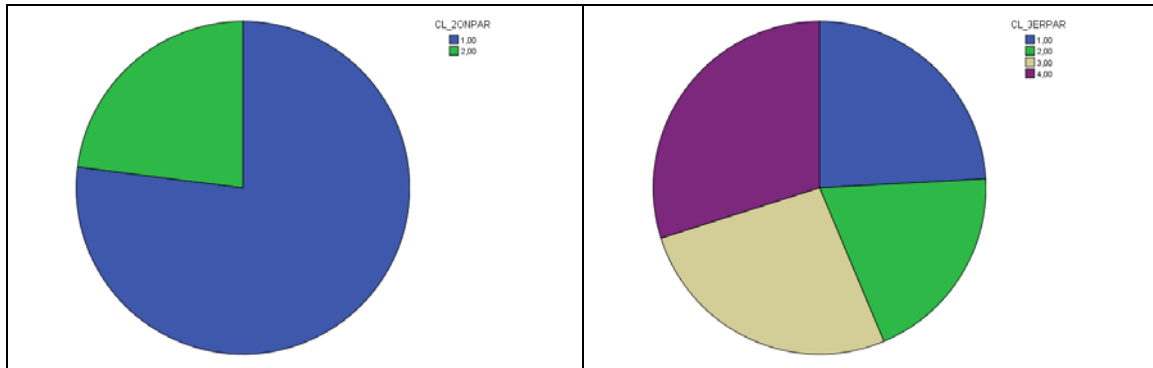


Figura 7.17 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del segundo y tercer parcial.

De lo que extraemos el siguiente Grafo de Markov de transiciones:

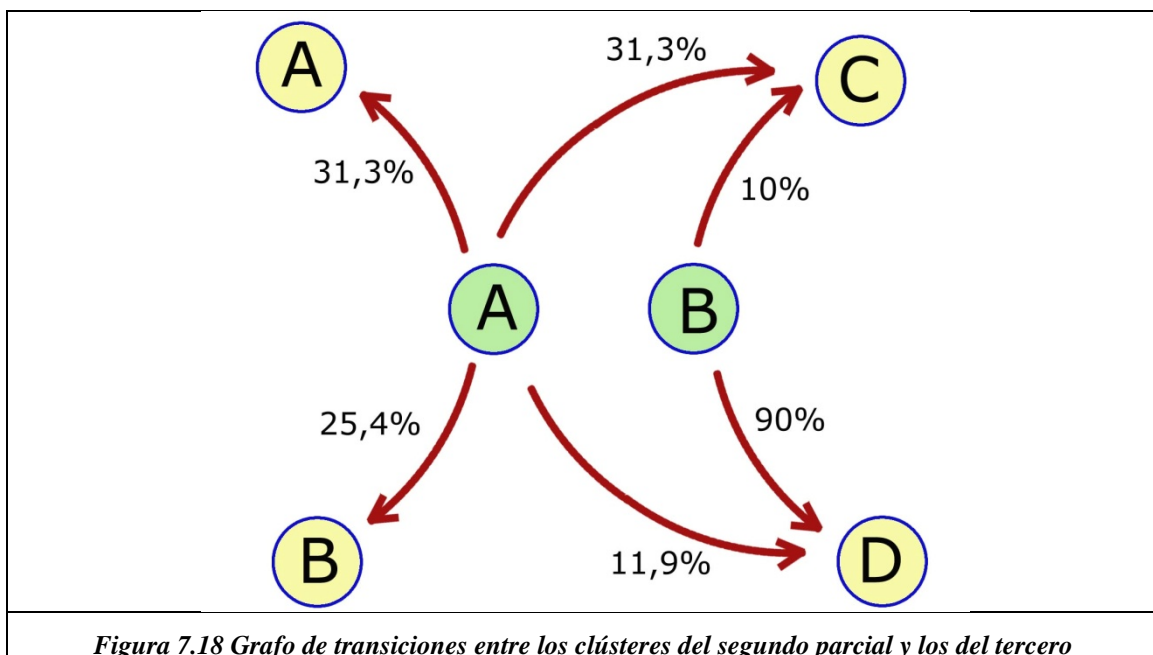


Figura 7.18 Grafo de transiciones entre los clústeres del segundo parcial y los del tercero

Una observación simple al grafo nos revela que desde el clúster B del segundo parcial (color verde) no se sale hacia A o B del tercer parcial (color amarillo), que no sería deseable, por ser grupos de mejor rendimiento. Por tanto, los buenos alumnos se mantienen buenos y solo un 10% empeora un poco.

No obstante, a efectos de poder evaluar la intervención del centro, seguimos con el método propuesto porque de momento no nos hemos encontrado con ninguna patología estadística.

7.3.3 Cálculo de la Matriz esperada ME a partir de la Matriz primero-segundo

Usaremos la notación de 3 clústeres para poder aplicar 2 veces la matriz de transiciones del primer al segundo parcial. Esta matriz, recordémoslo, es una matriz de 2 x 3. La inmersión de la matriz 2 x 3 en una 3 x 3 es un recurso formal para poder multiplicar y deducir, de este modo, el estado final. Se hubieran podido calcular manualmente las probabilidades finales, sin recurrir a las matrices, pero preferimos reflejar el

procedimiento estándar porque en casos de sistemas con mayor número de clústeres los cálculos son más complicados.

De este modo, la matriz M_E resultante se obtiene de la aplicación por 2 veces de la matriz de transición entre el primer y segundo parcial, es decir:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0.966 & 0.034 & 0 \\ 0.143 & 0.857 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0.966 & 0.034 & 0 \\ 0.143 & 0.857 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0.998844 & 0.001156 & 0 \\ 0.970862 & 0.029138 & 0 \end{pmatrix}$$

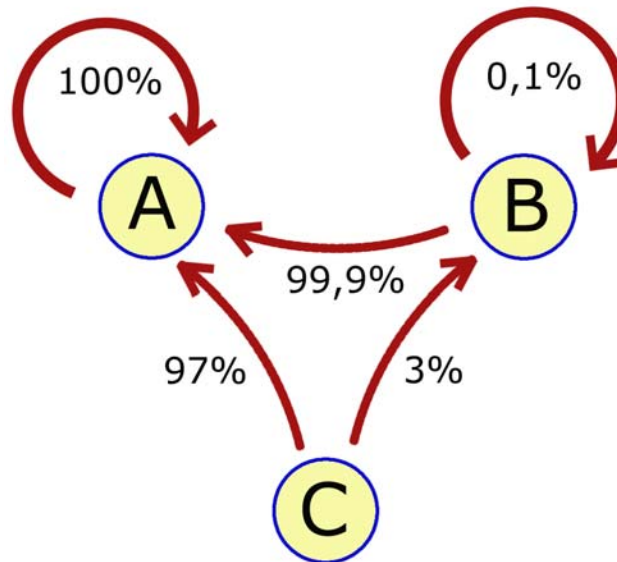


Figura 7.19 Grafo de transiciones de la matriz esperada.

7.3.4. Transiciones entre el primer y tercer parcial: Grafos y matrices de transición

La evolución obtenida desde el primer parcial directamente al tercero ha sido la que refleja el siguiente diagrama de barras:

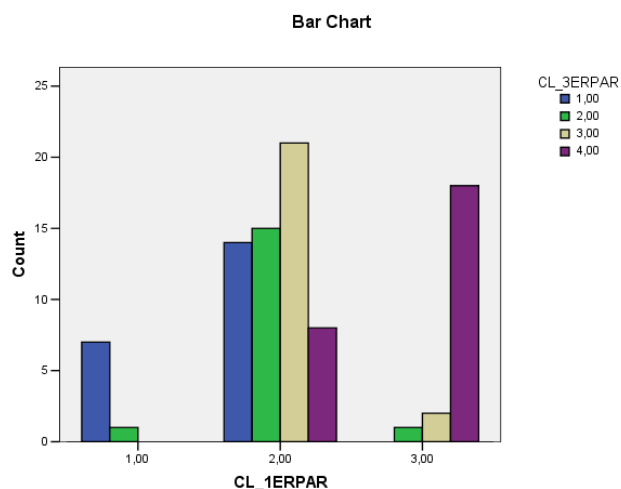


Figura 7.20 Diagrama de barras de transiciones entre los clústeres del primero al tercero

Realizamos la prueba CROSSTABS, obteniendo:

CL_1ERPAR * CL_3ERPAR Crosstabulation

			CL_3ERPAR				Total
			1,00	2,00	3,00	4,00	
CL_1ERP AR	1,00	Count	7	1	0	0	8
		% within CL_1ERPAR	87,5%	12,5%	,0%	,0%	100,0%
	2,00	Count	14	15	21	8	58
		% within CL_1ERPAR	24,1%	25,9%	36,2%	13,8%	100,0%
	3,00	Count	0	1	2	18	21
		% within CL_1ERPAR	,0%	4,8%	9,5%	85,7%	100,0%
Total		Count	21	17	23	26	87
		% within CL_1ERPAR	24,1%	19,5%	26,4%	29,9%	100,0%

Por lo que la matriz final de Markov, M_O , es una 3 x 4:

$$\begin{pmatrix} 0.875 & 0.125 & 0 & 0 \\ 0.241 & 0.259 & 0.362 & 0.138 \\ 0 & 0.048 & 0.095 & 0.857 \end{pmatrix}$$

Omitimos los cálculos al lector porque, como puede ver, la discrepancia entre M_E y M_O nos lleva a una patología conceptual y estadística. El cálculo de los estados finales según M_E nos llevará a un vector de estado con 2 ceros como mínimo (estados inexistentes) que, además, son una patología con la Chi cuadrado porque tenemos frecuencias menores del 5%. Conceptualmente la patología también es obvia: una matriz que pronostica la desaparición de estados (de 3 a 2) no es comparable con una que pronostica la aparición de estados (de 3 a 4).

De este modo, lo que nos viene a decir el método es que hay, efectivamente, un gran cambio en la dinámica del grupo entre una transición y otra, cosa que ya observábamos con las gráficas descriptivas del sistema consistente de clústeres.

Hemos llegado, por tanto, a una patología en el método que nos obliga a redefinir el sistema consistente de clústeres. Aunque sea un abandono del método propuesto, vamos a probar 3 vías de solución para encontrar un modelo que se adapte a nuestro caso de estudio:

- Dentro del sistema consistente no biunívoco de clústeres, realizar el tratamiento estadístico para una agrupación de clústeres (en particular, de los 4 del tercer parcial). Es decir, estadísticamente, agrupamos clústeres que tienen un perfil parecido, buenos y malos.
- Renunciar al sistema consistente no biunívoco y forzar un sistema biunívoco de 3 clústeres mediante un algoritmo de agrupación K-MEANS (lo cual implica tomar como conjunto guía los clústeres del primer parcial).
- Renunciar igualmente al sistema consistente no biunívoco, pero forzar un sistema biunívoco de 4 clústeres mediante un algoritmo de agrupación K-MEANS (lo cual implica tomar como conjunto guía los 4 clústeres del tercer parcial).

Cualquiera de estas tres vías puede dar un buen resultado o, por el contrario, suponer trabajo en balde. La única decisión heurística que se ha tomado en este momento ha sido

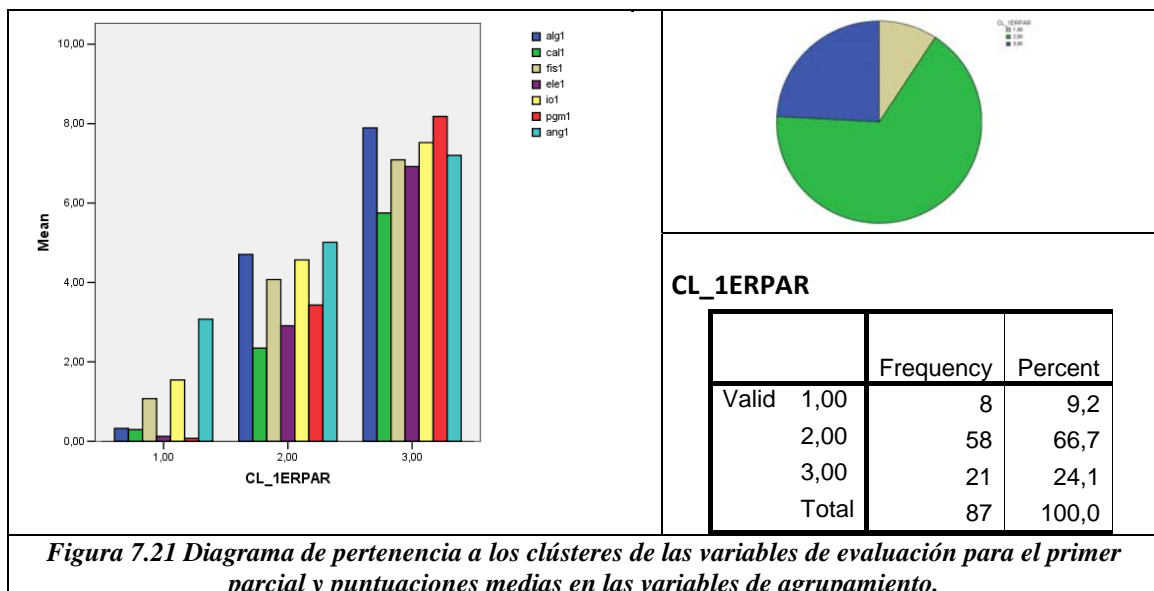
descartar la opción “sistema biunívoco de 2 clústeres” porque tiene mucha similitud con la solución a). Es decir, dado que *probablemente* se conseguiría un resultado similar al del apartado a), porque se basa en una reducción a 2 clústeres del parcial más discrepante (el cuarto) se omite este camino y se opta por explorar las reducciones a 3 y 4 clústeres.

Estas tres vías son las que se desarrollarán en los apartados siguientes y nos servirán para ilustrar cómo se aplica el método expuesto en esta tesis de manera no ortodoxa, en los casos en que es difícil encontrar un buen modelo que ajuste a la realidad. Debe tenerse en cuenta que se ha abandonado el algoritmo jerárquico para tratar ahora de forzar un número determinado de clústeres con las pruebas K-MEANS.

7.4 Desarrollo de la reducción a 2 clústeres en el tercer parcial para las ingenierías tecnológicas

Es obvio que un investigador tantee en primer lugar con esta opción porque la patología metodológica que padecemos en este caso se deriva de enfrentar unos pocos clústeres (los 2 del segundo parcial) con demasiados (los 4 del segundo). En tal situación, es lógico preguntarse si cabe una reinterpretación de los 4 clústeres que nos permita, conceptualmente, agruparlos en 2, a fin de poder hacer enfrentarlos a los 2 del segundo parcial.

Dado que se trata de una operación conceptual (no estrictamente matemática) el sistema consistente de clústeres debe ser mirado en su totalidad. Para ello, debemos considerar los clústeres de los estudiantes de ingenierías tecnológicas tal como los hemos construido en el capítulo 5:



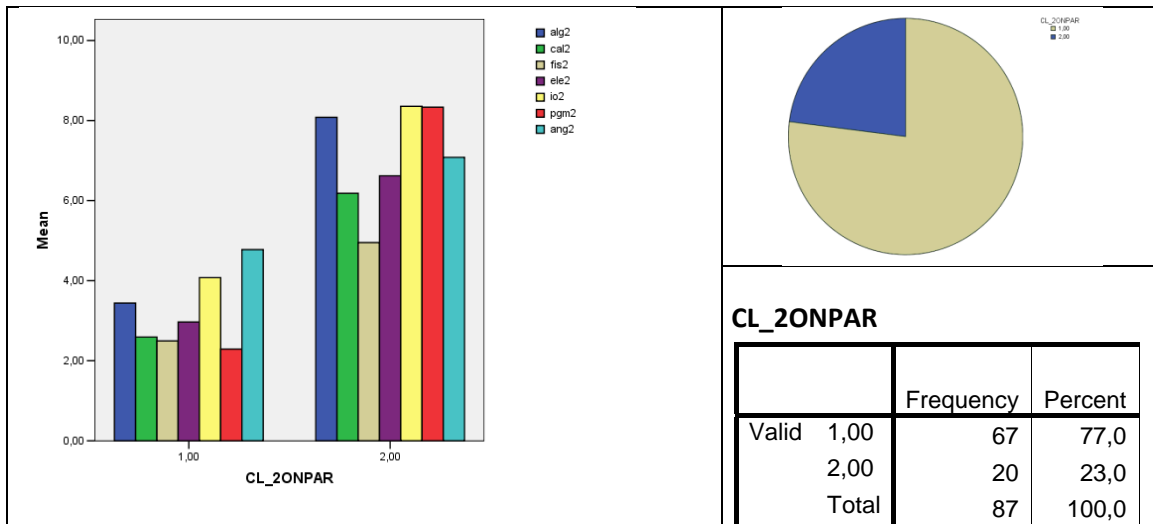


Figura 7.22 Diagrama de pertenencia a los clústeres de las variables de evaluación para el segundo parcial y puntuaciones medias en las variables de agrupamiento.

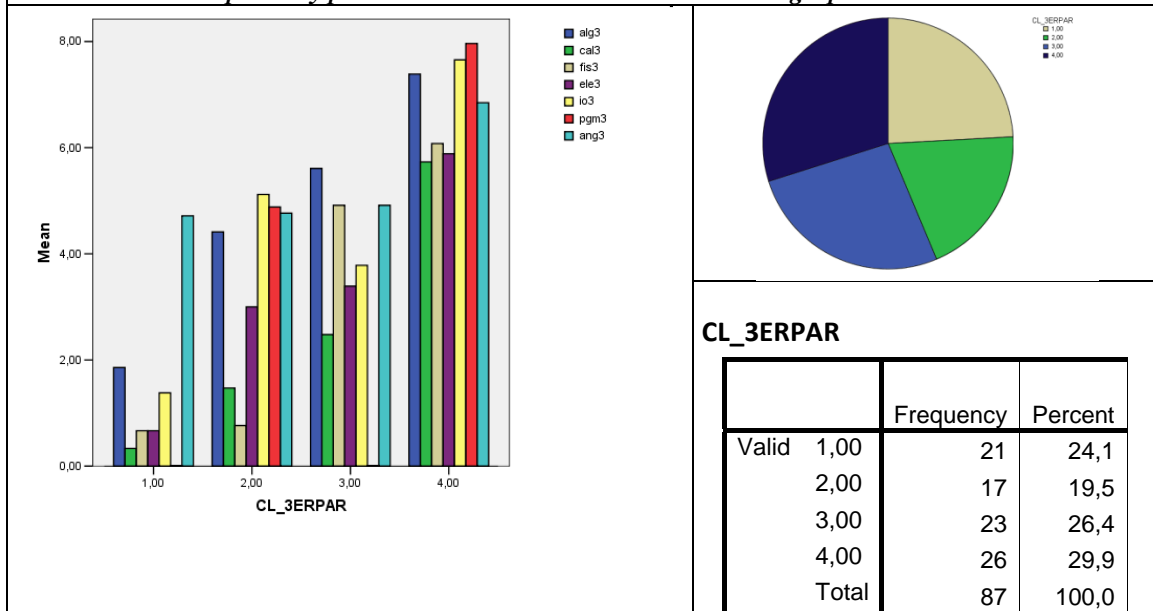


Figura 7.23 Diagrama de pertenencia a los clústeres de las variables de evaluación para el tercer parcial y puntuaciones medias en las variables de agrupamiento.

Reestructuración de los clústeres del tercer parcial

Si consideramos la figura 7.23 de clústeres de rendimiento en el tercer parcial, es razonable pensar que los alumnos de los clústeres 2 y 3 entran en un grupo de alumnos con futuro. Es decir, el diagnóstico podría resumirse hablando de *alumnos aplicados* y *alumnos fundamentales*.

Para los *aplicados*, clúster 2, tenemos la descripción siguiente:

- Aprueban álgebra, introducción a los ordenadores, programación e inglés
- Sacan notas bajas en cálculo, física y electrónica

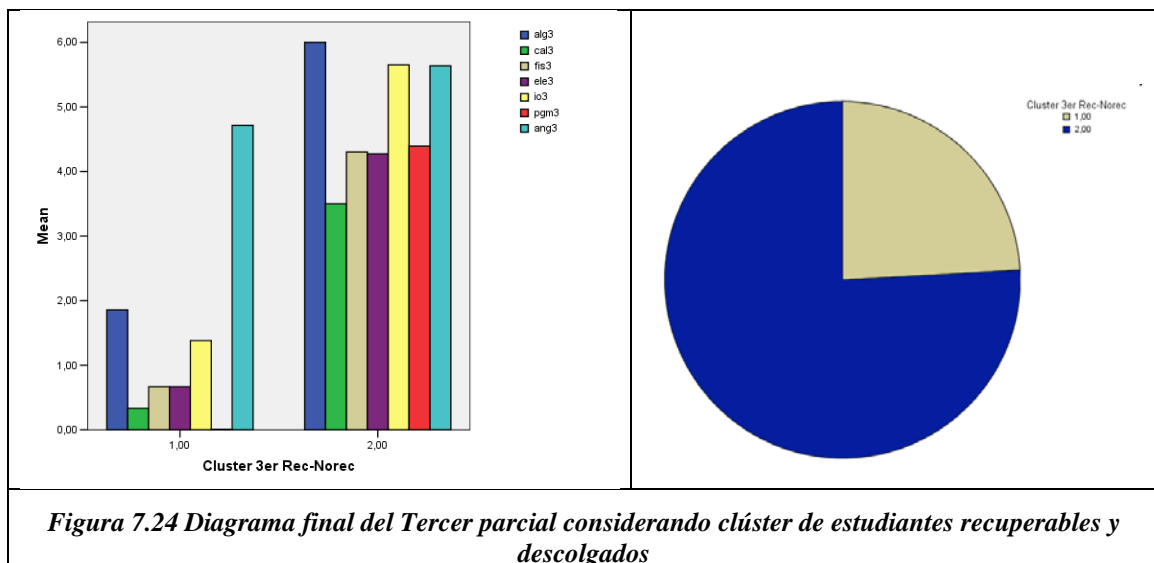
Para los *fundamentales*, clúster 3, tenemos la descripción siguiente:

- Aprueban álgebra, física e inglés

- b) Han dejado la programación
- c) Sacan notas bajas en cálculo, electrónica e introducción a los ordenadores

Una interpretación optimista nos indica que estos alumnos de los clústeres 2 y 3 son estudiantes que recuperan en verano la parte que les falta (o que se han dejado para poder aprobar el curso académico) y acaban pasando de curso con pocas asignaturas pendientes.

Bajo este punto de vista, el cuadro general de transiciones nos permite hablar de *alumnos recuperables* y *alumnos descolgados* en las ingenierías tecnológicas. Por lo que la evolución del sistema consistente de clústeres se reduce a comprobar el estado final de estos dos grandes clústeres.

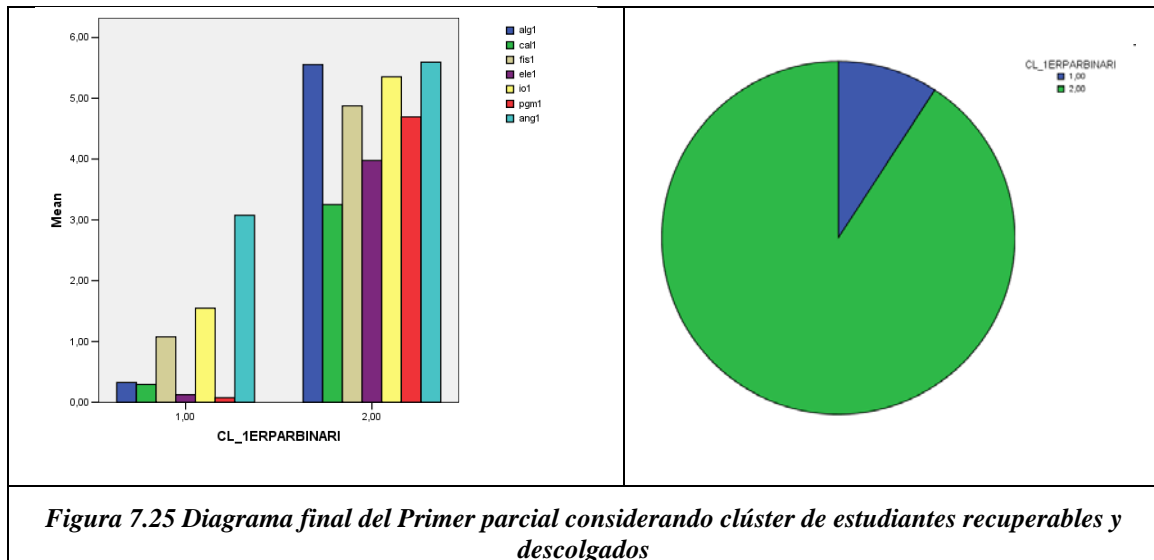


Clúster 3er Rec-Norec					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	21	24,1	24,1	24,1
	2,00	66	75,9	75,9	100,0
	Total	87	100,0	100,0	

Veamos, por ejemplo, un dato de muestra que avala esta división:

- a) El clúster número 1 tiene unas medias inferiores a 5 incluso para la asignatura de inglés, por lo que casa adecuadamente con la idea de alumnos que difícilmente podrán aprobar.
- b) El clúster número 2 es más heterogéneo (no olvidemos que, en realidad, procede de unificar 3 clústeres) pero la proximidad de las notas a 5 recoge la idea de alumnos que podrán aprobar.

Si consideramos ahora el mismo criterio para ser aplicado a los clústeres del primer parcial, tenemos que los nuevos clústeres son:



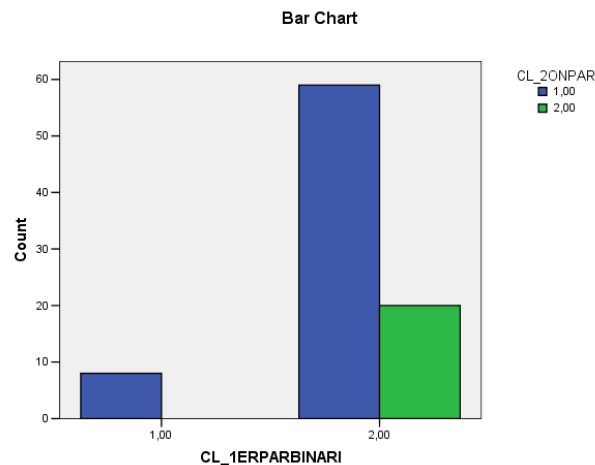
Para el procesamiento de datos en SPSS hemos creado una nueva variable: CL_1ERPARBINARI. Con esta expresión queremos indicar que hemos forzado la muestra para obtener sólo 2 valores de pertenencia a clústeres. Con tal suposición, la distribución de individuos de la muestra queda de la siguiente manera:

CL_1ERPARBINARI

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1,00	8	9,2	9,2	9,2
2,00	79	90,8	90,8	100,0
Total	87	100,0	100,0	

7.4.1 Diagrama resultante de evolución de los clústeres

Tenemos que proceder ahora de la misma manera que en los otros casos para llegar a una matriz de transiciones. Lo primero que haremos es observar la tabla de las evoluciones de los clústeres, ahora simplificada a 2 estados:



Posteriormente, siguiendo el método, realizaremos un CROSSTABS para obtener las probabilidades de transición:

CL_1ERPARBINARI * CL_2ONPAR Crosstabulation

			CL_2ONPAR		Total
			1,00	2,00	
CL_1ERPARBINARI	1,00	Count	8	0	8
		% within CL_1ERPARBINARI	100,0%	,0%	100,0%
	2,00	Count	59	20	79
		% within CL_1ERPARBINARI	74,7%	25,3%	100,0%
Total		Count	67	20	87
		% within CL_1ERPARBINARI	77,0%	23,0%	100,0%

Y, finalmente, obtendremos el Grafo de Markov siguiente:

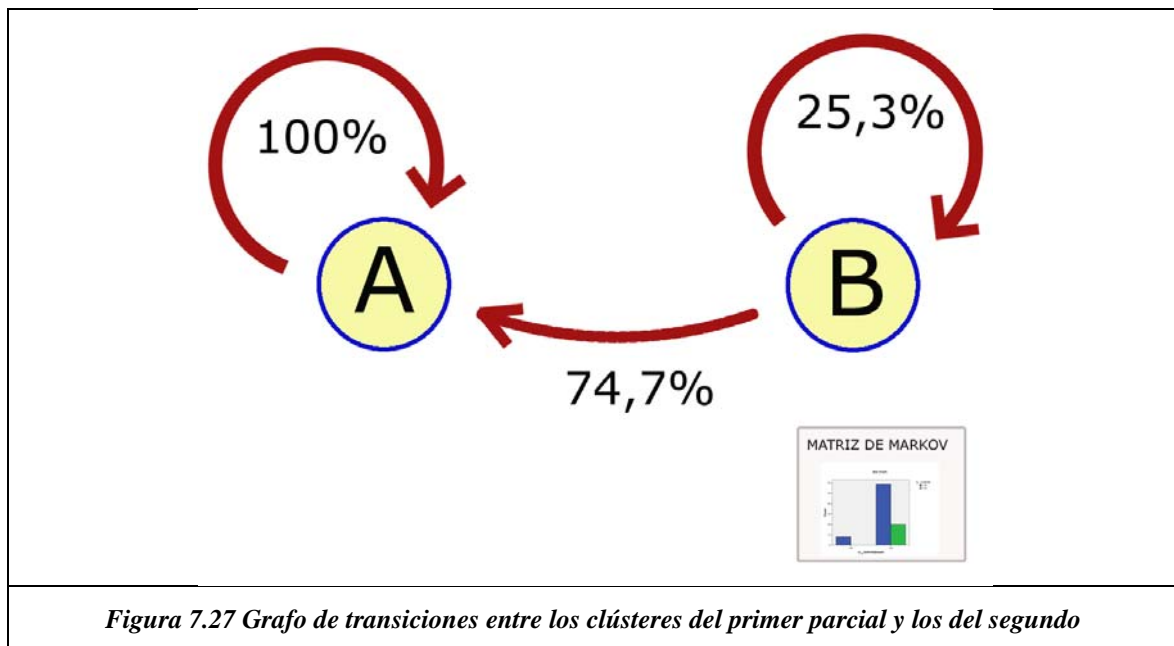


Figura 7.27 Grafo de transiciones entre los clústeres del primer parcial y los del segundo

Nos queda un estado absorbente, A, y un segundo estado con probabilidades aparentemente no demasiado sesgadas (que nos lleven de nuevo a una patología). Un sistema consistente de clústeres como éste, en el que al menos una parte sí es biunívoca (la de la transición entre segundo y tercer parcial) nos hace pensar en el éxito de este modelo de encaje. Sin embargo, esta solución no es más que el primer tanteo de un algoritmo heurístico de *trial and error* y, obviamente, nada está garantizado.

7.4.2 Comparación final entre el estado esperado y el pronosticado

Del mismo modo que hemos procedido con los estudiantes de multimedia, para poder comprobar la significación de los resultados, se opta por estudiar el Estado final pronosticado con el objeto de compararlo con el resultado observado y así medir la desviación producida por la intervención del centro.

Con las suposiciones que hemos hecho, el estado inicial es:

Clúster	Individuos
A	8
B	79

El producto de la matriz de transición por sí misma, que hemos denominado M_E , es:

$$\begin{pmatrix} 1 & 0 \\ 0.747 & 0.253 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0.747 & 0.253 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.935991 & 0.064009 \end{pmatrix}$$

Por tanto, las frecuencias esperadas se obtienen multiplicando el vector de estados iniciales por M_E :

$$(8 \quad 79) \times \begin{pmatrix} 1 & 0 \\ 0.935991 & 0.064009 \end{pmatrix}$$

Ello nos da las frecuencias esperadas:

Clúster	Individuos
A	81,943289
B	5,056711

Que son las que usaremos para contrastar con los resultados obtenidos al final, considerando sólo los 2 clústeres finales que hemos tomado en nuestra clasificación. Éstos son:

Clúster	Individuos
A	21
B	66

Como puede verse, nuevamente estamos ante un caso patológico: la discrepancia entre frecuencias esperadas y observadas es tan grande que no nos hemos movido de la situación que ha motivado esta exploración de caminos alternativos. La prueba Chi cuadrado es significativa, como era de esperar, pero la información que aporta no es útil para la investigación.

Hace falta, pues, explorar las otras dos opciones que nos quedan entre las que hemos enunciado.

7.5 Corrección a sistema biunívoco de 3 clústeres

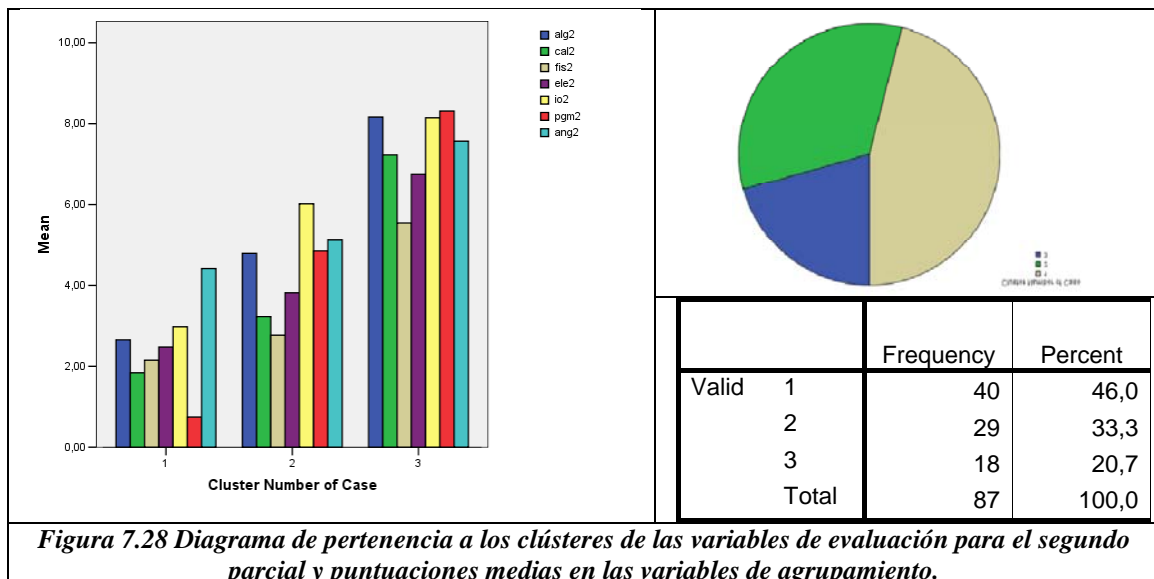
Por dos veces nos hemos encontrado que nuestro método nos ha llevado a un camino que no da información. Este desenlace ha sido provocado porque hemos simplificado demasiado la realidad al considerar solamente 2 clústeres en el sistema consistente. El comportamiento académico de los estudiantes de ingenierías tecnológicas tiene matices que a una clasificación tan simple se le escapan.

Así las cosas, es obvio que un investigador en esta situación se plantee forzar un número mayor de clústeres con la esperanza de conseguir un buen ajuste. El siguiente candidato es 3. La opción es tentadora porque es la que, de forma natural, ha surgido para explicar la conducta de los estudiantes de ingeniería multimedia.

7.5.1 Obtención de los clústeres para el segundo y tercer parcial, forzando a 3 grupos

Mediante el paquete estadístico SPSS realizamos una prueba de agrupación de tipo K-MEANS en la que se fuerza a 3 el número de clústeres. Dicha operación se hace para el segundo parcial y para el tercero. Dado que el primer parcial ya presentaba de forma natural 3 clústeres, el resultado es un sistema consistente biunívoco.

Hay que recordar que, una vez hallados los clústeres, se hacía una corrección de las etiquetas para maximizar las correlaciones entre ellos (a fin de no perder el significado para las interpretaciones). Una vez hecha dicha corrección (recordemos que se trataba de una reordenación de los clústeres) se obtiene:



Podemos observar que el clúster de los buenos alumnos prácticamente se ha mantenido (de 20 a 18 alumnos) y el clúster de los alumnos menos competitivos se ha repartido en 2.

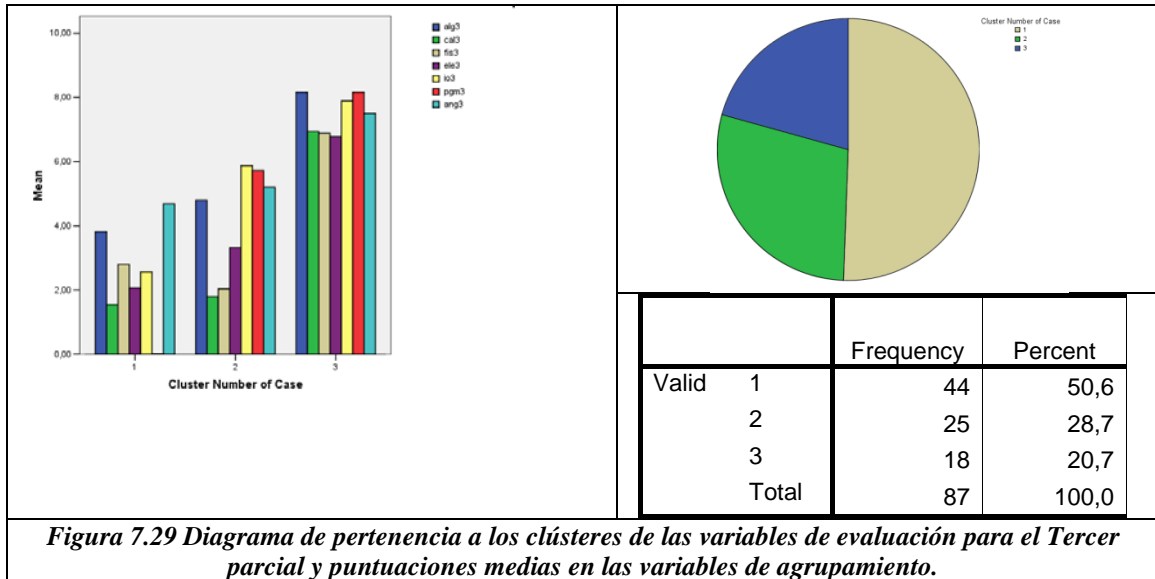


Figura 7.29 Diagrama de pertenencia a los clústeres de las variables de evaluación para el Tercer parcial y puntuaciones medias en las variables de agrupamiento.

El clúster de los alumnos de alto rendimiento se ha reducido (de 26 a 18 alumnos) pero han subido las notas. Los clústeres 1 y 2 se han agrupado en el 1.

7.5.2 Transiciones entre el primer y segundo parcial

Con estos datos, la transición entre el primer parcial (3 clústeres naturales) y el segundo (3 clústeres forzados) se plasma en la siguiente figura:

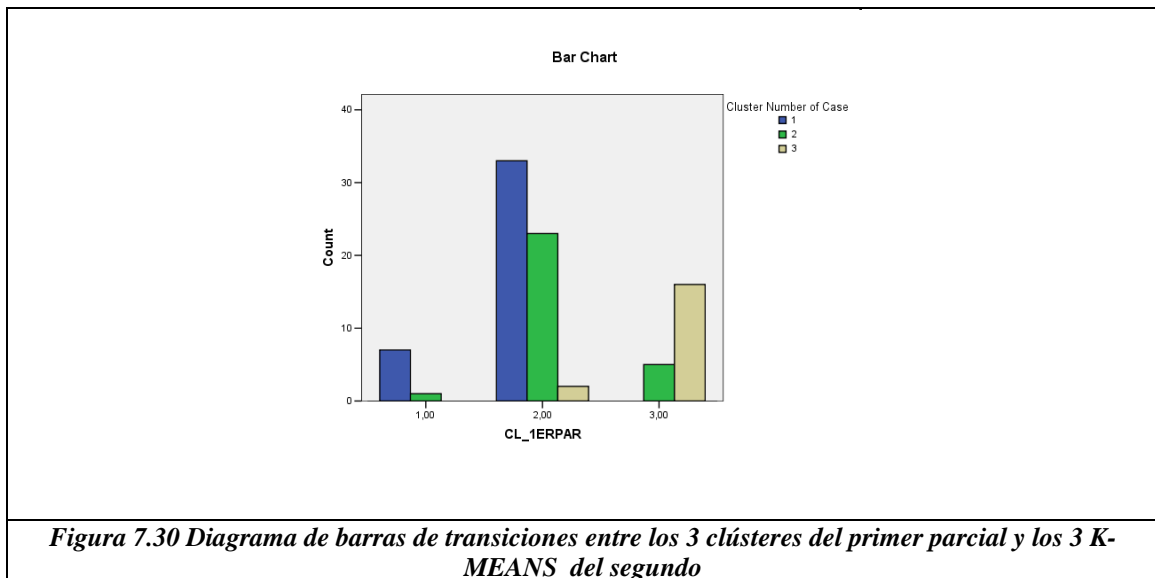


Figura 7.30 Diagrama de barras de transiciones entre los 3 clústeres del primer parcial y los 3 K-MEANS del segundo

Por lo que vemos, la presión a que nos referíamos la primera vez que tratábamos este sistema de clústeres se nota en los gráficos. Apreciamos que se produce un bajón en las notas, dado que los clústeres 2 y 3 del primer parcial pierden terreno; dicho terreno lo gana el clúster 1.

Siguiendo con el método, procedemos a realizar un CROSSTABS:

CL_1ERPAR * Cluster Number of Case Crosstabulation

			Clúster Number of Case			Total
			1	2	3	
CL_1ERP AR	1,00	Count	7	1	0	8
		% within CL_1ERP	87,5%	12,5%	,0%	100,0%
		% of Total	8,0%	1,1%	,0%	9,2%
	2,00	Count	33	23	2	58
		% within CL_1ERP	56,9%	39,7%	3,4%	100,0%
		% of Total	37,9%	26,4%	2,3%	66,7%
Total	3,00	Count	0	5	16	21
		% within CL_1ERP	,0%	23,8%	76,2%	100,0%
		% of Total	,0%	5,7%	18,4%	24,1%
Total		Count	40	29	18	87
		% within CL_1ERP	46,0%	33,3%	20,7%	100,0%
		% of Total	46,0%	33,3%	20,7%	100,0%

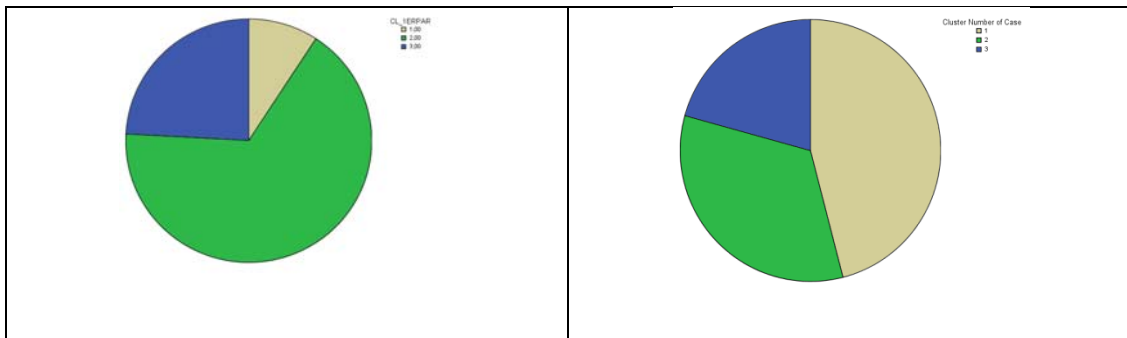


Figura 7.31 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del primero y segundo parcial.

Por motivos de brevedad, omitimos la matriz de estados, si bien su confección es inmediata a partir de la tabla anterior.

7.5.3 Calculo de la Matriz esperada ME a partir de la Matriz primero-segundo

Ahora podemos calcular la matriz M_E haciendo el producto de la primera transición por sí misma:

$$\begin{pmatrix} 0.875 & 0.125 & 0 \\ 0.569 & 0.397 & 0.034 \\ 0 & 0.238 & 0.762 \end{pmatrix} \times \begin{pmatrix} 0.875 & 0.125 & 0 \\ 0.569 & 0.397 & 0.034 \\ 0 & 0.238 & 0.762 \end{pmatrix} = \begin{pmatrix} 0.83675 & 0.159 & 0.00425 \\ 0.723768 & 0.236826 & 0.039406 \\ 0.135422 & 0.275842 & 0.588736 \end{pmatrix}$$

Sin embargo, la evolución obtenida desde el primer parcial directamente al tercero ha sido la siguiente:

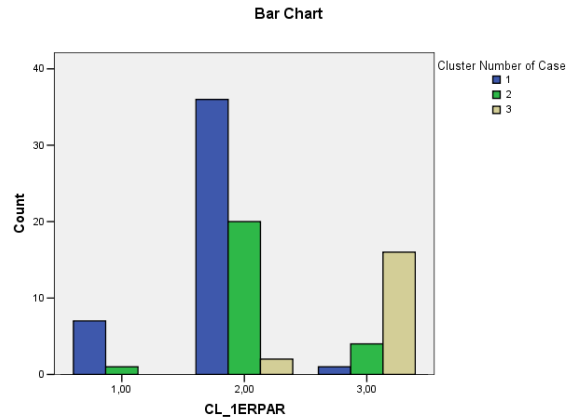


Figura 7.32 Diagrama de barras de transición entre el primer y el tercer parcial

De donde obtenemos con una prueba CROSSTABS las probabilidades de transición:

CL_1ERPAR * Cluster Number of Case Crosstabulation

		Clúster Number of Case			Total	
		1	2	3		
CL_1ERP AR	1,00	Count				
		7	1	0	8	
		% within CL_1ERPAR	87,5%	12,5%	,0%	100,0%
		% of Total	8,0%	1,1%	,0%	9,2%
	2,00	Count				
		36	20	2	58	
		% within CL_1ERPAR	62,1%	34,5%	3,4%	100,0%
		% of Total	41,4%	23,0%	2,3%	66,7%
	3,00	Count				
	1	4	16	21		
	% within CL_1ERPAR	4,8%	19,0%	76,2%	100,0%	
	% of Total	1,1%	4,6%	18,4%	24,1%	
Total	Count					
		44	25	18	87	
		% within CL_1ERPAR	50,6%	28,7%	20,7%	100,0%
	% of Total	50,6%	28,7%	20,7%	100,0%	

Es decir, que la matriz final observada de Markov, M_0 , es:

$$M_0 \begin{pmatrix} 0.875 & 0.125 & 0 \\ 0.621 & 0.345 & 0.034 \\ 0.048 & 0.19 & 0.762 \end{pmatrix}$$

Con lo cual hay que compararla con la pronosticada anteriormente, M_E :

$$M_E \begin{pmatrix} 0.83675 & 0.159 & 0.00425 \\ 0.723768 & 0.236826 & 0.039406 \\ 0.135422 & 0.275842 & 0.588736 \end{pmatrix} M_0 \begin{pmatrix} 0.875 & 0.125 & 0 \\ 0.621 & 0.345 & 0.034 \\ 0.048 & 0.19 & 0.762 \end{pmatrix}$$

Resolveremos este problema de la misma manera que hemos procedido en multimedia: trasladando la hipótesis sobre matrices a la comparación de estados finales. Así pues, el estado inicial es:

Clúster	Individuos
A	8
B	58
C	21

Por tanto, las frecuencias esperadas se obtienen aplicando dicho vector a la matriz M_E :

$$(8 \quad 58 \quad 21) \quad M_E \begin{pmatrix} 0.83675 & 0.159 & 0.00425 \\ 0.723768 & 0.236826 & 0.039406 \\ 0.135422 & 0.275842 & 0.588736 \end{pmatrix}$$

Ello nos da las frecuencias esperadas de los individuos en un sistema forzado a 3 clústeres:

Clúster	Individuos
A	51,516406
B	20,80059
C	14,683004

Que son las que usaremos para contrastar con los resultados obtenidos al final, es decir:

Clúster	Individuos
A	44
B	25
C	18

Hasta aquí, todo parece marchar bien. Pero, a la hora de realizar la prueba Chi cuadrado, nos encontramos con la sorpresa de que el resultado no es estadísticamente significativo:

fila2			
	Observed N	Expected N	Residual
1,00	44	51,5	-7,5
2,00	25	20,8	4,2
3,00	18	14,7	3,3
Total	87		

Test Statistics	
	fila2
Chi-Square(a)	2,687
Df	2
Asymp. Sig.	,261

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 14,7.

Hemos obtenido, pues, una patología diferente a las que habíamos visto hasta ahora. La clasificación (el sistema consistente de clústeres) sirve para conseguir unas matrices de estados tratables, pero el modelo que nos proporciona no es suficientemente sensible para, dentro de estas clasificaciones, poder apreciar la intervención del centro.

La última opción que le queda a nuestro método es la reducción a 4 clústeres.

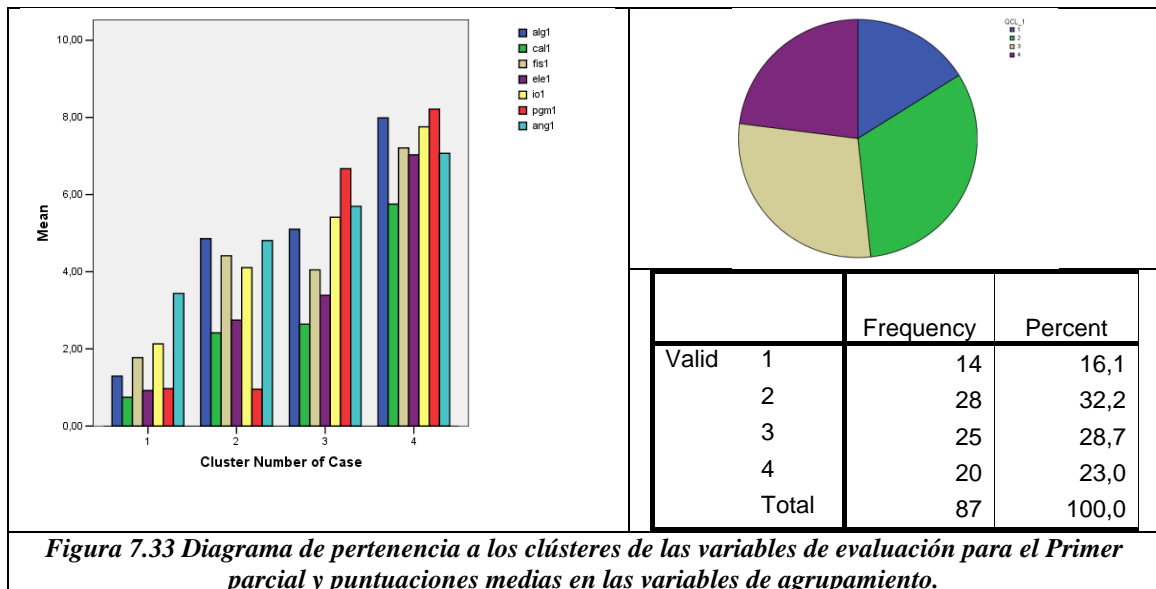
Como resumen de los tanteos realizados hasta el momento, tenemos la siguiente tabla:

Opción	Descripción	Resultado obtenido
Desarrollo inicial de un sistema consistente no biunívoco	Es el que resulta al aplicar el método de forma natural	Estadísticamente patológico. Resultados significativos pero que no aportan información
Reagrupación de clústeres manteniendo el sistema consistente biunívoco	Resulta de reinterpretar algunas agrupaciones a fin de reducir las discrepancias entre el número de clústeres	Estadísticamente patológico. Resultados significativos pero que no aportan información
Sistema consistente biunívoco forzando a un modelo de 3 clústeres	Resulta de forzar un número intermedio de clústeres que aparentemente puede funcionar	Estadísticamente estéril. Los resultados no son significativos. No acaba de casar con la situación que estamos analizando

7.6 Modelo explicativo de reducción a 4 clústeres

7.6.1 Construcción de los 4 clústeres de cada parcial

Estadísticamente, el primer paso no tiene demasiada complicación, ya que se trata de ejecutar una agrupación con el procedimiento K-MEANS forzado a 4. Obtenemos los siguientes resultados:



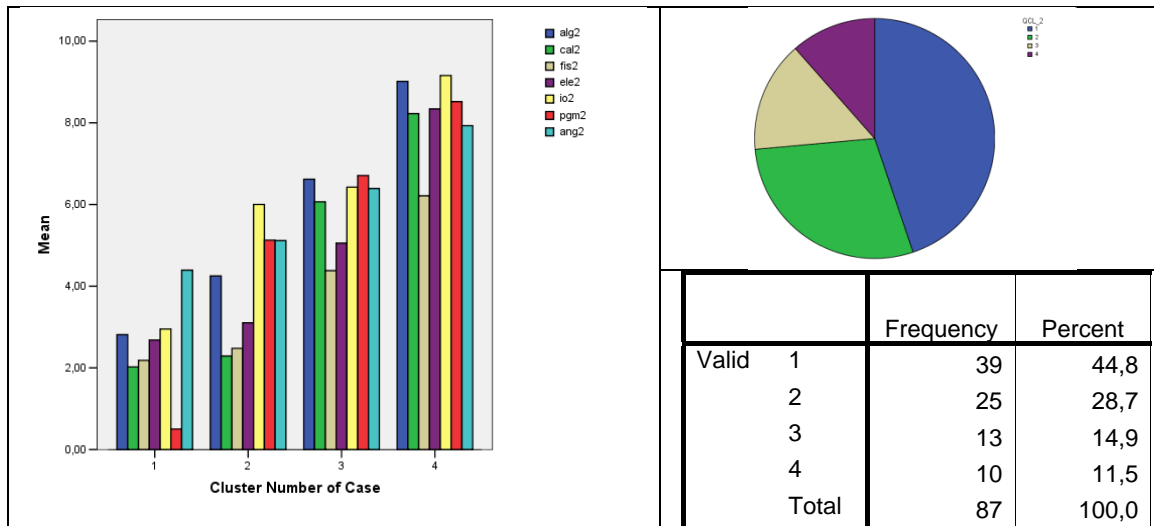


Figura 7.34 Diagrama de pertenencia a los clústeres de las variables de evaluación para el Segundo parcial y puntuaciones medias en las variables de agrupamiento.

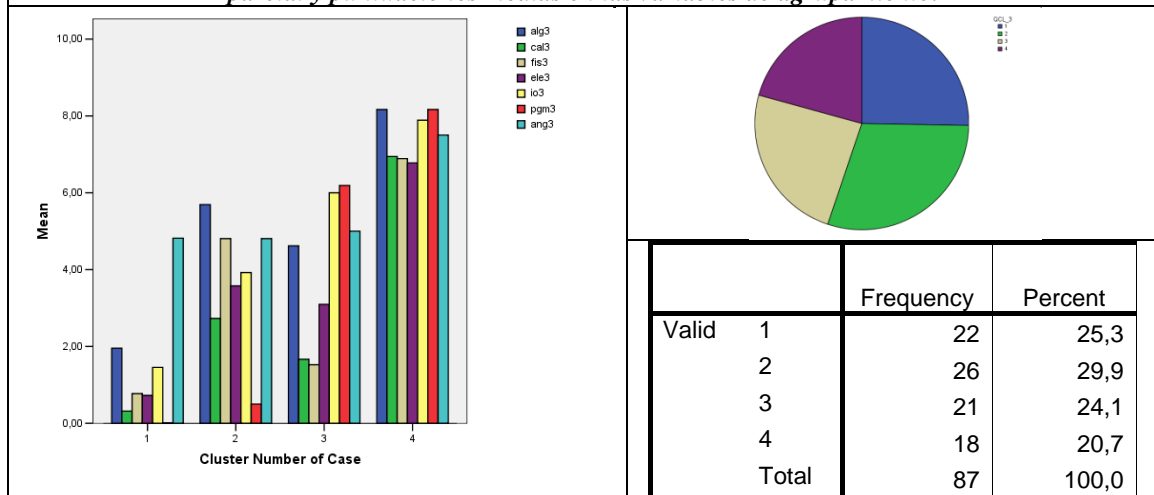


Figura 7.35 Diagrama de pertenencia a los clústeres de las variables de evaluación para el Tercer parcial y puntuaciones medias en las variables de agrupamiento.

7.6.2 Transiciones entre el primer y segundo parcial: Grafos y matrices de transición

Una vez considerados estos esquemas, la transición entre primer y segundo parcial queda plasmada en la siguiente figura:

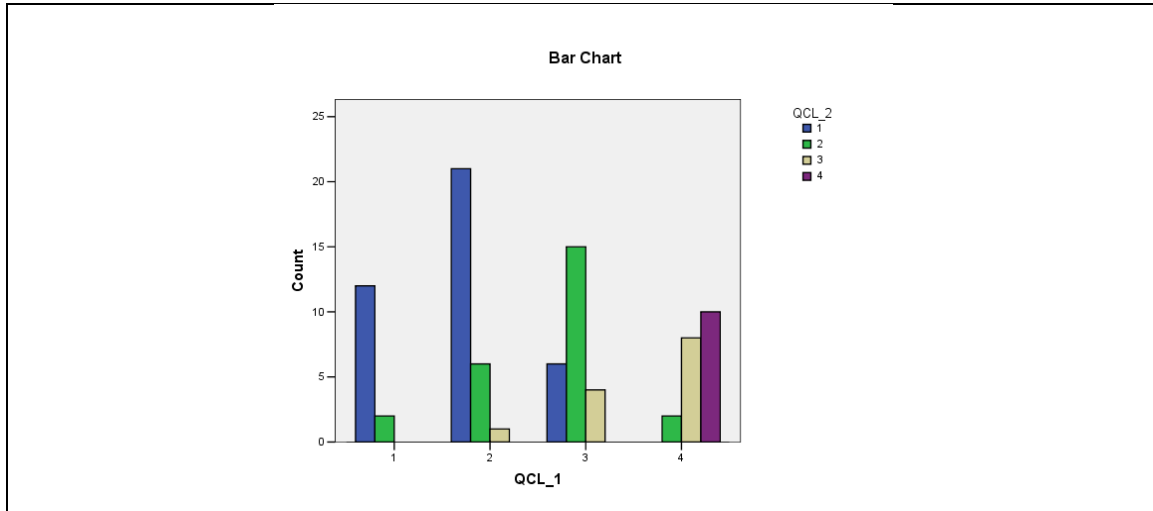


Figura 7.36 Diagrama de barras de transiciones entre los 4 Quick clústeres del primer parcial y los 4 Quick clústeres del segundo

Como siempre, realizamos el CROSSTABS para la obtención de las probabilidades de transición:

QCL_1 * QCL_2 Crosstabulation

			QCL_2				Total
			1	2	3	4	
QCL_1	1	Count	12	2	0	0	14
		% within QCL_1	85,7%	14,3%	,0%	,0%	100,0%
		% of Total	13,8%	2,3%	,0%	,0%	16,1%
2	2	Count	21	6	1	0	28
		% within QCL_1	75,0%	21,4%	3,6%	,0%	100,0%
		% of Total	24,1%	6,9%	1,1%	,0%	32,2%
3	3	Count	6	15	4	0	25
		% within QCL_1	24,0%	60,0%	16,0%	,0%	100,0%
		% of Total	6,9%	17,2%	4,6%	,0%	28,7%
4	4	Count	0	2	8	10	20
		% within QCL_1	,0%	10,0%	40,0%	50,0%	100,0%
		% of Total	,0%	2,3%	9,2%	11,5%	23,0%
Total		Count	39	25	13	10	87
		% within QCL_1	44,8%	28,7%	14,9%	11,5%	100,0%
		% of Total	44,8%	28,7%	14,9%	11,5%	100,0%

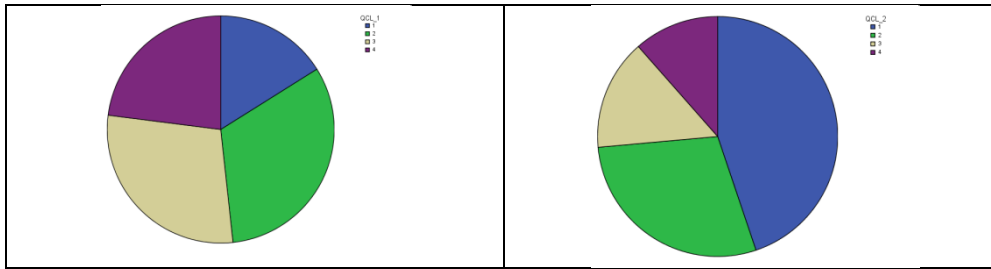


Figura 7.37 Diagrama de sectores del porcentaje que representa cada uno de los clústeres del primero y segundo parcial.

Con lo cual obtenemos la siguiente matriz de Markov, que ahora es una matriz 4 x 4:

$$M_{12} = \begin{pmatrix} 0.857 & 0.143 & 0 & 0 \\ 0.75 & 0.214 & 0.036 & 0 \\ 0.24 & 0.6 & 0.16 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \end{pmatrix}$$

Su correspondiente Grafo de estados es el que se plasma en la figura:

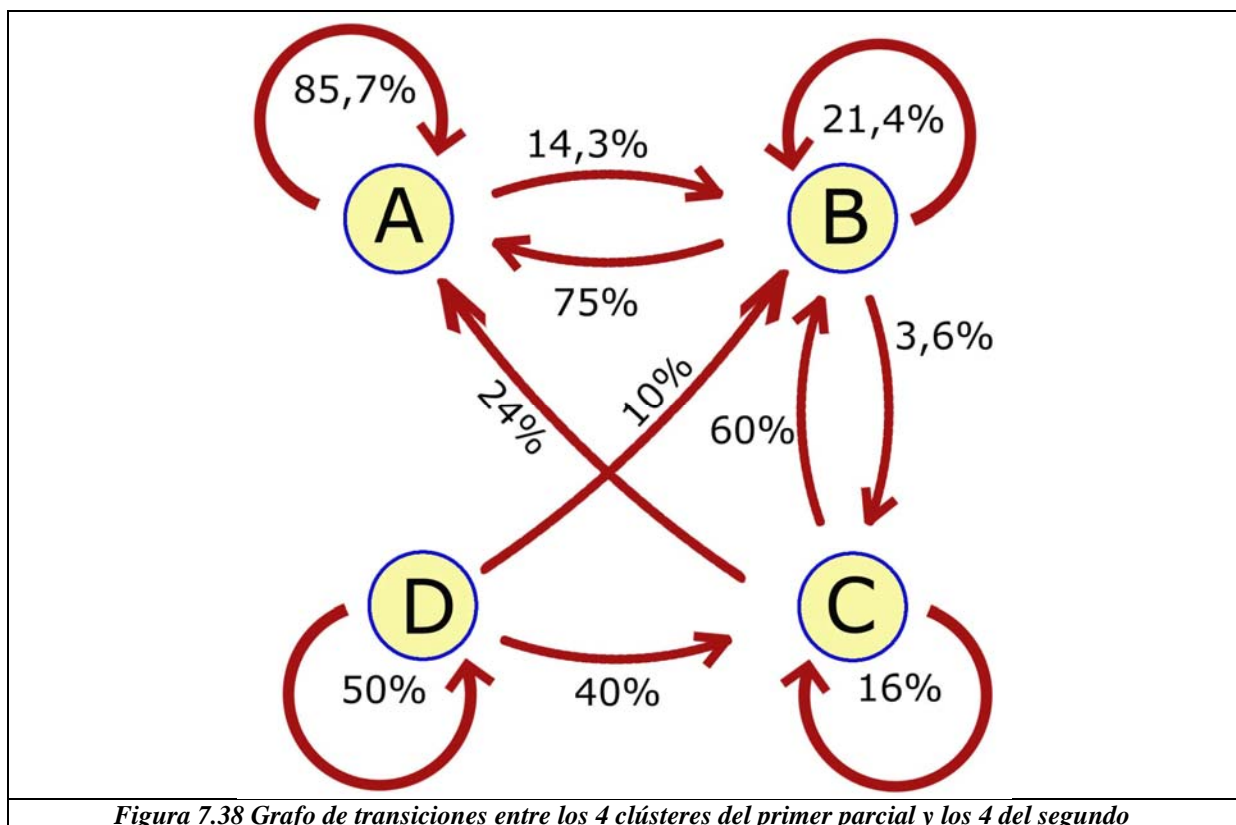


Figura 7.38 Grafo de transiciones entre los 4 clústeres del primer parcial y los 4 del segundo

Bajo la suposición de no intervención del centro, tenemos que multiplicar esta matriz por sí misma, para obtener la matriz de transición del primer al tercer parcial, M_E . Por tanto, las frecuencias esperadas se obtienen multiplicando el vector de los estados iniciales por M_E :

$$M_E M_{12^2} = \begin{pmatrix} 0.857 & 0.143 & 0 & 0 \\ 0.75 & 0.214 & 0.036 & 0 \\ 0.24 & 0.6 & 0.16 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \end{pmatrix}^2 = \begin{pmatrix} 0.841699 & 0.153153 & 0.005148 & 0 \\ 0.81189 & 0.174646 & 0.013464 & 0 \\ 0.69408 & 0.25872 & 0.0472 & 0 \\ 0.171 & 0.3114 & 0.2676 & 0.25 \end{pmatrix}$$

(14 28 25 20)

Lo que nos da el siguiente estado final esperado para el tercer parcial:

Clúster	Individuos
A	55,288706
B	19,73023
C	6,981064
D	5

Que son los que usaremos para contrastar con los resultados obtenidos al final, es decir:

Clúster	Individuos
A	22
B	26
C	21
D	18

7.6.3 Prueba de significación Chi cuadrado para pronunciarse sobre la intervención académica del centro

Después de este largo camino de tanteo, la agradable sorpresa de la reducción a 4 clústeres es que, a la hora de la prueba Chi cuadrado, los resultados no pueden ser mejores. La situación de partida es la que se refleja en la siguiente tabla:

datos

	Observed N	Expected N	Residual
1,00	22	55,3	-33,3
2,00	26	19,7	6,3
3,00	21	7,0	14,0
4,00	18	5,0	13,0
Total	87		

Y los resultados obtenidos en la prueba Chi cuadrado son:

Test Statistics

	Datos
Chi-Square(a)	83,997
df	3
Asymp. Sig.	,000

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 5,0.

Con lo que, por fin, se consigue un camino que encaja bien con el fenómeno que se quiere explicar y, además, detecta la intervención académica del centro. Éste es, por tanto, el sistema consistente de clústeres que hemos estado buscando para los estudiantes de ingenierías tecnológicas.

Con este resultado estadístico, que confirma el éxito del camino heurístico tomado para la resolución del problema, el método expuesto en esta tesis llega a su fin. Sólo queda ahora la interpretación de resultados con el enfoque y los parámetros que ha establecido la solución hallada. Dicha labor se realiza en el siguiente capítulo.

Capítulo 8

Conclusiones

8.1 Recapitulación: esquema general de la tesis

De la presente investigación hemos resaltado una doble aportación a la hora de atacar un problema dado: la contribución de contenido (distinción entre dos tipos de estudiantes de ingeniería) y la contribución metodológica (presentación de un método de análisis de la información basado en la evolución de clústeres).

Para cada una de estas aportaciones, hemos realizado sendas exposiciones en los dos primeros capítulos:

- a) **Capítulo 1:** Presentación de la problemática de los estudiantes de ingenierías, percepción del profesorado de los mismos, aspectos contextuales, iniciativas institucionales, historia de los estudios multimedia, etc.
- b) **Capítulo 2:** Presentación de la metodología basada en el análisis de la evolución de los clústeres, relación con métodos precedentes (cadenas de Markov), notación utilizada (sistemas consistentes, biunívocos, etc.) y comentarios conceptuales (interpretación, observaciones).

Así las cosas, hemos llegado al **Capítulo 3** donde hemos formalizado el problema de investigación (objetivos, enunciado, hipótesis y variables) y hemos recogido la doble vertiente anunciada en los capítulos 1 y 2. Para fijar ideas, podemos recordar cómo hemos definido el objetivo de la tesis:

Objetivo general de la investigación

Distinguir los perfiles académicos de los estudiantes de Multimedia y los de ingenierías estrictamente tecnológicas pero aportando una metodología de análisis de la realidad académica novedosa, basada en la evolución de clústeres a lo largo del curso.

Dado que la contribución de contenido se deriva de los resultados de la parte metodológica, se ha desarrollado ésta en los capítulos 4, 5, 6 y 7 de la siguiente manera:

- a) **Capítulo 4:** Exploración de los datos y detección de los clústeres que pueden entrar en juego en el sistema consistente. Etiquetación e interpretación de los mismos. Correcciones a los diferentes clústeres obtenidos y discusión de sus características. Exposición de otras lecturas que podrían llevarse a cabo con los sistemas de clústeres obtenidos.
- b) **Capítulo 5:** Construcción del sistema consistente de clústeres que va a ser usado en la investigación. Tratamiento realizado a las variables objeto de estudio, ordenaciones de clústeres para una mejor interpretación y tratamiento estadístico. Establecimiento del sistema de clústeres definitivo para las ingenierías multimedia y para las tecnológicas.
- c) **Capítulo 6:** Estudio descriptivo de los sistemas consistentes de clústeres. Análisis y exposición de las evoluciones de los clústeres para las dos ingenierías.
- d) **Capítulo 7:** Estudio inferencial de los sistemas consistentes de clústeres, realizado bajo el enfoque de las cadenas de Markov. Pruebas de significación para aportar la solidez requerida a los datos observados.

Los diversos interrogantes sobre las características de los estudiantes han sido respondidos en estos capítulos de análisis y construcción de los sistemas de clústeres. La contribución metodológica de esta investigación ha quedado ilustrada en ellos, sirviendo como guía para investigaciones posteriores. Es obvio que el método puede ser ampliado, variado o desarrollado mediante diferentes enfoques o técnicas; de hecho, hemos indicado algunas derivaciones posibles que no se han llevado a cabo debido a la necesidad de una delimitación en todo estudio como el presente.

Después de esta visión de conjunto, sólo nos queda ahora añadir las conclusiones que el autor elaborará basándose en los análisis realizados y añadiendo a ellos el conocimiento derivado de su experiencia profesional. De manera sintética se expondrán las conclusiones a las que se llega con la unión de ambas fuentes de conocimiento.

8.2 Diferencias entre alumnos de Ingeniería multimedia y alumnos de Ingenierías tecnológicas

Como hemos podido observar en los clústeres obtenidos, tanto en los primeros tanteos para su construcción como en los sistemas definitivos que han entrado en juego, hay una diferencia sustancial entre las estructuras de datos de los alumnos que cursan Ingeniería multimedia y los que cursan las demás Ingenierías tecnológicas. Hay que señalar, por tanto, que los comentarios informales del profesorado del centro, sus diferentes percepciones, y su consecuente adaptación de las metodologías docentes a los diferentes perfiles de estudiantes no partían de una quimera, sino de una realidad que, simplemente, no se había sometido a observación científica con anterioridad.

En tal observación han aparecido algunas informaciones ya conocidas o sospechadas, sin embargo, también han aflorado datos nuevos importantes. De ellos hemos destacado en la investigación dos especialmente relevantes:

- a) La componente matemática-abstracta por encima de la física-experimental en el perfil de los estudiantes de ingeniería multimedia.
- b) La mayor regularidad en los clústeres obtenidos de los alumnos multimedia en comparación con los de ingenierías tecnológicas.

Estas dos características nos llevan a cuestiones que podrían desarrollarse en investigaciones posteriores. La primera nos remite a la figura emergente un ingeniero multimedia predisuesto o que aspira a trabajar en el futuro con “información pura”, con cierto alejamiento o animadversión hacia lo experimental, lo cual se ha reflejado en sus perfiles de rendimiento. La segunda nos remite a una cuestión más difícil de desgranar: ¿Por qué los alumnos multimedia forman grupos más estables que los de las otras tecnologías? ¿Se debe a una mayor capacidad de reacción de los segundos ante los fracasos en las evaluaciones? ¿Se debe a que los estudiantes de multimedia podrían dividirse en un grupo ciertamente vocacional, preparado y dispuesto a un cierto sacrificio personal en sus estudios, y en otro que acude a La Salle para la adquisición de una cultura tecnológica de la cual sólo mostrará interés por temas concretos? En todo caso, las preguntas escritas quedan para que otros investigadores las respondan.

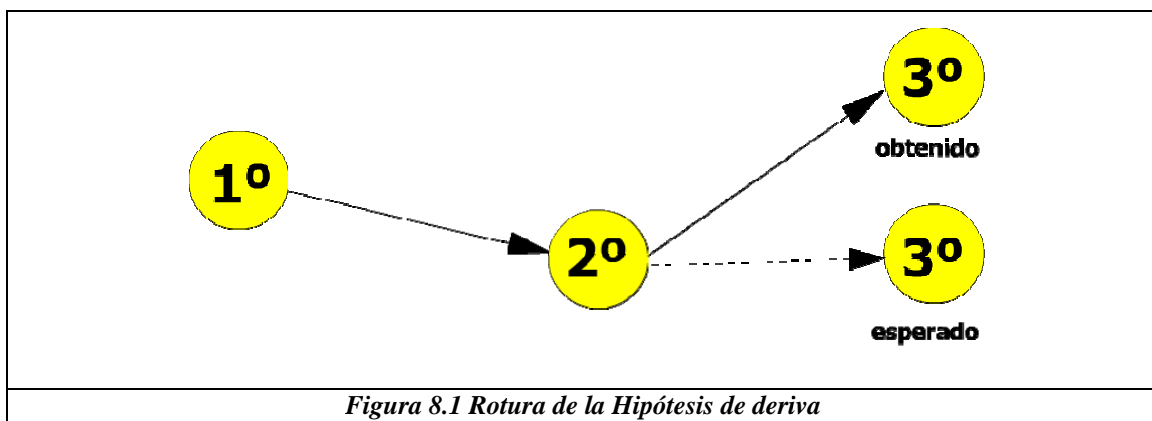
Estas cuestiones relativas al rendimiento pueden estudiarse con detalle en los gráficos de los capítulos precedentes. A modo de síntesis, una figura comparativa entre el número de clústeres para los dos perfiles en los diferentes parciales, plasma la diferencia estructural:

Parcial	Multimedia				Tecnológicas				
	Estructura			Clústeres	Estructura			Clústeres	
Primero	Alto	Medio	Bajo	3	Alto	Medio	Bajo	3	
Segundo	Alto	Medio	Bajo	3	Muy Alto		Muy Bajo		
Tercero	Alto	Medio	Bajo	3	Muy Alto	Medio Alto	Medio Bajo	Muy Bajo	4

8.3 Semejanzas entre alumnos de Ingeniería multimedia y alumnos de Ingenierías tecnológicas

A pesar de estas diferencias estructurales en los grupos de rendimiento, la tendencia de las notas en ambos tipos de estudios tiene un patrón común. Ello se debe en gran medida a que los estudios en ingenierías de La Salle, en el primer parcial, se orientan a repasar conceptos clave e igualar el nivel mínimo de los alumnos, ya que recibe estudiantes de múltiples centros, cada uno con unas exigencias de aprendizaje distintas. Esta orientación provoca que los estudiantes avanzados tanto de multimedia como de tecnológicas tengan puntuaciones bastante elevadas en el primer parcial. Los demás pueden superarlo con más o menos éxito pero, de todas formas, es en el segundo parcial en donde va a producirse una mayor discriminación entre los diferentes alumnos.

Se da, por tanto, una bajada de notas o un aumento de la desviación en el segundo parcial, la cual se recupera en el tercer parcial debido a la acción del centro y a la propia maduración de los alumnos que no abandonan los estudios. Esta recuperación en el tercer parcial contradice la “hipótesis de deriva”, es decir, el resultado esperado para los alumnos en la tercera evaluación a raíz de lo observado en el paso de la primera a la segunda. Si la “hipótesis de deriva” fuese cierta, el aprobado a final del curso sólo sería asequible para unos pocos alumnos entre los dos centenares de matriculados en las dos titulaciones.



¿Cómo se explica el cauce de alumnos hacia un resultado más satisfactorio? La experiencia del autor de esta tesis señala tres factores principales que han influido en dicho cambio. Dos de ellos provienen del centro: el sistema de evaluación continuada y las tutorías de seguimiento y asesoramiento. El tercero hay que buscarlo en el propio alumno: la maduración producida en el desarrollo (anual) de la asignatura. Podemos comentar con más detalle estas tres influencias:

Maduración

Se observa una maduración no solo cognitiva sino también social-organizativa en el colectivo de estudiantes. Puede decirse que el estudiante de ingeniería “aprende a serlo” durante el primer curso de la titulación. Las tutorías y algunos mecanismos informales de recogida de información dejan constancia de un cambio notable de actitud. No sería de recibo atribuir el éxito general en el tercer parcial exclusivamente a la intervención directa del centro, ya que hay que considerar que tal intervención es una componente que se suma a esta maduración observada. Otra cosa es plantearse si el centro propicia y ayuda a dicha maduración y, en este caso, la respuesta es afirmativa. La cultura de centro de la cual se impregnan estos alumnos parte de un diseño anual de las asignaturas, de mecanismos de colaboración y trabajo en equipo, de sistemas de participación (y, por tanto, responsabilidad) en proyectos de transferencia de tecnología.

Otro aspecto importante que ayuda a la maduración es la incidencia importante de las prácticas en el desarrollo del aprendizaje de las asignaturas más tecnológicas.

Si a estas características le añadimos que se trata de un centro organizado en unos pocos edificios con espacios comunes (servicio de bar, gimnasio, salas de pc's, etc.) y con unas normas de absoluta libertad de movimiento y organización para los alumnos, a la vez que con exigencia de responsabilidad, podemos entender que se trate de un espacio pensado y orientado para pasar de la juventud a la adultez. Este paso va a ser, pues, tanto personal como profesional.

Evaluación continua

La evaluación continua es defendida por diversos profesores debido a que permite tener un control más exacto del alumno y sus necesidades. Sirva la siguiente cita para ilustrar tal punto de vista:

"La evaluación continua ofrece al profesor, con un concepto dinámico de la perfección, la experiencia diaria con cada alumno, que beneficiará a los demás alumnos y a las futuras programaciones. Frente al sin sentido de marcarse objetivos a largo plazo, disponer de los medios y lanzarse a la tarea, esperando pasivamente el resultado final (D. B. Sánchez)."

En el caso de La Salle se ha combinado dicho sistema con la acción tutorial. Si bien supone un esfuerzo extra para el profesor, se obtiene la recompensa de la inmediatez, al ser capaz de tomar medidas adecuadas a tiempo, reconduciendo en muchos casos un posible abandono de los estudios.

Desde el punto de vista del estudiante, el primer cambio respecto a la evaluación final es que el alumno no tiene que focalizar su esfuerzo solo ante un examen o control sino que lo distribuye a lo largo del curso. Ello le supone las siguientes ventajas:

- La asimilación de los contenidos de forma gradual por parte del estudiante
- La información que recibe el estudiante sobre su propio ritmo de aprendizaje
- El conocimiento que llega a adquirir el estudiante de la forma de evaluar del profesor
- El aumento de la motivación en el estudiante si percibe este sistema como una oportunidad de éxito.
- Poder modificar el ritmo de la clase del profesor en función del seguimiento del alumnado.

Tutorías

Las tutorías se entienden en La Salle como un elemento de trato individual con el objetivo de la integración del alumno. Es decir, en estudiantes con dificultades en el estudio se busca la recuperación del ritmo adecuado y, en los casos más acentuados, evitar el abandono. Dichas sesiones de tutoría se realizan mediante entrevistas de orientación para las cuales se han recogido previamente datos sobre el rendimiento y la actitud general del alumno.

Como fruto de las mismas, el estudiante obtiene orientación tanto en asignaturas concretas (cómo afrontar las pruebas de evaluación, por ejemplo) o de su curso en general (elaboración de planes para superar el máximo número posible de asignaturas). Una orientación de este tipo es valiosa de por sí en unos estudios que encierran dificultad, como son las ingenierías tecnológicas; pero, por otro lado, también tiene una relevancia especial en los estudios de multimedia ya que, como se ha indicado, ingresan en ellos alumnos con características muy diferentes y, por lo general, con algunas lagunas importantes en la formación científica de base.

En el caso particular de los alumnos con muchas dificultades hay que recordar que es preceptivo que aprueben un mínimo de créditos en el primer curso ya que, caso contrario, no podrán continuar cursando sus estudios. Este colectivo es uno de los particularmente beneficiados por el sistema tutorial de La Salle en el sentido que hemos indicado de las estrategias sobre el curso en general. Es obvio que dicho sistema requiere una dedicación y un compromiso notable del profesorado por cuanto lleva de observador implicado en el proceso de formación. Al respecto, diversos listados detallan las características que deben cumplir las tutorías en la formación superior y podemos tomar como referencia el de la Universidad de Navarra puesto que La Salle cumple los puntos señalados en el *check list* que propone:

- Acompañamiento del estudiante en la acogida por parte de la universidad.
- Ayuda en la adaptación al nuevo contexto universitario. Integración del alumnado a la vida académica.
- Ayudar en las dificultades con las que el alumno se encuentre en su proceso formativo, sugiriéndole estrategias de aprendizaje y gestión del tiempo que mejoren su rendimiento académico.
- Ayudar al alumno en adquirir un compromiso de trabajo personal.
- Estimulación de la participación en las actividades extraacadémicas.

- Analizar y valorar con el alumno las calificaciones que va obteniendo en sus trabajos y exámenes.
- Aconsejarle en el modo de afrontar el nuevo curso
- Ayudarle a elegir las asignaturas optativas más adecuadas a sus intereses y preferencias.
- Aconsejarle en cuanto al tipo de prácticas en instituciones o empresas que más van a contribuir en el desarrollo de sus competencias profesionales.
- Ponerle en contacto con el profesor que coordina los programas de movilidad estudiantil de su titulación, para que pueda valorar con el alumno qué programa es más útil para su formación, ayudándole a elegir una universidad extranjera que imparta enseñanzas afines a su titulación.
- Informarle acerca de los estudios de postgrado que puedan ofrecerle una formación especializada que complemente la recibida en su titulación.
- Informarle acerca de las becas y ayudas que pueda pedir.
- Orientarle y apoyarle en su proceso de inserción laboral e informarle acerca de los recursos que la universidad le ofrece para ayudarle en esta tarea.

Hay que resaltar que esta metodología docente se desarrolla en La Salle desde hace muchos años, por lo que puede considerarse precursora de las directrices establecidas en el enfoque formativo de Bolonia, aplicadas a estudios de ingeniería.

8.4 Consecuencias derivadas de nuestro análisis: Titulaciones adaptadas al nuevo EEES

Uno de los motivos que llevó al doctorando a realizar esta tesis es su preocupación por la dificultad de los estudiantes en cuanto a su adaptación a la Universidad en su primer año académico. Además su inicio coincidió en el momento que las Universidades se estaban planteando la adaptación de sus títulos al nuevo Espacio Europeo de Educación Superior.

Las conclusiones de este trabajo se han aprovechado para definir algunos aspectos sobre la metodología de enseñanza-aprendizaje que ya se están aplicando a los nuevos grados que se están impartiendo en el Centro:

- En los dos primeros cursos las asignaturas se han definido anuales para aprovechar la maduración del alumno y la ayuda que pueden proporcionar los tutores.
- Se está aplicando un sistema de evaluación continuada en todos los cursos para mejorar el rendimiento de los alumnos.
- Se ha implantado una coordinación transversal en cada curso académico para todos los grados.
- Se ha mejorado el sistema de gestión de notas parciales para facilitar el seguimiento del alumnado por parte del profesor.
- Cada grupo de primer curso tiene asignado un tutor y a partir de segundo los alumnos tienen asignado un tutor en función del grado que estudian.
- Seminarios para los alumnos de primer curso con el objetivo de ayudarles en el proceso de maduración como estudiantes universitarios.
- Incidencia ya desde el primer curso en el trabajo colaborativo en equipo.

8.4 Directrices para el futuro.

Después de la *normalización* que ha supuesto la homologación de los estudios de multimedia se espera para los próximos años una aproximación entre los perfiles de los alumnos de Ingeniería multimedia y los de Ingenierías Tecnológicas. Ello es debido a que los estudios de multimedia se definían hasta ahora como un conocimiento espurio, que cubría una necesidad del mercado laboral, pero que padecía confusiones epistemológicas como la adscripción misma a diferentes departamentos o áreas de conocimiento.

En la actualidad queda establecido que el titulado multimedia será un ingeniero con preparación para el desarrollo y la creación de proyectos comunicativos apoyados por las tecnologías de la información y la comunicación, con un conocimiento básico de las temáticas de ingeniería y un complemento de otras disciplinas. Ello le distingue, por ejemplo, de un titulado en comunicación audiovisual, que no es un ingeniero sino un especialista en comunicación, o un diseñador, que es un especialista en composición visual y otras disciplinas relacionadas con el diseño. El ingeniero multimedia se distinguirá de los demás, por ejemplo, en aspectos de programación y análisis de aplicaciones, lo cual no va a limitarle para poder intervenir en otras facetas de las aplicaciones de tipo más general.

Por otra parte, los factores de recuperación del rendimiento de los alumnos (maduración, tutorías y evaluación continuada) deberán mantenerse si se quiere compaginar un nivel exigente en la formación de los titulados con un aumento de sus posibilidades de éxito en su carrera formativa. Ésta es una recomendación que se ha querido resaltar puesto que ha sido discutida frecuentemente en el centro y la presente tesis ha constatado de manera experimental los efectos de tales iniciativas docentes.

Bibliografía

- A GIMENO, J.; PÉREZ GÓMEZ, A. (Coords.) (1989): *La enseñanza: su teoría y su práctica*. Akal. Madrid.
- A JONASSEN, D. (ED) (1996): *The handbook of research on educational communications and technology*. Scholastic. New York.
- A WILSON, B. (Coord.) (1996): *Constructivist learning environments*. Educational Technology Publications. New Jersey.
- AGUILERA DEL PINO, A.M. (2005): *Análisis de tablas de contingencia bidimensionales*
- AHUMADA, W.; ESPEJO, R.; OTERO, J.; PÉREZ DE LANDAZÁBAL, M.A. (2007): *Condiciones de entrada de los estudiantes que ingresan a las carreras de Ingeniería*.
- AILON, N.; CHIEN, S.; DWORK, C. (2005): *On Clusters in Markov Chains*. Princeton University.
- ARNAL, J.; DEL RINCÓN, D.; LATORRE, A. (1992): *Investigación educativa. Fundamentos y metodología*. Editorial Labor S.A. Barcelona
- ARNAL, J.; DEL RINCÓN, D.; LATORRE, A. (1992): *Bases metodológicas de las investigación educativa*. Experiencia Ediciones. Madrid
- ARY, D.; JACOBS, CH.; RAZAVIEH, A. (1987): *Introducción a la Investigación Pedagógica*. McGraw-Hill. México
- AUSUBEL, D.; NOVAK, J.; HANESIAN, H. (1978): *Psicología Educativa*. Trillas. México.
- BARTOLOME, M. (1988): *Investigación-Acción, Innovación pedagógica y calidad de los centros educativos*. Bordon. Barcelona
- BARTOLOME, M. (1988): *Aspectos Metodológicos de investigación educativa*. NARCEA. Madrid.
- BARTOLOME, M. (1990): *Investigación cooperativa: Vías de innovación en la universidad*. PPU. Barcelona
- BEST, M. (2001) *The New Competitive Advantage*, pp. 60-90(31) Oxford Scholarship Online Monographs. Oxford
- BISQUERRA, R. (1989): *Métodos de investigación educativa. Guía práctica*. Ediciones Ceac, S.A. Barcelona
- CALDEIRO, G. P. (2005): *La teoría del aprendizaje significativo*.

- CASTAÑEDA GARCIA, J.A. (2005): *El comportamiento del usuario en internet: análisis de los antecedentes y consecuencias de la fidelidad*. Universidad de Granada.
- DUFY, T.; CUNNINGHAM, D. (1996): *Constructivism: Implications for Design and delivery of instruction*
- DUNN D.; LANDWEHR J. (1980): *Analyzing Clustering Effects Across Time*. Journal of the American Statistical Association, Vol. 75, No. 369. (Mar., 1980), pp. 8-15.
- FEIRSTEIN, G. A. (1999): *El cambio cognitivo durante la interacción docente-alumno: una propuesta de investigación*. Trabajo de suficiencia investigadora. Universidad Autónoma de Barcelona. Barcelona.
- FRÜHWIRTH-SCHNATTER, S.; PAMMINGER, C. (2007): *Bayesian Clustering of Categorical Time Series Using Finite Mixtures of Markov Chain Models with Application to Wage Mobility in Austria*. Francia
- GALAN MORENO, M. (2003): *Apuntes de estadística descriptiva*. ETSA.
- GILLICK, B.; MEDHAT GABER, M.; KRISHNASWAMY, S.; ZASLAVSKY, A. (2006): *Visualisation of Cluster Dynamics and Change Detection in Ubiquitous Data Stream Mining*. Monash University.
- HONEBEIN, P. (1996): *Seven goals for the design of Constructivist learning environments*.
- HUI-MIN CHEN, COOPER M. (2001): *Using Clustering Techniques to Detect Usage Patterns in a Web-Based Information System*. University of California at Berkeley.
- JONASSEN, D. (1991): *Objectivism vs. Constructivism*. An Educational Technology, Research and Development. 39(2). 5-14.
- KERLINGER, F.N. (1985): *Investigación del Comportamiento (Técnicas y Metodología)*. Interamericana. México
- KERLINGER, F.N. (1985): *Enfoque Conceptual de la Investigación del Comportamiento*. Interamericana. México
- LEIGH, N.; SILLS A. (2007): *Relative Frequencies of Blue Stragglers in Globular Clusters*. McMaster University.
- MARGALEF, J.; VICENT, L.; SENMARTÍ, X.; BOU, G.; ANGUERA, J. (2007): *The Teaching of Electronics Engineering: Guidelines for the Design of*

-
- Training In and Out of the Classroom setting.* ASEE/IEEE Frontiers in Education Conference
- MARGALEF, J. (2007): *La docència d'enginyeria electrònica: directrius per al disseny de la instrucció presencial i a distància.* Universitat Ramon Llull
 - MCGUIGAN, F.J. (1977): *Psicología Experimental. Enfoque metodológico.* Trillas. México
 - NEUSTADTL, A. (2007): *Analysis of variant-Anova*
 - NOTEBOOM, B. (2004): *Innovation, learning and cluster dynamic.* Tilburg University
 - PÉREZ GÓMEZ, A. (1989): *Paradigmas contemporáneos de investigación didáctica*
 - PÉREZ-OLMOS, I.; TALERO-GUTIÉRREZ, C.; GONZÁLEZ-REYES, R.; MORENO, C. (2006): *Ritmos circadianos de sueño y rendimiento académico en estudiantes de medicina*
 - PIAGET, J. (1986): *Estudios sociológicos.* Planeta-Agostini. Barcelona.
 - POIRIER, J. (2000): *Analyse de la variance et de la régression. Plans d'expérience.*
 - POZO, J. I. (1997): *El cambio sobre el cambio: hacia una nueva concepción del cambio conceptual en la construcción del conocimiento científico*
 - RAMONI, M.; SEBASTIANI, P.; COHEN, P. (2001): *Bayesian clustering by dynamics.* Douglas H. Fisher, Jr. Massachusetts
 - REINHARD, P.; SURAUD, E. (2003): *Introduction to Cluster Dynamics.* Wiley-VCH. Alemania
 - RINCÓN, L. (2003): *Manual de probabilidad y estadística para computación.* UNAM. México
 - RODRIGO, M. J.; ARNAY, J. (1997): *La construcción del conocimiento escolar.* Paidós. Barcelona.
 - SALGADO, P. (2007): *Medidas de asociación para variables cualitativas*
 - SEGOVIA, C.; ABURTO, L.; GOIC, M. (2006): *Caracterización del proceso de fuga de clientes utilizando información transaccional.*
 - SIT, V. (1995): *Analyzing ANOVA Designs.* Biom. Info. Hand. 5. Res. Br., B.C. Min. For., Victoria, B.C. Work. Pap. 07/1995

- SMOKE, K. L. (1932): *An objective study of concept formation*. Psychological Monographs. 42. 4.
- TARPY, R. (1989): *Principios básicos del aprendizaje*. Debate. Madrid.
- VAN DALEN, D.B.; MEYER, W.J. (1983): *Manual de Técnica de la Investigación Educativa*. Paidós Ecuador. Buenos Aires.
- VICÉNS OTERO, J.; MEDINA MORAL, E. (2005): *Análisis de datos cualitativos*.
- VICENTE G.; COTERON, A.; MARTINEZ, M.; ARACIL, J. (1997): *Application of the factorial design of experiments and response surface methodology to optimize biodiesel production*. Madrid.
- YELA, M. (1980): *La evaluación del conductismo. Análisis y modificación de conducta*. 6 (11-12). 147-180.
- ZHI-HUA FAN; DA-KE HUANG; JUAN-ZI LI; KE-HONG WANG (2004) *Morphing cluster dynamics: time series clustering with multipartite graph*

