

UNIVERSIDAD POLITÉCNICA DE VALENCIA

Departamento de Estadística e Investigación Operativa Aplicadas y Calidad



**ESTUDIO DEL CUESTIONARIO DE EVALUACIÓN DEL
PROFESORADO DE LA UPV MEDIANTE OPINIÓN DE LOS
ESTUDIANTES. TRATAMIENTO ESTADÍSTICO**

TESIS DOCTORAL

Presentada por:

Mónica Martínez Gómez

Directores:

Dr. José Miguel Carot Sierra

Dr. José Manuel Jabaloyes Vivas

Valencia, Enero 2005

A Manu

Agradecimientos

La realización de una tesis doctoral es un trabajo arduo que no es posible sin la colaboración de muchas y muy diversas partes. A lo largo de todo su desarrollo he encontrado el apoyo, amparo y colaboración de muchas personas a las que me gustaría expresar, desde estas páginas, mi más sincero agradecimiento.

En el ámbito académico quiero dar las gracias a mis directores, Jose Miguel Carot y José Manuel Jabaloyes, por el esfuerzo y dedicación que les ha supuesto la supervisión de este trabajo. A ambos les agradezco, no sólo sus comentarios y aportaciones, sino también su paciencia y comprensión. Además, quiero agradecer, especialmente, al Dr. Carot por mostrarme una forma de trabajar que considero muy valiosa y por haber seguido apoyándome desde el otro lado del océano y, de la misma manera, al Dr. Jabaloyes por creer siempre en el buen fin de este trabajo, incluso en los momentos en los que yo perdí la fe.

Asimismo, les debo mucho a las personas con las cuales he desarrollado mi trabajo en el seno del Departamento de Estadística, Investigación Operativa Aplicadas y Calidad, en particular, a los integrantes del grupo de investigación GADIC.

Para la recopilación de las bases de datos, he contado con la colaboración de Pedro Rodríguez, al que quiero reconocer su amabilidad y continua disposición a brindarme la información que le solicité, pese a las dificultades que le supuso.

También, quiero mostrar mi gratitud a Amparo Fernández, del Instituto de las Ciencias de la Educación, por sus consejos, buen trato y paciencia con la que ha atendido mi continua demanda de información.

Asimismo, agradecer el soporte y financiación de este trabajo, concedido por el Vicerrectorado de Personal Docente e Investigador de esta Universidad.

Por último, desde lo más profundo de mi corazón, quiero reconocer a mi familia y amigos todo el amor, apoyo y comprensión que me han brindado. A mis padres, a mi tía y a mi hermano, muy especialmente hoy, quiero agradecerles que estén siempre a mi lado y que sepan, en todo momento, escuchar, perdonar y alentar. Y a mi marido, su cariño y amparo incondicional, sin el cuál no habría sido capaz de concluir este trabajo.

Índice

1.	Justificación	1
1.1.	<i>Fundamentación Teórica</i>	1
1.2.	<i>Enfoques de Evaluación de la Docencia Universitaria</i>	5
1.3.	<i>Fuentes y Formas de Evaluación de la Docencia Universitaria</i>	6
2.	Objetivos	11
2.1.	<i>Contenido de la memoria</i>	14
3.	Antecedentes	17
3.1.	<i>Marco Legal</i>	17
3.2.	<i>La Evaluación del Profesorado por parte de los Alumnos</i>	21
3.3.	<i>Formas de Evaluación del Profesorado por los Alumnos</i>	29
3.4.	<i>Utilización de los resultados</i>	33
4.	Análisis del Instrumento de Evaluación del Profesorado de la UPV a partir de la Opinión de los Estudiantes.....	45
4.1.	<i>Introducción</i>	45
4.2.	<i>Características del Cuestionario</i>	47
4.3.	<i>Proceso de Evaluación</i>	50
4.4.	<i>Sistemas de Tratamiento de la Información</i>	53
4.5.	<i>Descripción del Estudio</i>	55
4.6.	<i>Análisis de Fiabilidad y Validez</i>	60

5.	Dimensionalidad del Cuestionario	67
5.1.	<i>Introducción</i>	67
5.2.	<i>Métodología</i>	68
5.2.1.	<i>Análisis Factorial Exploratorio</i>	68
5.2.2.	<i>Análisis Factorial Confirmatorio</i>	89
5.3.	<i>Resultados y Discusión</i>	115
5.3.1.	<i>Análisis Factorial Exploratorio</i>	115
5.3.2.	<i>Análisis Factorial Exploratorio Trabajando con Valores Medios</i> ...	137
5.3.3.	<i>Análisis Factorial Confirmatorio</i>	142
5.4.	<i>Conclusiones</i>	159
6.	Obtención de Grupos Homogéneos	161
6.1.	<i>Introducción</i>	161
6.2.	<i>Metodología</i>	161
6.2.1.	<i>Definición y diseño de la investigación</i>	162
6.2.2.	<i>Selección de Medidas de Similitud</i>	165
6.2.3.	<i>Supuestos del Análisis Clúster</i>	167
6.2.4.	<i>Selección del Algoritmo para obtención de conglomerados</i>	168
6.2.5.	<i>Análisis Discriminante</i>	182
6.3.	<i>Resultados y Discusión del Análisis Clúster</i>	191
6.4.	<i>Análisis de la Estabilidad de los Grupos Homogéneos en distintos periodos evaluativos</i>	200
6.5.	<i>Caracterización de los Clústers</i>	205

6.6. <i>Análisis Discriminante</i>	207
6.8. <i>Obtención de Grupos Homogéneos tras Análisis Discriminante</i> ..	216
6.9. <i>Conclusiones</i>	217
7. Identificación y caracterización de tipologías docentes.....	219
7.1. <i>Introducción</i>	219
7.2. <i>Métodología CHAID Exhaustivo</i>	221
7.3. <i>Metodología Análisis Discriminante</i>	235
7.4. <i>Métodología SIMCA</i>	235
7.4.1. <i>Procedimiento de aplicación del Método SIMCA</i>	240
7.4.2. <i>Clasificación con SIMCA</i>	248
7.4.3. <i>Sensibilidad y Especificidad del Método SIMCA</i>	251
7.4.4. <i>Ventajas del Método SIMCA</i>	252
7.5. <i>Resultados y Discusión</i>	252
7.5.1. <i>Definición de Tipologías Docentes</i>	252
7.5.2. <i>Identificación de los ítems de mayor poder discriminante en las tipologías docentes obtenidas</i>	272
7.5.3. <i>Identificación de los ítems de mayor poder discriminante entre las tipologías docentes obtenidas</i>	288
7.5. <i>Conclusiones</i>	292
8. Conclusiones y Valoración Final	295
8.1. <i>Conclusiones Generales</i>	295
8.2. <i>Limitaciones del Estudio</i>	300

8.3.Líneas Futuras de Investigación.....	302
9 . Bibliografía	303
ANEXOS.....	335
<i>Anexo 1. CEDA de la UPV.....</i>	<i>337</i>
<i>Anexo 2.Posibles tipos de relaciones entre variables</i>	<i>339</i>
<i>Anexo 3.Escalado de datos.....</i>	<i>341</i>
<i>Anexo 4.Resultados del AF.....</i>	<i>343</i>
<i>Anexo 5.Resultados del Análisis Clúster</i>	<i>351</i>
<i>Anexo 6.Resultados del LDA efectuado en cada nodo y SIMCA.....</i>	<i>363</i>

Resumen

La mejora de la calidad de las instituciones universitarias constituye el reto más importante en los próximos años para las universidades y el instrumento potencial para ello lo constituye la evaluación institucional, en general, y la evaluación de la actividad docente, en especial.

El cuestionario de opinión de los estudiantes, es el instrumento de evaluación de la actividad docente más generalizado en las universidades españolas.

El objetivo general del trabajo es desarrollar una metodología estadística adecuada para extraer, analizar e interpretar la información contenida en el Cuestionario de Evaluación Docente mediante Opinión de los Alumnos (CEDA) de la UPV, con la finalidad de optimizar su utilización práctica.

El estudio se centra en la aplicación de distintas técnicas multivariantes y ha sido estructurado en tres partes:

- Evaluación de la fiabilidad, validez y dimensionalidad del instrumento. La técnica multivariante a utilizar para ello es el Análisis Factorial Exploratorio. A continuación, se utiliza el Análisis Factorial Confirmatorio como técnica para verificar que el cuestionario capta las dimensiones básicas, que se pretenden evaluar y analizar la relación y adecuación de los ítems del cuestionario con cada de dichas dimensiones.
- Determinación de la capacidad del cuestionario para identificar perfiles diferenciados de profesorado a partir de la calidad percibida por los estudiantes. Con este objetivo se han utilizado y comparado distintas técnicas de clasificación: el análisis cluster jerárquico, no jerárquico y en dos etapas. A continuación, se procede a identificar los ítems del cuestionario que mejor discriminan entre las tipologías docentes obtenidas.
- Caracterización de tipologías docentes en función de distintas características descriptivas referentes a la asignatura y al profesor. En esta fase la metodología estadística utilizada es una combinación de los árboles de

decisión y análisis discriminante. De este modo se han podido describir cada una de las tipologías docentes encontradas. Finalmente, se comparan las tipologías en base a los ítems del cuestionario. Para ello se utilizará el método de clasificación SIMCA (Soft Independent Modelling of Class Analogy).

Con la utilización conjunta de las técnicas propuestas, se pretende optimizar el uso del CEDA como herramienta de medida e indicador de calidad de la enseñanza en la universidad, para introducir actuaciones de mejora continua en los procesos educativos de la UPV.

Summary

Quality improvement of university institutions represents the most important challenge in the next years, and the potential tool to achieve it is based on the institutional evaluation in general, and specially the evaluation of the teaching performance.

The opinion questionnaire from the students is the most generalised tool used to evaluate the teaching performance at Spanish universities.

The general objective of this thesis is to develop a statistical methodology suitable to extract, analyse and interpret the information contained in the Questionnaire of Teaching Evaluation from Student Opinion (CEDA) of the UPV, aimed at optimising its practical use.

The study is centred in the application of different multivariate techniques and has been structured in three parts:

- Evaluation of the reliability, validity and dimensionality of the tool. The multivariate method used for this purpose is the Exploratory Factorial Analysis. Afterwards, the Confirmatory Factorial Analysis is used as a method to verify that the questionnaire captures the basic dimensions that have to be assessed.
- Determination of the capacity of the questionnaire to identify different profiles of lecturers based on the quality perceived by students. This target is conducted with different multivariate classification techniques: hierarchical cluster analysis, non-hierarchical and two-stage analysis. Moreover, those items that best discriminate among the teaching typologies obtained are identified in the questionnaire.
- Identification of the teaching typologies according to different descriptive characteristics referent to the subject and lecturer, with the use of decision trees. Once identified these typologies, a new discriminant analysis is conducted aimed at identifying those items that best characterise each typology. Finally, a study is carried out with the classification method SIMCA

(Soft Independent Modelling of Class Analogy) in order to determine the discriminant loading of every item among the identified teaching typologies, allowing the identification of those that best distinguish the different classes obtained.

With the combined use of the proposed techniques, it is expected to optimise the use of CEDA as a measuring tool and an indicator of the teaching quality at the university, that would allow the introduction of actions for the continuous improvement in the teaching processes of the UPV.

Resum

La millora de la qualitat de les institucions universitàries constitueix el repte més important en els pròxims anys i l'instrument potencial n'és l'avaluació institucional en general i l'avaluació de l'activitat docent en especial.

El qüestionari d'opinió dels estudiants és l'instrument d'avaluació de l'activitat docent més generalitzat en les universitats espanyoles.

L'objectiu general del treball és desenvolupar una metodologia estadística adequada per a extraure, analitzar i interpretar la informació continguda en el Qüestionari d'Avaluació Docent mitjançant Opinió dels Alumnes (CEDA) de la UPV, amb la finalitat d'optimitzar la seua utilització pràctica.

L'estudi se centra en l'aplicació de distintes tècniques multivariants i ha sigut estructurat en tres parts:

- Avaluació de la fiabilitat, validesa i dimensionalitat de l'instrument. La tècnica multivariant que s'hi fa servir és l'Anàlisi Factorial Explorador. A continuació, s'hi utilitza l'Anàlisi Factorial Confirmatori com a tècnica per a verificar que el qüestionari capta les dimensions bàsiques que es pretenen avaluar.
- Determinació de la capacitat del qüestionari per a identificar perfils diferenciats de professorat a partir de la qualitat percebuda pels estudiants. Per a assolir aquest objectiu s'hi empren els anàlisis cluster jeràrquic, no jeràrquic i en dues etapes, com a tècniques multivariants de classificació. A continuació, es procedeix a identificar els ítems del qüestionari que millor discriminen entre les tipologies docents obtingudes.
- Identificació de tipologies docents en funció de distintes característiques descriptives referents a l'assignatura i al professor, on s'utilitza la tècnica dels arbres de decisió. Una vegada identificades aquestes tipologies, s'hi realitza un nou anàlisi discriminant per a identificar els ítems que millor caracteritzaven cadascuna d'elles. Finalment, s'hi aborda un estudi per a

conéixer el poder discriminant dels distints ítems entre les tipologies docents identificades, de manera que puguem identificar aquells que millor discerneixen entre les possibles classes obtingudes. Per a això s'utilitzarà el mètode de classificació SIMCA (Soft Independent Modelling of Class Analogy).

Amb la utilització conjunta de les tècniques proposades, es pretén optimitzar l'ús del CEDA com a ferramenta de mesura i indicador de qualitat de l'ensenyança en la universitat, per a introduir actuacions de millora contínua en els processos educatius de la UPV.

1. Justificación

1.1. Fundamentación Teórica

La evaluación constituye una poderosa herramienta para introducir en las instituciones educativas actuaciones de mejora continua que afecten tanto a los procesos propios de su organización como a los propiamente educativos. No obstante, evaluar es una tarea compleja, como se desprende de la multitud de enfoques de este concepto que podemos encontrarnos en la literatura revisada.

Los conceptos de evaluación y mejora están interrelacionados, ya que la evaluación adquiere su pleno significado cuando se orienta a la mejora y la mejora necesita de la evaluación para poder identificar la situación de partida y la de llegada.

El cambio, la innovación, la reforma y, en general, la mejora institucional la entendemos como uno de los objetivos básicos en la investigación sobre centros educativos en el momento presente.

Evaluar la Universidad significa valorar su actividad docente (aprendizaje de alumnos/as, rendimiento docente, etc.), su actividad investigadora (producción científica, calidad de las investigaciones, etc.) y su actividad de gestión (servicios disponibles para el alumnado y profesorado, gerencia, administración, etc.) De esta manera podremos identificar sus puntos débiles y fuertes y, en consecuencia, podremos incorporar las actuaciones necesarias destinadas a mejorar la calidad de la misma.

La evaluación de la Universidad, tal y como señala Luxán (1998), actualmente se sitúa en el núcleo del debate sobre el papel del Estado. Aunque la preocupación por la calidad en la educación universitaria no es nueva, el interés en el ámbito empresarial por los temas de calidad y la progresiva extensión de estos conceptos al sector público ha hecho reaparecer con fuerza este tema en el mundo universitario (Peña, 1997). Sin embargo, los estudios realizados hasta el momento sobre la evaluación de los sistemas educativos carecen de cierta coherencia y sistematización.

Como establece Ginés (1998), en los últimos años existe una preocupación creciente por la calidad de las instituciones universitarias por razones fundamentalmente políticas y económicas, entre las que cabe destacar:

- Proporcionar formación de calidad a una sociedad tecnológicamente avanzada.
- Controlar la eficiencia de las propias instituciones universitarias, dado el importante número de recursos proporcionados a su desarrollo.
- Proporcionar unos servicios de calidad (diversidad, flexibilidad, etc.) a los propios “clientes” en aras de satisfacer sus necesidades personales.
- Adaptar la formación a la realidad del mercado de trabajo.

La enseñanza universitaria en nuestro país ha sufrido cambios importantes en los últimos años. Se ha producido un cierto cuestionamiento de la funcionalidad y el rendimiento de las instituciones universitarias, junto con un incremento de las expectativas de la sociedad respecto a la actuación y los servicios de las universidades públicas y una mayor exigencia de los diferentes usuarios de dichos servicios. Se observa también una internacionalización gradual del sistema universitario, una creciente competitividad de las instituciones de educación superior y un aumento de la relación entre las universidades y muchas organizaciones públicas y privadas.

Por ello, se detecta actualmente en la universidad española una preocupación por mantener y aumentar la calidad de la docencia, de la investigación y, en general, la de todos los servicios que presta. Las numerosas actividades relacionadas con la calidad universitaria que se han emprendido en los últimos años son una prueba de ello, en especial los procesos de evaluación interna y externa que se llevan a cabo en dichas instituciones en el marco del Plan Nacional de Evaluación de la Calidad de las Universidades.

El desarrollo de un plan de evaluación coherente, tendrá unas consecuencias inmediatas para la Universidad, en la medida en que le ayuda a tomar conciencia de sus posibilidades, limitaciones y oportunidades.

La evaluación de la Universidad constituye una herramienta clave de las políticas universitarias, pues permite conocer la calidad de los programas de enseñanza ofertados,

detectar sus posibles carencias e impulsar estrategias de acción para conseguir mejoras concretas.

Las Universidades Españolas, entre ellas la Universidad Politécnica de Valencia, son conscientes de que la mejora de la calidad de las instituciones universitarias constituye el reto más importante en los próximos años y el instrumento potencial para ello lo constituye la evaluación institucional en general y la evaluación de la actividad docente en especial.

La evaluación de la función docente, desde la perspectiva de las políticas educativas, constituye una herramienta útil, al permitirnos conocer el grado de cumplimiento de su eficiencia. A través de ella, se pone de relieve la calidad del trabajo docente desarrollado en el seno de la institución. También se estimula la mejora de la actividad profesional, al permitir al profesorado tomar conciencia de las posibilidades y debilidades profesionales.

Existen distintas formas de efectuar la evaluación docente. No obstante, tal y como se desprende del estudio efectuado por Tejedor (1991) sobre los Planes de Evaluación en las Universidades Españolas: “el instrumento de evaluación de la actividad docente más generalizado en las universidades españolas es el cuestionario de opinión de los estudiantes en el que los alumnos puntúan sobre una escala de opinión diferentes aspectos docentes del profesor”.

La evaluación de la actividad docente mediante el cuestionario de opinión de estudiantes, ha sido extensamente investigada en educación, existiendo opiniones muy dispares sobre las conclusiones que pueden desprenderse mediante este sistema. En este sentido, El-Hassan (1995) señala que los temas que pueden afectar a la validez de las puntuaciones obtenidas en el cuestionario son fundamentalmente la dimensionalidad, validez, fiabilidad, generalidad de las puntuaciones y la investigación de los errores potenciales que pueden afectar la validez de estas puntuaciones.

El Consejo de Universidades, en su resolución de 20 de junio de 1990, estableció los criterios generales de la actividad docente del profesorado universitario. La Comisión de Gobierno de la UPV, constituyó una Comisión de Evaluación de la Docencia, con la

finalidad de evaluar al profesorado en condiciones objetivas, tal y como se establece en el Plan Nacional de Calidad de las Universidades. Los instrumentos de evaluación establecidos por dicha comisión son cuatro: el cuestionario de los alumnos, los autoinformes, el informe del director de Centro y el informe del director del Departamento, siendo el más relevante de todos ellos las encuestas de opinión de los estudiantes.

En la UPV, el empleo de encuestas de opinión de alumnado sobre la actuación docente del profesorado, se remonta al estudio llevado a cabo sobre “Evaluación de Competencias Docentes del Profesorado de la UPV, por el Instituto de la Ciencias de la Educación (ICE), en el marco del Plan Nacional de Investigación Educativa correspondiente al año 1981, antes de que la Ley de Reforma Universitaria prescribiese el pase de encuestas.

En el curso académico 1987-1988, el Rectorado de la UPV encargó al ICE la elaboración de un procedimiento que organizase y regulase la aplicación de dichas encuestas. A tal efecto, se creó un grupo de trabajo integrado por profesores, directivos de los distintos departamentos, representantes del alumnado y técnicos del ICE que elaboraron un conjunto de propuestas que sirvieron de punto de referencia en los procesos de evaluación de la actuación docente del profesorado. Finalmente, en julio de 1994, el ICE presentó a la Junta de Gobierno el documento “Evaluación del Profesorado”, aprobándose por unanimidad la encuesta institucional que viene aplicándose desde el curso 1994-1995.

Inicialmente, en el documento presentado a la Junta de Gobierno se hacía hincapié en que la principal garantía para obtener un aprovechamiento eficaz de los resultados en estos procesos de evaluación docente del profesorado, era plantearlos desde una perspectiva formativa, abierta, flexible y dirigida fundamentalmente a la mejora. Sin embargo, actualmente y dentro del marco del Plan Nacional de la Calidad de la Universidades Españolas (PNECU) y posteriormente dentro del marco del II Plan de Calidad de las Universidades (PCU), se ha generalizado la utilización de los resultados con fines sumativos y para la toma de decisiones administrativas (promociones, reconocimiento de tramos docentes, etc.), lo cual plantea la necesidad de verificar la adecuación del instrumento de medida utilizado.

Por otro lado, la UPV, al acogerse al II Plan de Calidad de las Universidades debe continuar y profundizar en la implantación de sistemas y planes institucionales de calidad, lo que propiciará la introducción paulatina de nuevos indicadores de las actividades y servicios de las universidades y, a su vez, el desarrollo de un proceso de evaluación institucional en dos fases: una evaluación interna y una evaluación externa. Además, siguiendo lo establecido en la guía de autoevaluación del Consejo de Universidades sobre el PCU, la evaluación de la enseñanza deberá utilizar un enfoque centrado sobre la Titulación, integrando dentro de esta unidad de análisis, todos los aspectos fundamentales que están relacionados con la planificación, ejecución y evaluación de los procesos de enseñanza-aprendizaje.

En este contexto, el Comité de Calidad de la UPV consciente de que hasta el momento el indicador más utilizado para conocer el estado de calidad de la docencia en la Universidad y obtener valoraciones individuales del rendimiento docente de todo el profesorado es el cuestionario de opinión de estudiantes y, considerando la relevancia que está tomando este instrumento, plantea la necesidad de desarrollar una metodología para el análisis e interpretación de los resultados obtenidos.

A partir de estas inquietudes arranca el presente trabajo. Se pretende, por un lado, efectuar un proceso de revisión del instrumento de manera que se verifique si se está utilizando realmente para el fin que se había diseñado y, por otro lado, desarrollar una metodología estadística adecuada para extraer, analizar e interpretar la información contenida en el mismo con la finalidad de optimizar su utilización práctica.

1.2. Enfoques de Evaluación de la Docencia Universitaria

En el ámbito universitario podemos encontrar distintos enfoques de la Evaluación de la Actividad Universitaria. Nos parece interesante resaltar aquí, la diferenciación establecida por Jornet (1991), teniendo en cuenta tres aspectos desde los que situar los principales enfoques de la evaluación universitaria:

- Control de evaluación.

- Objeto de la Evaluación.
- Finalidad de la Evaluación.

A partir de estos aspectos, los tres grandes enfoques definidos por Jornet (1991) y Jornet et al. (1996) serían:

- Institucionales.
- De Programa.
- De Profesorado.

En cada uno de estos enfoques, debemos distinguir a su vez dos perspectivas dentro de la evaluación de la actividad Universitaria, que están profundamente interrelacionadas:

- La evaluación sumativa, cuya finalidad es la sanción administrativa, bien sea de carácter positivo o negativo.
- La evaluación formativa, que pretende facilitar a cada uno de los miembros en el proceso estudiado, un factor de información que permita encaminarnos a un rumbo de mejora (Rotger, 1990).

De los tres grandes enfoques definidos por Jornet (1991) y Jornet et al. (1996), la evaluación del profesorado constituye la herramienta fundamental para lograr la calidad docente, en opinión de la mayoría de autores revisados. En este sentido, De la Orden (1987), señala que “el profesor no es un componente del sistema, sino que constituye el factor esencial en la coordinación operativa de todos los demás”.

1.3. Fuentes y Formas de Evaluación de la Docencia Universitaria

Existen diversas formas de evaluar la docencia universitaria, cada una de ellas con aspectos a favor o en contra, que pueden condicionar su utilización.

Según De la Orden (1992), las formas de evaluación de la docencia universitaria más utilizadas son: el juicio de los estudiantes, la opinión de colegas y superiores académicos, la medida del rendimiento académico de los alumnos y la productividad investigadora, tanto en términos de calidad como de cantidad.

Rippey (1981) establece tres categorías en las mediciones de la docencia: percepciones, procesos y productos. Cada una de ellas puede considerarse como se define a continuación:

- Percepciones: Auto-encuestas, evaluaciones por iguales, encuestas a estudiantes y evaluaciones administrativas.
- Proceso: Experiencia en la materia, habilidades pedagógicas, carisma-estimulación, popularidad, civismo, empatía, esfuerzo y juicio.
- Producto: Salidas cognitivas, afectivas y motoras.

Mateo (1987) señala las siguientes fuentes de información en la evaluación del profesorado universitario, a la vez que realiza un estudio comparativo de sus cualidades, defectos, condiciones para su uso, tipo de evidencia y propósito:

- Evaluación por parte de los estudiantes
- Rendimiento de los estudiantes como medida de la competencia del profesor
- Observación en clase
- Evaluación por parte de los colegas
- Autoevaluación
- Evaluación mediante equipos de expertos
- Evaluación a partir de modelos de enseñanza

Escudero (1991) señala las siguientes fuentes:

- Los estudiantes
- Los colegas del departamento y/o facultad

- Autoinformes
- Administradores
- Directivos académicos
- Materiales curriculares
- Registros académico-administrativos
- La clase como unidad de análisis
- Los expertos externos (académicos y pedagógicos)
- Los exalumnos:
 - Grupos mixtos
 - Simulaciones controladas

De igual forma, este autor reseña los procedimientos para obtener la información:

- La observación directa
- La observación en vídeo
- Registros de resultados académicos
- Pruebas de rendimiento
- Pruebas de competencia (también a profesores)
- Test de todo tipo
- Escalas diversas
- Comentarios escritos
- Entrevistas (individuales y grupales)
- Debates (grupos de discusión)

Sin embargo, como ponen de manifiesto diversos autores (Arubayi, 1986; Moses, 1986; Rutherford, 1987; Centra, 1988; Escudero et al., 1996; García Ramos y Congosto, 2000), la evaluación del profesor por parte de los alumnos se ha convertido en una de las fuentes de información más importantes sobre los docentes y, en consecuencia, una de las fundamentales sobre las que se apoya la evaluación del sistema educativo, al ser el colectivo del profesorado uno de los pilares fundamentales en los que se apoya la calidad de las universidades (Mateo et al., 1996a).

Las evaluaciones de los profesores mediante opinión de los estudiantes han sido extensamente estudiadas en educación, siendo uno de los campos más proliferos de investigación. Sin embargo, como señalan Dowell y Neal (1982), estos estudios se caracterizan por ser extensos, contradictorios y de calidad muy dispar. En los últimos años se ha seguido investigando en este campo, aunque el número de estudios que han apareciendo es mucho menor, pese a que en la actualidad pueden aplicarse en los análisis, métodos estadísticos más complejos, novedosos y apropiados.

En este sentido, se plantea la necesidad de seguir investigando en el campo de la evaluación docente, especialmente en relación a los cuestionarios de opinión de los alumnos, con la finalidad de definir una metodología para analizar e interpretar los resultados obtenidos mediante este procedimiento, de manera que la información extraída de los mismos, pueda ser utilizada por la universidad para establecer actuaciones de mejora continua.

2. Objetivos

De acuerdo a lo expuesto anteriormente, el objetivo fundamental del presente trabajo es analizar la adecuación del Cuestionario de Evaluación de la Docencia a partir de la Encuestas de Opinión de los Estudiantes de la Universidad Politécnica de Valencia y desarrollar una metodología estadística adecuada para extraer, analizar e interpretar la información contenida en el mismo con la finalidad de optimizar su utilización práctica.

Para ello, se ha estructurado el trabajo en tres grandes fases, cada una de las cuales presenta unos objetivos específicos y unas metodologías particulares.

En primer lugar se aborda el problema de evaluar la dimensionalidad del cuestionario para conocer la estructura subyacente del mismo, con la finalidad de extraer y verificar la información contenida. La técnica multivariante a utilizar para identificar las dimensiones subyacentes existentes en el cuestionario es el Análisis Factorial Exploratorio. A continuación, se utiliza el Análisis Factorial Confirmatorio como técnica para verificar que el cuestionario capta las dimensiones básicas, que se pretenden evaluar.

En la segunda fase, se pretende determinar la capacidad del Cuestionario para identificar perfiles diferenciados de profesorado a partir de la calidad percibida por los estudiantes. Para ello se emplea el análisis cluster jerárquico, no jerárquico y en dos etapas, como técnicas multivariantes de clasificación. A continuación, se procede a identificar los ítems del cuestionario que mejor discriminan entre las tipologías docentes obtenidas.

En una tercera fase, tras haber obtenido los distintos grupos homogéneos mediante la aplicación del análisis de conglomerados, en función de las puntuaciones medias obtenidas en los ítems, se procede a identificar tipologías docentes en función de diversas características descriptivas referentes a la asignatura y al profesor. Para ello, se utiliza la técnica de los árboles de decisión. Una vez identificadas estas tipologías, se realiza un nuevo análisis discriminante para identificar los ítems que mejor caracterizaban cada una de ellas. Finalmente, se aborda un estudio para conocer el poder discriminante de los distintos ítems entre las tipologías docentes identificadas, de manera que podamos identificar

aquellos que mejor discernen entre las posibles categorías obtenidas. Para ello se utilizará el método de clasificación SIMCA (Soft Independent Modelling of Class Analogy).

Estos objetivos generales se concretan en los siguientes objetivos específicos:

- Extraer las dimensiones subyacentes evaluadas mediante el cuestionario.
- Analizar la adecuación de los ítems incluidos en el cuestionario en relación a las dimensiones que se desean analizar.
- Utilizar el análisis clúster con fines exploratorios para identificar grupos homogéneos de calidad docente.
- Simplificar las observaciones originales en conglomerados de características definidas y perfiladas.
- Identificar la relación existente entre las observaciones a partir de los conglomerados definidos y de la estructura subyacente existente en los datos.
- Determinar las variables que son relevantes para caracterizar los objetos a agrupar.
- Segmentar los datos iniciales y establecer grupos homogéneos en función de determinadas variables predictoras, utilizando la técnica de árboles de clasificación.
- Determinar las variables o ítems del cuestionario que tienen mayor relevancia en cada uno de los grupos homogéneos obtenidos tras la segmentación y clasificación óptima de la variable dependiente, utilizando el análisis discriminante dentro de cada uno de los grupos establecidos.
- Determinar las variables o ítems incluidos en el cuestionario que mejor discriminen entre las tipologías docentes obtenidas. Para ello se utilizarán los valores del poder discriminante y poder de modelización de cada uno de los modelos creados para la aplicación del método de clasificación SIMCA.

Cabe resaltar aquí, que siempre que se plantea el tema de la calidad de la enseñanza universitaria, inevitablemente se asocia a evaluación, de ahí que los profesores

universitarios siempre hayan estado sometidos a procedimientos de evaluación más o menos formalizados. Hasta hace poco tiempo, las prácticas evaluativas utilizadas en nuestras instituciones se limitaban mayoritariamente a la recogida de la opinión de los alumnos sobre la actuación docente en el aula mediante cuestionarios. Sin embargo, en las últimas décadas, conscientes de que las tareas y funciones del profesor son muy diversas y son muchos los factores implicados en cada una de ellas, no resulta adecuado utilizar un único procedimiento para recoger la información. Al contrario, la calidad de una estrategia evaluativa depende de la variedad y adecuación de las fuentes, agentes, procedimientos e instrumentos utilizados, lo que implicaría una serie de limitaciones en la utilización del Cuestionario de Evaluación Docente mediante la Opinión de los Alumnos (CEDA) como instrumento de medida de la calidad docente, siendo necesario complementar esta herramienta con otras técnicas, aspecto en el que coinciden todos los autores que, como se verá en el capítulo 3, se han aproximado al tema y, tal y como establece, por ejemplo, la Agència per a la Qualitat del Sistema Universitari a Catalunya (AAVV, 2002): “será un elemento más del procedimiento, ...nunca el único, ni si quiera el básico”.

En cualquier caso, conviene no olvidar que la evaluación del profesorado mediante el CEDA es uno de los principales Indicadores de Calidad exigidos a las universidades, particularmente en la coyuntura de la convergencia europea siendo necesario establecer las herramientas estadísticas necesarias para el análisis, evaluación e interpretación de los resultados que puedan desprenderse del mismo.

Para terminar, debe remarcarse el carácter eminentemente práctico de los objetivos propuestos. Se pretende finalmente, proporcionar recomendaciones prácticas sobre las variables más adecuadas para la medida de la Calidad percibida por los alumnos en la UPV, y los métodos multivariantes más adecuados para tratar este tipo de información.

2.1. Contenido de la memoria

Esta memoria esta organizada en tres partes. En la primera parte, constituida por los capítulos 1, 2 y 3, se expone el problema que ha generado la investigación, se formulan los objetivos propuestos y se analiza, finalmente, los antecedentes bibliográficos relativos a la evaluación del profesorado a partir de la opinión de los estudiantes, ofreciendo una panorámica general del problema de la evaluación docente, de las distintas formas utilizadas para evaluar y de las posibles formas en que deben utilizarse los resultados obtenidos.

Conforme a la justificación y objetivos expuestos en el capítulo 1 y capítulo 2 respectivamente, y después de la revisión sobre los antecedentes a la investigación efectuada en el capítulo 3, este trabajo se estructura a partir de aquí en otros 4 capítulos que constituyen la segunda parte de la memoria, de la forma que se expone a continuación.

En el capítulo 4 se define el proceso de elaboración y las características del cuestionario de evaluación docente utilizado actualmente en la UPV. A continuación se describen las bases de datos que van a ser utilizadas para el estudio y el proceso de obtención de las mismas.

En el capítulo 5 se estudia la dimensionalidad del cuestionario mediante la utilización del análisis factorial con fines exploratorios y los modelos de ecuaciones estructurales con fines confirmatorios, para verificar la estructura subyacente del mismo.

En el capítulo 6 se desarrolla, en primer lugar, la aplicación del análisis clúster para la obtención de grupos homogéneos en distintos periodos evaluativos. A continuación, se utiliza el análisis discriminante para determinar las variables que caracterizan cada uno de los grupos obtenidos.

En el capítulo 7 se desarrolla la aplicación de la técnica de los árboles de clasificación y del análisis discriminante para determinar la mejor segmentación posible de los datos y las variables que caracterizan cada uno de los grupos obtenidos. A continuación, se emplea el método SIMCA para determinar las variables o ítems incluidos en el cuestionario que mejor disciernen entre las tipologías docentes obtenidas.

Comentar, que en estos tres capítulos (5, 6 y 7), dado que uno de los objetivos fundamentales de este trabajo es desarrollar una metodología basada fundamentalmente en métodos multivariantes, se dedica un apartado específico a hablar de las herramientas estadísticas utilizadas, haciendo una descripción detallada de las mismas.

Finalmente, en la tercera parte de la memoria, se formulan las conclusiones generales de la tesis, aunque en cada uno de los capítulos ya se ha realizado una exposición más detallada de las conclusiones más importantes de cada una de las etapas del estudio.

En los anexos se encuentran los resultados numéricos que, por su extensión, no se ha creído conveniente incluir en el texto.

3. Antecedentes

3.1. Marco Legal

El marco legislativo que rige y orienta el sistema educativo español está formado por la Constitución Española (1978) y seis leyes orgánicas que desarrollan los principios y derechos establecidos en ella: la Ley Orgánica reguladora del Derecho a la Educación (LODE), de 1985; la Ley Orgánica de Ordenación General del Sistema Educativo (LOGSE), de 1990; la Ley Orgánica de la Participación, la Evaluación y el Gobierno de los centros docentes (LOPEG), de 1995; la Ley Orgánica de Universidades (LOU), de diciembre de 2001; Ley Orgánica de las Cualificaciones y de la Formación Profesional (LOCFP), del 2002 y la Ley Orgánica de calidad de la Educación (LOCE), del 2002

La Ley Orgánica 6/2001 de Universidades (LOU) regula el sistema universitario, derogando la anterior Ley de Reforma Universitaria (LRU) de 1983, con el objetivo de mejorar la calidad y la excelencia del desarrollo de la actividad universitaria.

En el R.D 1947/1995 de 1 de diciembre (B.O.E. 9-12-95), el Ministerio de Educación y Ciencia establecía, por iniciativa del Consejo de Universidades, el Plan Nacional de Evaluación de la Calidad de las Universidades (PNECU) con los siguientes objetivos:

- Promover la evaluación institucional de la calidad de las universidades.
- Elaborar metodologías homogéneas para la evaluación de la calidad, integradas en la práctica vigente en la Unión Europea.
- Proporcionar información objetiva que pueda servir de base para la adopción de decisiones de las distintas organizaciones en el ámbito de su respectiva competencia.

Este Plan tenía una duración de cinco años, revisable anualmente, y se ejecutaba a través de convocatorias anuales de proyectos de evaluación institucional en los que podían participar universidades públicas y privadas. Dos actividades precedieron su implantación:

Antecedentes

en primer lugar, el Programa Experimental de Evaluación de la Calidad del Sistema Universitario (1992-1994), cuyo objetivo fue poner a prueba una metodología de evaluación institucional inspirada en las experiencias internacionales, y en segundo lugar, el Proyecto Piloto Europeo de (1994-1995), que reunió en una misma orientación metodológica la evaluación de la enseñanza superior de 17 países participantes.

El PNECU establece dos tipos de proyectos de evaluación por los que las universidades pueden optar:

- Proyectos globales: abarcan una o varias universidades, consideradas en su conjunto.
- Proyectos temáticos: se refieren a una titulación o a un conjunto de titulaciones del mismo campo científico-docente.

Posteriormente se estableció un nuevo tipo de proyecto: acciones especiales destinadas a la creación de una unidad técnica de evaluación en la universidad correspondiente y a la elaboración de metodologías de evaluación o acciones de mejora de la calidad.

El Real Decreto establece la metodología que deben seguir los proyectos de evaluación. Debe ser una metodología mixta, de autoevaluación y evaluación externa. Esta metodología, junto con los protocolos y tablas de datos para el PNECU, se recogen en la Guía de Evaluación del Consejo de Universidades.

Tras finalizar el periodo de vigencia del PNECU, se estableció el II Plan de la Calidad de las universidades mediante el Real Decreto 408/2001 el 20 de abril 2001 que tiene una vigencia de seis años. Su voluntad explícita de fomentar la implantación de sistemas de calidad en la institución universitaria que aseguren la mejora continua, sigue la línea iniciada por el anterior Plan Nacional de Evaluación de la Calidad de las Universidades y, pone su énfasis en la transparencia y la información al ciudadano y, en consonancia con los países europeos, abre una vía hacia la acreditación de las titulaciones.

El PCU ha sido establecido con los siguientes objetivos:

- Continuar con la evaluación institucional y fomentar la implantación en las universidades de sistemas de calidad integral para la mejora continua.
- Promover la participación de las Comunidades Autónomas en el desarrollo y gestión del Plan, propiciando la creación de agencias autonómicas dedicadas a tal fin, con el objeto de crear una Red de Agencias de la Calidad Universitaria coordinada por el Consejo de Coordinación Universitaria.
- Desarrollar metodologías homogéneas con las existentes en la Unión Europea, que permitan establecer estándares contrastados para valorar la calidad alcanzada.
- Implantar un sistema de información a las universidades, a las administraciones públicas y a la sociedad, basado en la evaluación por resultados y apoyado en un catálogo de indicadores, que pueda servir de base para la toma de decisiones en el ámbito de sus respectivas competencias.
- Establecer un sistema de acreditación de programas formativos, grados académicos e instituciones que permita garantizar la calidad en conformidad con estándares internacionales, abarcando también programas de doctorado y formación de postgrado.

La UPV concurrió al II Plan de la Calidad de las Universidades con un Proyecto de Calidad Plurianual, en donde se presentó la planificación de las evaluaciones y revisiones de resultados para las seis convocatorias que establece el Plan (sexenio 2001-2006). Estas evaluaciones afectan a la totalidad de titulaciones que componen la UPV que cumplen con los requisitos establecidos.

La tendencia y evolución del entorno internacional y de las directrices marcadas por este II Plan de la Calidad de las Universidades, conllevan a que la evaluación se centre en la titulación, pero de una manera integral, comprendiendo tanto los aspectos docentes como los de investigación y gestión, directamente relacionados con la enseñanza. Se podría decir que es éste un escalón previo a la acreditación de programas formativos a través de la evaluación del proceso enseñanza-aprendizaje y todos los parámetros que en éste influyen.

Antecedentes

La Universidad Politécnica de Valencia, que participó en el anterior PNECU desde la 1ª convocatoria con un total de 41 unidades, en la 1ª convocatoria del II PCU con 3 unidades y en la 2ª convocatoria con las siguientes:

- Escuela Politécnica Superior de Gandía
 - Titulación de Turismo
 - Las titulaciones de Ingenierías Técnicas de Telecomunicación
- Escuela Politécnica Superior de Alcoy
 - Titulación de Administración y Dirección de Empresas
 - Las titulaciones de Ingenierías Técnicas Industriales
- Centro de Transferencia de Tecnología
- Centro de Formación de Postgrado
- Biblioteca General

Además, se está realizando la revisión de los resultados de evaluaciones anteriores de las siguientes unidades:

- Escuela Técnica Superior de Ingeniería del Diseño
- Departamento de Ingeniería Textil y Papelera
- Departamento de Ingeniería Química y Nuclear

En la figura 3.1 se muestra la evolución de las unidades participantes en cada una de las convocatorias del PNECU y en las primeras del PCU.

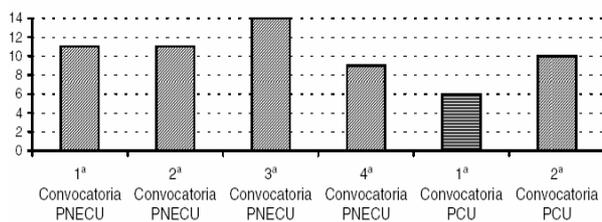


Figura 3.1. Evolución del nº de unidades participantes en el PNECU / PCU

El II Plan de la Calidad de las Universidades introdujo en la UPV, como ya se ha comentado a nivel general, nuevos objetivos con respecto al anterior Plan:

- Continuar con la evaluación institucional, fomentando la implantación en las universidades de sistemas de calidad para la mejora continua.
- Promover la participación de las Comunidades Autónomas (CCAA) con el objeto de crear una Red de Agencias de la Calidad coordinada por el PCU.
- Continuar con el desarrollo de metodologías homogéneas con las existentes en la Unión Europea, que permitan establecer estándares contrastados para la acreditación de la calidad alcanzada.
- Implantar un sistema de información basado en la evaluación por resultados y apoyado en un catálogo de indicadores.
- Establecer un sistema de acreditación de programas formativos, grados académicos e instituciones que permita garantizar la calidad.

En este contexto, el desarrollo de una metodología que permita extraer y analizar la información obtenida en los instrumentos utilizados para evaluar, permitirá optimizar la implantación gradual de una cultura de calidad en la UPV, identificar las debilidades y articular planes de mejora orientados a superarlas.

3.2. La Evaluación del Profesorado por parte del Alumnado

El énfasis actual en la evaluación se manifiesta de forma inequívoca en el sector universitario. En efecto, en la última década, la preocupación por la evaluación universitaria (profesores, instituciones y programas de postgrado fundamentalmente) constituye un rasgo esencial de la educación superior en los países más desarrollados. En España, la evaluación del profesorado universitario es, actualmente, una práctica generalizada y la tendencia a converger con las corrientes europeas y americanas en su preocupación por la calidad de la educación se refleja nítidamente en la multiplicación de congresos nacionales e

internacionales y en las publicaciones pertinentes. La literatura al respecto, es muy extensa y, a veces, contradictoria siendo conveniente realizar un repaso de la misma.

La evaluación del profesorado es un campo que ha sido investigado desde muy antiguo: ya en el siglo XIX, Kratz (1889), publicaba un trabajo “Characteristics of the Best Teachers as Recognized by Children” sobre la efectividad de los profesores a partir de la opinión de estudiantes, que se puede reconocer como pionero en el tema (Good y Murray, 1990). A principio de siglo, también aparecen algunos trabajos sobre docencia aunque es a partir de los años veinte, cuando comienzan los estudios sobre la docencia en las instituciones de educación superior. Los temas que abarcan estos trabajos son de muy diversa índole, desde las escalas utilizadas en los instrumentos de medida hasta los posibles factores exógenos que pueden influir en las valoraciones obtenidas.

En relación a las escalas de valoración, la literatura es muy abundante. Marsh (1987a), señala que los programas de evaluación del profesorado mediante opinión de los alumnos se introdujeron, entre otras, en Harvard, la Universidad de Washington, la Universidad de Purdue y la Universidad de Texas a mediados de los años veinte. Cook (1989), Marsh (1987b) o McKeachie (1990) indican que la primera escala para estudiantes, publicada en 1927, fue la Purdue Rating Scale of Instruction de Remmers, aunque Good y Murray (1990) señalan que fue Elliot en 1915 el primero que establece un instrumento para medir la efectividad docente. Estos últimos autores indican que Elliot en 1915, presentó una escala de valoración que incluía 42 rasgos que fueron seleccionados en investigaciones previas y que fueron consideradas esenciales para la docencia efectiva. Los juicios se presentaban de forma numérica y se obtenía una puntuación total sumando las puntuaciones de cada rasgo. También señalan que Boyce en 1915, desarrolló la escala Boyce Card, muy popular durante los años veinte y treinta, que incluía 5 secciones (características personales, características sociales y profesionales, gestión escolar, técnicas de enseñanza y resultados) y 44 aspectos o cuestiones de evaluación. Por otro lado, establecen que otro tipo de escala fue elaborado para la automejora por Rugg, en 1920 y fue una de las primeras de su clase, muy elaborada en comparación con otras escalas de autoevaluación. Los mismos autores indican un segundo tipo de escalas de valoración, que Barr en 1931, denomina Escalas de Calidad, que producían una descripción más cualitativa que cuantitativa de la docencia,

utilizando una escala de 3 a 5 puntos y que requería que el opinante informara de la presencia/ausencia de los rasgos o del grado en que estaban presentes. Estas escalas, más objetivas, fueron también denominadas “checklists” e incluían checklist generales de las características del profesor; checklist de actividades y listados de ítems estándares que consisten en listados de actividades específicas de profesor-alumno, en las que un determinado observador anotaba su presencia o ausencia. Sin embargo, el contenido de estos cuestionarios distaba de ser uniforme, tanto en el número de ítems como en las características o respuestas individuales.

Los trabajos que han ido apareciendo sobre diversos temas relacionados con los cuestionarios de evaluación docente han sido numerosos desde principios del siglo XX, entre los que destacaremos: Remmers (1931, 1934) en la Universidad de Washington, investigaron los cuestionarios de estudiantes en los años treinta y cuarenta; Barr (1948) cita 138 estudios sobre eficacia docente escritos entre 1905 y 1948; Doyle (1983) indica un patrón cíclico en la actividad investigadora con una mayor actividad en la década inicial y en la década de los setenta; Wolf (1990a), resume 220 estudios sobre evaluación de estudiantes de la efectividad docente que fueron escritos entre 1968 y 1974.

La utilización de los cuestionarios comenzó a difundirse en distintas instituciones, oscilando su uso, como se comentará con mayor detalle posteriormente, entre lo formativo (que incluía el desarrollo docente) y lo sumativo (que jugaba un papel importante en la toma de decisiones). Durante la década de los cincuenta fueron desarrollándose miles de cuestionarios de evaluación del profesorado. Sin embargo, como señalan Tejedor et al. (1988), el aspecto de rendición de cuentas no se utilizaba en las Universidades Españolas, debido al sistema de acceso a la docencia por oposición vitalicia, aunque en la actualidad casi todas las universidades tienen un sistema de evaluación propio.

La investigación sobre la evaluación de estudiantes es fundamentalmente un fenómeno de los setenta y los ochenta, aunque como ya se ha señalado, Remmers inició el primer programa de investigación sistemática en este campo en 1928 y debe ser señalado como el padre de la investigación sobre evaluación por estudiantes de la efectividad docente (Marsh, 1987a; Cook, 1989; McKeachie, 1990). En la década de los veinte,

Remmers publicó su escala Purdue multirrasgo y propuso tres principios para el diseño de estos instrumentos (Marsh, 1987a):

- El listado de rasgos debe ser lo suficientemente corto para evitar efectos de halo, es decir, juzgar en base a un solo rasgo, y evitar que el estudiante se aburra.
- Los rasgos deben ser añadidos por expertos como los más importantes.
- Los rasgos deben ser susceptibles a la observación y juicio de los estudiantes.

Marsh (1987a) indica que en los estudios de Remmers en 1928 ya se examinan aspectos de la fiabilidad y validez, los efectos de halo, los errores, la relación entre calificaciones de curso y encuestas a estudiantes, etc. Este autor resume algunos de los aspectos más relevantes de este periodo:

- Remmers (1931,1934) fue el primero en reconocer que la fiabilidad de las encuestas a estudiantes puede estar basada en el acuerdo de diferentes estudiantes sobre el mismo profesor y que la fiabilidad de la respuesta media varía en el número de estudiantes y varía en forma análoga a la relación entre longitud del test y fiabilidad del mismo en la fórmula de Spearman-Brown.
- Remmers publicó el primer análisis factorial de respuestas medias de los estudiantes con sus 10 rasgos e identificó dos rasgos de orden superior que denominó Empatía y Madurez Profesional (Smalzried y Remmers, 1943; Creager, 1950).
- En 1949, Remmers, Martin y Elliot, encontraron que cuando los estudiantes se asignan aleatoriamente a diferentes secciones del mismo curso, el rendimiento medio de la sección corregido por la aptitud inicial mostraba una correlación positiva con las encuestas medias de clase sobre la efectividad docente, proporcionando así una base para el paradigma de validez multisección.
- En 1950, Drucker y Remmers encontraron que las encuestas de los alumnos diez años después de su graduación en la Universidad de Purdue, estaban correlacionadas sustancialmente con las encuestas de estudiantes actuales, en el

caso de los profesores comunes que habían tenido los dos grupos. Los exalumnos y los actuales estudiantes mostraron también un gran acuerdo sobre la relativa importancia que pusieron sobre los 10 rasgos de la Escala Purdue.

- En el primer estudio a gran escala multiinstitucional, Remmers et al. (1949) y Elliot, (1950) correlacionaron las respuestas de los estudiantes de 14 colleges y universidades con una amplia variedad de variables (por ejemplo, sexo, rango, habilidad escolar, años en la escuela,...). Aunque se encontraron algunas relaciones, los resultados sugieren que las características demográficas tienen poco o ningún efecto sobre las encuestas.

Como señalan Avi-Itzhak y Kremer (1986), a partir de la década de los setenta el control y la evaluación del profesorado se convirtieron en aspectos importantes de la administración educativa. Hubo un aumento creciente de la burocratización en la escuela pública lo que hizo que se enfocara el estudio de la efectividad docente y el desarrollo de estándares para el profesorado. El incremento del salario del docente y la creciente demanda de una rendición de cuentas por parte del mismo hicieron posible el desarrollo de la evaluación del profesorado (Good y Murray, 1990).

Por su parte, Seldin (1989) tras efectuar diversos estudios entre 1983 y 1988 afirma que:

- La docencia en el aula es la consideración más importante en la evaluación del rendimiento general del profesorado.
- Las encuestas sistemáticas a estudiantes son la segunda fuente de información más importante en la determinación del rendimiento docente en el aula.
- Los comités de Facultad son cruciales en la evaluación docente.
- La autoevaluación ha tomado un considerable soporte, y las visitas al aula han ganado importancia significativa.
- Desde 1983, únicamente ha habido cambios limitados en la evaluación del rendimiento general, pero han tenido lugar cambios considerables en la evaluación de la docencia en el aula.

Debido a la creciente utilización de los cuestionarios de evaluación del profesorado por parte de los estudiantes en casi todas las universidades del mundo, se han ido generando infinidad de estudios sobre los mismos y sobre la docencia en los niveles medio y superior, publicándose multitud de trabajos, sobre todo en Estados Unidos. Sin embargo, como señala Marsh (1987a), casi toda la investigación se ha centrado en evaluaciones de la docencia de colleges/universidades, mientras que una parte muy reducida de esta investigación ha sido dirigida hacia el nivel precollege, aunque en general resulta difícil extrapolar los resultados de uno a otro nivel (Good y Brophy, 1986). Estas investigaciones se situaron inicialmente en USA y Canadá, aunque en los últimos años han ido desarrollándose paulatinamente en todo el mundo (Mahmoud, 1991; Watkins y Akande, 1992; Borich y Madden, 1997). Son interesantes en este sentido las revisiones de los trabajos publicados realizadas por Costin et al., (1971), Feldman (1978) y Marsh (1984; 1987a). Para dar una idea, Marsh (1987a y b), indica que el descriptor "student evaluation of teacher performance" fue introducido en el sistema ERIC (*Educational Resources Information Center*), que es la base de datos educativa más grande y diversa del mundo, con más de un millón de records que cubren todos los campos y niveles de la educación, en 1976. Entre 1976 y 1984 hubo 1055 estudios con esta etiqueta y aproximadamente la mitad aparecieron desde 1980. De igual forma, entre 1982 y septiembre de 1993 se han identificado 1013 trabajos con el descriptor Student Evaluation of Teacher Performance (SETP) en el ERIC, siendo 133 referidos a validez.

En general, se pueden identificar dos líneas en los trabajos publicados, en función de su adecuación metodológica: una línea caracterizada por una alta exactitud metodológica y otros en los que no se requiere ese grado de precisión. Todos los estudios no pueden obtener una única valoración y Aleamoni (1981) indica que las valoraciones sobre las encuestas de evaluación oscilan entre "fiables, válidas y útiles" a "no fiables, no válidas e inútiles".

De entre todos los trabajos desarrollados a nivel internacional, cabría destacar los desarrollados por autores como Aleamoni, 1981; Arubayi, 1986; Braskamp et al., 1984; Cohen, 1980; Costin et al. 1971; Doyle. 1975, 1983; Feldman, 1976, 1977, 1978, 1979, 1983, 1984; Kulik y Kulik, 1974; Marsh, 1980, 1982a, 1982b, 1984, 1987a, 1987b;

McKeachie, 1979, 1986, 1990; McKeachie et al., 1980; Miller, 1971; Murray, 1980; Overall y Marsh, 1980, 1982; Orr, 1972; Remmers, 1963; Wolf, 1974; etc.

En nuestro país, podemos centrar las primeras experiencias de evaluación del profesorado por encuestas a estudiantes en la década de los ochenta. Como señala Tejedor (1991), a partir de la experiencia inicial de la Universidad Autónoma de Madrid en el curso 81-82, las universidades que mostraron con más fuerza su interés por el tema fueron las de Santiago, Zaragoza, Barcelona, Valencia, Cantabria y la Autónoma de Madrid, incorporándose con posterioridad otras (Complutense, Politécnica de Madrid, Oviedo, País Vasco, Granada, Málaga, Alicante, Extremadura, Murcia, Sevilla,...), de forma que casi todas las universidades españolas han puesto en marcha el proceso de evaluación del profesorado, centrandó la estrategia de recogida de información en el alumnado.

El mismo autor establece que el instrumento generalizado en la evaluación del profesorado en la universidad española son los cuestionarios de opinión.

Tejedor (1991) señala algunas de las circunstancias ocurridas en la Universidad española durante los últimos años y que considera que han contribuido a perfilar las condiciones actuales de la evaluación de la enseñanza universitaria:

- Aprobación de la LRU y posteriormente de la LOU.
- Establecimiento del PNECU y PCU.
- Realización de las primeras experiencias evaluativas, de carácter exclusivamente formativo.
- La constitución de las Juntas de Personal, lo que supuso la preocupación de los sindicatos por el tema y, en consecuencia, un cambio en la orientación básica del proceso, al anteponerse ahora criterios sumativos a los estrictamente formativos. Este planteamiento culmina con la aprobación en agosto de 1989 del Decreto sobre Retribuciones (BOE, 9 de septiembre) y de la Resolución que lo desarrolla, que claramente distorsionaron el sentido inicial del proceso evaluativo.
- Se ha producido una "politización" del tema.

- Comienza a evaluarse la productividad investigadora.

Además de las encuestas de evaluación docente mediante opinión de los estudiantes, es conveniente que a la Comisión Académica, encargada de realizar la evaluación, le lleguen otro tipo de informes sobre el profesorado (autoinformes, informes del departamento, del centro,...). Con todos ellos deberá tomar las decisiones que estime oportuno. La evaluación del profesorado puede así mantener su razón de ser formativa aunque se utilicen los resultados, completados con otras fuentes, para la aplicación de la normativa sobre retribuciones.

Son múltiples las aportaciones que se han realizado hasta el momento en nuestro país, como se muestra en la extensa relación que se cita a continuación: Abalde et al., 1995; Aparicio, 1991; Aparicio et al., 1982; Aparicio et al., 1994; Benedito et al., 1989; Borrell, 1995; Cajide, 1994; Centra, 1974, 1988; Centra y Creech, 1976; De la Orden, 1993; De Miguel, 1988a, 1988b, 1989a, 1989b, 1991, 1997, 1998; De Miguel et al., 1991; Escudero, 1979, 1986, 1987, 1988, 1989a, 1989b, 1991, 1996, 1999; Escudero et al., 2000; Fernández et al., 1991; Fernández Ballesteros, 1995; Fernández Díaz, 1988; Fernández Sánchez, 1988, 1992; Fernández Sánchez y Mateo, 1991, 1994; Fernández Sánchez et al., 1995, 1996; Fernández Sánchez et al., 1997; Ferrández, 1991; Ferrández, et al., 1995a, 1995b ; García Ramos, 1987, 1988, 1989, 1997; García Ramos et al., 1995a, 1995b; García Ramos y Congosto, 2000; García Valcarcel et al., 1991; González González et al., 1999; González Such, 1997; González Such, et al, 1990 y 1999; González Such et al., 1993; González Such et al., 1995; Ibáñez-Martín, 1990; Jiménez, 1985; Jornet, 1991; Jornet y Suárez, 1989, 1997; Jornet, et al. 1989; Jornet, et al., 1993; Jornet, et al., 1995; Luxán, 1998; Mateo, 1987, 1990; Mateo et al., 1991, 1992; Mateo et al., 1996a, 1996b; Mora, J.G., 1995, 1998, 1999; Muñoz, 1997; Muñoz, et al 1991; Pérez Juste, 1994, 2000; Pérez Juste y Martínez Aragón, 1989; Pérez Juste et al., 1995; Prieto, 1982; Rodríguez Espinar, 1986, 1987, 1991, 1996; Salvador, 1989; Salvador y García, 1989; Salvador, 1990; Salvador y Sanz, 1987, 1988a, 1988b; Sobrado, 1991; Tejedor, 1985, 1990, 1991; Tejedor et al., 1988; Tejedor y García Valcácer, 1996; Tejedor y Montero, 1990; Tourón, 1989; Valera y López, 2000 y Villar Ángulo, 1983a, 1983b, 1987, 1990 y 1991.

En general la mayor parte de las aportaciones están relacionadas con evaluaciones reales desarrolladas en nuestras Universidades, si bien, se centran en estudios o investigaciones sobre las características métricas de los instrumentos utilizados o sobre los modelos de evaluación. Como puede observarse, la mayor parte de estas publicaciones se producen en los años en los que Tejedor reseña la implantación y desarrollo de los sistemas de evaluación de la docencia universitaria en diferentes universidades del estado.

3.3. Métodos de Evaluación del Profesorado por parte de los Alumnos

Existen cuatro formas de recoger la información generada por los alumnos: el rendimiento, los cuestionarios de evaluación docente mediante opinión de los estudiantes, la entrevista y los comentarios abiertos. La forma más utilizada de recogida de información son los Cuestionarios de Evaluación Docente mediante opinión de los Alumnos (CEDA), por lo que centraremos el estudio sobre ellos.

La mayoría de los cuestionarios piden a los alumnos que evalúen al profesor en determinados rasgos que se consideran relevantes para la enseñanza, incluyendo en ocasiones, ítems de naturaleza abierta (Tejedor et al., 1988).

Existen diferentes recomendaciones para la elaboración de los cuestionarios (Miller, 1987):

- El cuestionario debe ser corto, no más extenso de una página. Las formas cortas son menos fatigosas para el estudiante, que debe rellenar muchos cuestionarios por semestre.
- Oportunidad para flexibilidad e individualización.
- El cuestionario debe incluir preguntas que estén dentro del panorama del curso y de la experiencia del estudiante. Hay cinco áreas en las que los estudiantes pueden responder adecuadamente a las preguntas de los cuestionarios debido a

su contacto con el profesor, cada una de ellas debería estar representada por al menos una pregunta:

- Métodos pedagógicos.
- Equidad.
- Interés del profesor por el estudiante.
- Interés del profesor por la materia.
- Juicios normativos del profesor (cuestiones globales).

- Instrucciones claras.

El cuestionario de valoración en general se basa en que el estudiante da su opinión a partir de una escala numérica. Braskamp et al. (1984) proporcionan una diferenciación en función del tipo de escala utilizada:

- Escala Global. Incluyen áreas principales de docencia encontradas mediante la investigación, normalmente mediante el análisis factorial. Aunque no se trata de una escala continua, se asume como tal. Es la más utilizada.
- Escala basada en Metas. Valora los progresos del estudiante en áreas determinadas (conocimientos, etc.). A la vez se valora al profesor en las mismas metas para poder comparar resultados, como el sistema IDEA (Hoyt y Cashin, 1977).
- Escala denominada “Sistema Cafetería”. Consiste en un banco de ítems entre los que un profesor puede elegir un porcentaje variable de ellos, en función de la institución, para su utilización fundamentalmente formativa. Utilizan este sistema por ejemplo la Universidad de Purdue, la de Michigan o la de Illinois.

Ferrández (1991) incluye la clasificación de cuestionarios que realiza Aleamoni, atendiendo al tipo de ítems (por tanto a la generalidad de su uso), no a su contenido, y que resume de la siguiente forma:

1. Cuestionario para todos los instructores en todas las asignaturas.

2. Cuestionario con ítems prefijados e ítems a elegir.
3. Cuestionario en el que se seleccionan los ítems de un grupo de ellos (Sistema Cafetería).
4. Cuestionario para un sólo profesor en una asignatura determinada. Su único fin es el formativo.
5. Cuestionario en el que se selecciona el formato y/o los ítems.

Miller (1987) reseña cuatro alternativas básicas para los sistemas de evaluación por estudiantes. Estos son:

- Un único cuestionario que se utiliza para todos los estudiantes en todos los cursos. Ello ofrece simplicidad, posibilidad de comparaciones entre unidades, imparcialidad y datos de evaluación fundamentalmente sumativa, aunque en ocasiones puede permitir flexibilidad, individualización y evaluación formativa.
- Por otro lado estaría el "no sistema", en el que hay muy pocos procedimientos o políticas para usar o reportar los resultados de la evaluación. Ello permite el máximo grado de flexibilidad, individualización y evaluación formativa, aunque permite el mínimo nivel de evaluación sumativa.
- El enfoque "cafetería" desarrollado por la Universidad de Purdue durante los setenta. Es, como se ha dicho, un catálogo de ítems del que el profesor puede seleccionar para crear un cuestionario de evaluación. Sus ventajas son la gran flexibilidad, individualización, y algunos datos estadísticos de comparación, siendo sus desventajas el coste, la complejidad y unos datos de limitado uso sumativo. Un sistema modificado del enfoque "cafetería" consiste en tener una sección estándar de ítems aplicable a todos los cursos e instructores, con una sección de ítems opcionales que son elegidos por los profesores individuales. Este enfoque permite flexibilidad, individualización y asistencia en la evaluación formativa, además de proporcionar evaluación sumativa para otros fines.

- Un cuarto enfoque lo proporciona el sistema de evaluación docente de la Universidad de Washington (IAS), en el que se elaboran seis formas diferentes para proporcionar información diagnóstica para seis tipos de cursos: cursos grandes, cursos pequeños de discusión, seminarios, cursos de resolución de problemas, cursos de adquisición de habilidades y secciones. Este enfoque permite adaptar el cuestionario a la forma docente, así pueden ajustarse necesidades sumativas de tipo organizativo. Sin embargo, no permite flexibilidad o individualización.

Los ítems están derivados de los resultados de las investigaciones sobre efectividad docente y esencialmente derivados del paradigma proceso-producto (Tejedor et al., 1988). En cuanto al grado de especificidad de los ítems, estos pueden ser:

- Globales. Poseen una función esencialmente sumativa, ítems de alta inferencia ya que el estudiante debe generalizar sobre su experiencia y suelen evaluar globalmente al profesor, al curso o a la asignatura en general. Suelen incluirse al final del Cuestionario.
- De conceptos generales. Su grado de inferencia es algo menor y se refieren esencialmente a áreas de instrucción.
- Específicos. Son preguntas concretas a temas determinados, que requieren una inferencia menor por parte del alumno. Tienen un alto componente formativo.

Las escalas utilizadas suelen ser tipo Likert de cinco puntos o alternativas respuesta, que es considerada como más sencilla de completar y más fiable, aunque puede sesgar inconscientemente al estudiante hacia una puntuación alta (Miller, 1987); no obstante, en ocasiones se utiliza la de siete puntos, que requiere un juicio más preciso por parte del estudiante. Otro aspecto a tener en cuenta en esta decisión, cinco o siete puntos, son las etiquetas verbales que adjetivan la escala (French-Lazovik y Gibson, 1984).

El cuestionario de opinión utilizado actualmente en la Universidad Politécnica de Valencia, es de escala tipo global. Incluye 19 ítems entre los que podemos encontrar ítems globales, de tipo general y de tipo específico. La escala utilizada es de tipo Likert, con 5 alternativas de respuesta más una alternativa de no sabe no contesta, como se describe posteriormente en el capítulo 4.

3.4. Utilización de los Resultados

La utilización de los resultados es un aspecto primordial en la elaboración de cualquier plan de evaluación, ya que la validez de cualquier medida depende del uso que vaya a hacerse de ella.

El Joint Committee on Standards for Educational Evaluation (1981) señala los cuatro criterios por los que las evaluaciones deben ser juzgadas: utilidad, factibilidad, propiedad y precisión. Hay diversos colectivos que dependen de los resultados de esa evaluación, especialmente en la evaluación de la docencia y de las instituciones universitarias.

La evaluación de la docencia en educación superior fue diseñada y utilizada para mejorar la docencia (Goldschmid, 1976; Rotem y Glasman, 1977). Los cuestionarios de evaluación docente fueron desarrollados inicialmente para proporcionar información a los estudiantes sobre qué asignaturas podían elegir y distinguir al profesor bueno del malo (Arreola, 1987 a y b). Las puntuaciones obtenidas por los profesores en estos cuestionarios diseñados por las asambleas de estudiantes eran publicadas y distribuidas entre los mismos para ser utilizadas como guía de matrícula. Sin embargo, a menudo los estudiantes no podían escoger al profesor mejor puntuado, o bien, no disponían de las puntuaciones de los profesores peor valorados, por deber contar para su difusión con la aprobación de los mismos.

Actualmente, la utilización de los cuestionarios abarca desde fines estrictamente administrativos hasta fines investigadores, aunque con una mayor presencia de los primeros. Inicialmente los cuestionarios de opinión fueron utilizados por los

administradores con propósitos sumativos de promoción, traslado, juicios de mérito, etc. También se propuso su utilización, para conseguir aumentos de sueldo, aunque este uso fue muy impopular entre los profesores (Good y Murray, 1990), lo que hizo que no se extendiese esta forma de cuestionarios denominados “Cuestionarios de Mérito”. A partir de este momento, los cuestionarios de opinión comenzaron a utilizarse como medida de eficiencia docente.

Aunque la utilización de los resultados de la evaluación mediante opinión de los estudiantes puede ser interpretada mediante diversos enfoques, seguiremos resaltando aquí la dualidad propuesta por Scriven (1967) de usos formativos y sumativos de la evaluación. Cuando hablamos de utilización de los resultados de la evaluación necesariamente nos tenemos que referir tanto a los aspectos formativos como a los sumativos. De igual forma, para hablar de la utilización de los resultados de la evaluación desde la perspectiva de las encuestas de evaluación (EE), debemos tener en cuenta siempre los dos polos de formación/rendición de cuentas. De esta forma, para centrarnos en la discusión de los diferentes usos que pueden tener la evaluación de los cuestionarios partiremos de la dicotomía planteada por Scriven de Evaluación formativa y sumativa. También, para diferenciar los dos niveles de actuación nos centraremos, en primer lugar, en los usos de la evaluación en general a nivel universitario, tema desarrollado en Jorner et al. (1996).

La finalidad primigenia de las encuestas a estudiantes sobre el profesor fue la de mejorar la docencia (Goldschmid, 1976; Rotem, 1977), y ha ido pasándose de una finalidad esencialmente formativa a su utilización con fines de promoción y empleo (Avi-Itzhak y Kremer, 1986).

Como señala Barber (1990), los sistemas de evaluación del profesorado no son inherentemente formativos o sumativos. Si un sistema es formativo o sumativo lo determina la utilización de los datos resultantes.

Para ser efectivo, un sistema de evaluación incluye tanto evaluación sumativa como formativa, la primera para establecer recompensas diferenciales y promoción del personal, mientras que la formativa es para proporcionar feedback diagnóstico para entrenamiento (Murray, 1984).

Como se ha comentado anteriormente, dentro de la dicotomía formativo/sumativo, se distingue entre la evaluación de Instituciones, de Programas y de Profesores, diferenciando en las dos primeras un impacto global (equipamientos, aulas, profesores), mientras que la del profesorado incidirá fundamentalmente sobre los individuos.

Los resultados se pueden utilizar por el profesorado para la mejora individual de la instrucción, por los administradores y comités de promoción y empleo para la toma de decisiones y, por los estudiantes para la elección de la asignatura.

Por su parte, Marsh (1984 y 1987b) y McKeachie (1979 y 1986) señalan la siguiente utilización de los resultados:

1. Feedback diagnóstico al profesorado sobre la efectividad de su docencia que puede ser valiosa para mejorar su docencia.
2. Medida de la efectividad docente para usar en toma de decisiones administrativas.
3. Información para estudiantes para usar en la selección de cursos y profesores.
4. Medida de la calidad del curso, para ser utilizada en la mejora del mismo y en el desarrollo del currículum.
5. Una medida o descripción del proceso para la investigación en docencia.

Gillmore (1984) señala dos usos para los resultados de las encuestas de evaluación:

1. Como evidencia de docencia pobre o como evidencia de buena docencia en disputas de empleo.
2. Como un componente de la revisión periódica del profesorado.

Ambos usos son similares, ya que estos resultados proporcionan a la administración del centro la base para tomar decisiones de tipo sumativo, es decir, poder

determinar hasta qué punto el profesor cumple con sus obligaciones docentes. Por otra parte, otra función importante de la revisión periódica de los profesores por estudiantes es proporcionar a los profesores feedback, señalando los aspectos que pueden mejorar de su rendimiento y que están puntuados por debajo del estándar (Gillmore, 1984).

Por su parte, Cruse (1987) establece que los resultados pueden ser utilizados para mejorar las puntuaciones del profesor (puntuaciones formativas), evaluar al profesor con respecto a empleo, pago y rango (puntuaciones sumativas), enseñar al estudiante (énfasis en el aprendiente), o satisfacer al estudiante (énfasis en el consumidor).

Centra (1988) indica que los usos de las encuestas a estudiantes son: contribuir a la mejora docente; usos sumativos; ayudar a los estudiantes a elegir curso y determinar qué profesores cumplen unas competencias docentes definidas como mínimas, esperadas en cualquier profesor universitario.

Aunque los cuestionarios no ofrecen una medida absoluta de la habilidad docente del profesor o de su efectividad, han sido utilizados junto con otras formas de evaluar al profesor como base para tomar decisiones sobre promoción y empleo, aunque el aspecto más importante es proporcionar feedback a los profesores sobre su efectividad y habilidades docentes, esperando incentivos e ideas para la mejora personal (McBean y Lennox, 1985).

Tejedor y Montero (1990) señalan que la evaluación del profesorado por los alumnos se orienta a la estimación del nivel de calidad de la enseñanza universitaria para contribuir a su mejora teniendo en cuenta que la calidad de la instrucción representa la medida en que ésta se adapta al contexto: medios disponibles y capacidad de los estudiantes. El objetivo básico, según estos autores, es conseguir una utilidad efectiva del conjunto del proceso como recurso de perfeccionamiento docente, encaminados a la evaluación formativa.

Otra cuestión a considerar son las variables que influyen en las valoraciones de los estudiantes (tamaño de la clase, número de alumnos, notas...). Tejedor (1991) indica que este tema ha sido estudiado en muchas ocasiones aunque no en el ámbito español, y señala la conveniencia de realizar estudios para analizar en qué medida estas variables están

condicionando los resultados de la evaluación, con el objetivo de establecer acciones correctoras de la evaluación y poder contextualizar correctamente dichos resultados además de facilitar su interpretación.

Debido a que como hemos expuesto anteriormente, la utilización de los resultados de las encuestas de evaluación mediante opinión de los estudiantes debe efectuarse bajo la dualidad propuesta por Scriven de usos formativos y sumativos de la evaluación, vamos a profundizar un poco más en cada uno de estos aspectos.

3.4.1. Utilización de los resultados con fines formativos

Uno de los primeros usos que se dio a los resultados de la evaluación del profesorado fue con fines formativos. En este sentido, es necesario conocer como utiliza el profesor la información que le llega de las encuestas, cual es el feedback que recibe y de qué le sirve.

Los resultados de la evaluación de estudiantes es una de las fuentes de feedback más importantes que reciben los profesores. Un estudio de Davey y Sell (1985) demuestra que los profesores realizan relativamente poco esfuerzo dirigido a la mejora docente o el desarrollo profesional, tienen poco contacto con los colegas en lo relacionado con su docencia, cursos o su evaluación y que la estructura del curso, exámenes y los métodos, en general, reciben pocas revisiones excepto por las encuestas realizadas por los estudiantes.

De Nevé (1991) estudia el proceso de integración de la información que el profesor recoge de las EE y muestra que el proceso de feedback de éstas es mucho más complejo que lo que indica el modelo unidimensional, ya que, la satisfacción personal con una puntuación depende del nivel individual de aspiración, del conocimiento disponible de cómo mejorar una dimensión y la disposición a implementar cualquier nueva conducta docente, enmarcándose en una teoría subjetiva.

Tejedor (1991) indica que para explicar la mejora de la actuación docente que se produce en los profesores, se hace referencia a la teoría de "disonancia cognitiva" de

Festinger, según la cual la información proporcionada por los cuestionarios pone en marcha un mecanismo de retroalimentación de forma que se produce una cierta disonancia o insatisfacción que le induce a cambiar, produciéndose este efecto de mejora en aquellos ítems de referencia más específica y más relacionados con la asignatura, aumentando dicho efecto en la medida en que el profesor acepte los resultados y las orientaciones pertinentes para la mejora de su docencia. El efecto de mejora producido por los cuestionarios se deriva, como ya se ha apuntado, de los resultados que se obtienen en ítems específicos.

No obstante, los resultados sobre feedback no muestran resultados concluyentes. En general, se acepta que el feedback siempre produce efectos de mejora (Aleamoni, 1981; Cohen, 1980; Stevens y Aleamoni, 1985). En el mismo sentido, LHommedieu et al. (1990) señalan el efecto del feedback, aunque los resultados de su revisión no los persuade de que la utilización del feedback de las encuestas sea inefectiva para evaluar y mejorar la docencia, añadiendo que la literatura presenta un efecto positivo, aunque pequeño, del feedback únicamente escrito, presentando un efecto mayor si este feedback escrito se acompaña de consultas personales.

Fernández et al. (1996) indican que la sola devolución de la información no es suficiente para obtener resultados. McKeachie et al. (1980) encuentran poco efecto únicamente de las puntuaciones de los estudiantes sobre la mejora del profesor, mientras que encuentran una mejora significativa cuando las puntuaciones se comunican personalmente, cara a cara, en una sesión de orientación; Cohen (1980) señala que comparando diez directores de curso que tuvieron asistencia de orientación, con otros profesores que no la tuvieron se encuentran diferencias significativas en seis de los diez ítems estudiados.

Tejedor (1991) señala que el profesor puede tener dificultad de incorporar los resultados de la valoración de los estudiantes a la práctica diaria, por lo que aboga por la implantación desde la administración de medidas de apoyo al profesorado en la realización de su tarea, con la planificación de actividades formativas orientadas por criterios de practicidad, flexibilidad y adaptabilidad a las necesidades detectadas y que deben suponer, continúa diciendo, incentivos en su profesionalización.

García Ramos et al. (1995) indican que el efecto formativo de la evaluación implica directamente al centro, en el sentido que se pretende ofrecer al profesorado actividades, cursos o seminarios para mejorar o reforzar su desarrollo docente.

Marsh (1987a) cita dos estudios que sugieren que el feedback, complementado con una consulta a un asesor externo, puede ser una intervención efectiva para la mejora de la efectividad docente: Marsh et al., (1975); Overall y Marsh (1980).

Otro aspecto a considerar es la retroalimentación colectiva, con resultados globales, no individuales. La literatura sobre evaluación del aula incluye un desacuerdo aparente sobre el tiempo más apropiado para recoger y devolver los datos de evaluación a los profesores. Así, Murray (1980), indica que los resultados del feedback no son perceptibles hasta que no se incorporan en la revisión del profesorado, en un sistema de evaluación.

Marsh (1987a) establece que un aspecto importante es si la información que recogemos mediante la evaluación puede servir para otros propósitos y, en particular, si los mismos datos pueden ser usados para feedback diagnóstico para el profesor y para la toma de decisiones administrativas.

Otro aspecto a considerar es la utilización de los resultados varias veces. Algunos estudios (Aleamoni, 1978b; Centra, 1974; Miller, 1971) han mostrado mejoras modestas en la docencia, en particular en los profesores que utilizan los resultados por primera vez. Sin embargo, Centra (1988) afirma que el impacto formativo de los resultados disminuye considerablemente cuando se utilizan varias veces y, en consecuencia, los resultados deben ser utilizados únicamente para decisiones sobre personal.

Otro factor a considerar es cuando se realiza el pase de las encuestas, tanto a nivel de tiempo más adecuado en el curso, como de tiempo transcurrido entre la recogida de la información y la devolución de los resultados a los profesores. En este sentido, Hofman y Kremer (1983) y Ory y Parker (1989) entre otros, indican que los datos de evaluación deberían ser devueltos lo antes posible para facilitar la realización de las mejoras docentes antes de que acabe el curso.

3.4.2. Utilización de los resultados con fines sumativos

La utilización de los resultados de la evaluación por estudiantes como forma de evaluación sumativa del profesor ha ido creciendo, incluso se ha sugerido que la utilización de las encuestas a estudiantes por los administradores con fines de empleo se va a incrementar tanto por presiones como por competición entre escuelas por estudiantes, aumento de la evaluación centrada en el estudiante como consumidor, competición para subir posiciones entre el profesorado y aumento en el intento de mejorar la instrucción (Linsky y Straus, 1973). Su implantación con fines sumativos ha chocado en muchas ocasiones con la oposición de los sectores implicados a lo largo de su implantación (Avi-Itzhak y Kremer, 1986). La utilización de las encuestas a estudiantes con propósitos administrativos es un asunto delicado y corre el riesgo de que el profesorado lo pueda percibir como un posible recorte de su libertad académica (Goldschmid, 1976) y pueda volverse hostil o no cooperativo. Como ya se ha señalado anteriormente, uno de los requisitos de cualquier sistema de evaluación para que sea realmente útil es la colaboración y la aceptación de todos los sectores implicados.

Gels (1977) establece que debería abandonarse la evaluación sumativa de la docencia basándose en que causa amenaza, ansiedad y divisiones entre el profesorado. Otros autores afirman que las evaluaciones de la docencia son inevitablemente mal usadas y mal interpretadas por los comités de promoción y empleo (Scheck, 1978), o que la utilización de encuestas institucionales a estudiantes conduce a los profesores a un aumento en el nivel de las calificaciones y a bajar los estándares académicos en un intento de "comprar" evaluaciones favorables a los estudiantes (Renner, 1981).

El carácter sumativo de la evaluación se orienta a consecuencias de tipo laboral, por lo que servirá de apoyo a decisiones acerca de la contratación, estabilidad y promoción, asignación de sueldo, complementos y control de profesorado. Además, puede ser un elemento de apoyo sobre factores de organización docente como la asignación de profesores a asignaturas, información a estudiantes, etc., o de desarrollo profesional (Magnusen, 1987).

Otro aspecto a considerar en la utilización sumativa de la evaluación del profesorado es la elección del curso/asignatura por parte del alumno. Ello enlaza directamente con la cuestión de publicación de los resultados que trataremos más adelante.

Respecto a la evaluación de la actividad investigadora, se centra fundamentalmente en la vertiente sumativa. Jorner et al. (1996) establecen que el primer uso que debemos hacer de la evaluación es la identificación de indicadores de investigación, de manera que se puedan detectar los objetivos en un programa que la institución debe cumplir, instaurados por la administración central y por el gobierno de la institución. Estos objetivos tienen dos funciones: en primer lugar, como guías de acción por la administración central y, en segundo lugar, como objetivos o prioridades para las divisiones académicas (departamentos, etc.) (Levin, 1991). Los autores señalan que son utilizados diferentes tipos de recompensas, en general, económicas y se encuentran diferentes enfoques, algunos aplicando parámetros económicos a la educación superior (Bradburd y Mann, 1993; Altschuid y Zheng, 1995).

Otro aspecto a tener en cuenta debería ser la asistencia a reuniones y comités universitarios, aunque presumiblemente no aumentarán. Otro uso de los resultados de la evaluación de la investigación es la asignación de recursos de investigación bien al profesorado, aprobando proyectos de investigación posteriores o asignando bonos para viajes, años sabáticos, estancias en el extranjero o dotando al profesorado de medios materiales como ordenadores, libros, etc., bien al Departamento que recibe los fondos de la investigación y que repartirá entre los investigadores en función de su producción (Jorner et al., 1996).

En la utilización de los resultados de la evaluación se incluyen aspectos relacionados tanto con la ética como con su impacto, que junto a cuestiones sociales e institucionales (tradicción, tipo de contrato, tipo de institución, normalmente interactuantes, condicionan en gran medida la utilidad de la evaluación (Jorner et al., 1996).

Salthouse et al. (1978) encuentran efectos muy pequeños de las encuestas de estudiantes sobre la promoción de los profesores.

Bonetti (1995) señala que uno de los roles importantes de las encuestas de evaluación es invisible: crear en el pensamiento del profesor la aprensión de que su ejecución está siendo evaluada y que sus compañeros y superiores serán informados de las consecuencias de los resultados de las encuestas. Se juega así con el miedo y respeto profesional de las desconocidas y, posiblemente, serias consecuencias en el caso de una utilización muy mala. Algunas consecuencias de esta utilización serán:

- Que el profesor se autoimponga unos estándares de ejecución mínimos "red de seguridad", por debajo de los cuales los profesores no dejarán caer las respuestas.
- La relación de las EE con los aspectos de salario vinculado a la ejecución. La preocupación expresada con mayor frecuencia es que un sistema de estas características puede recompensar conductas inapropiadas o no relevantes como belleza física, carisma, magnetismo personal, apariencia sexual, personalidad.

Indudablemente estos factores pueden influir algunas veces en las evaluaciones de los estudiantes, en particular si el cuestionario pregunta por un vago "evaluación de la ejecución del profesor", antes que atender puntualmente otros aspectos de técnicas docentes o calidad.

Bonetti señala también otros aspectos relacionados con la utilización de los cuestionarios, como la recompensa, entendida como un premio a aquellos profesores que han conseguido, con persistencia y trabajo, mejorar a nivel de personalidad, antes que a aquellos ya dotados con ésta. Este argumento tiene dos aspectos. Primero, los rasgos como el carisma o el humor pueden ser aprendidos o cultivados, por lo que el profesor puede exponer una clase de forma amena. Segundo, el argumento ignora el fin del sistema de recompensas. El fin de estos sistemas basados en recompensas no es simplemente lograr una buena ejecución, sino alcanzar los niveles más elevados de escolarización. El fin del docente no es enseñar aspectos sino introducir al estudiante en ese aspecto.

Marsh (1987a y 1987b), por otro lado, señala que los resultados de las opiniones de los estudiantes, como indicadores de la efectividad docente, proporcionarán una base para decisiones administrativas informadas y, por tanto, para aumentar la constancia de que la calidad docente será reconocida y recompensada.

Apodaka et al. (1990) indican que si las expectativas del profesorado ante la evaluación son de tipo fiscalizador se darán fuertes resistencias, produciéndose un rechazo a todo tipo de autorreflexión sobre los resultados de la evaluación, citando el Decreto sobre Retribuciones del Profesorado Universitario como posible fuente de desconfianza hacia el proceso evaluador.

Otro de los usos lo sugiere Cruse (1987) que establece que como las puntuaciones de estudiantes sobre los profesores son al fin, apuntes fiables de la popularidad, pueden presumiblemente, proporcionar la llave de la contabilidad para incrementar la matrícula en tiempos de "suministros" limitados y de competición entre instituciones de aprendizaje.

Los resultados indican (Feldman, 1979; Nimmer y Stone, 1991) que las encuestas a estudiantes deben utilizarse siempre con cautela si van a servir como base en la toma de decisiones administrativas o de tipo sumativo. Algunos autores incluso ponen en entredicho la utilización de las encuestas con fines sumativos, por diferentes razones, como su falta de validez, su facilidad de estar influidas por otras variables e incluso las dudas en algunos centros concretos sobre su legalidad, y se aboga por otras prácticas de evaluación del profesorado como la evaluación por iguales, aunque como veremos otros autores consideran que los resultados de la investigación proporcionan suficientes datos empíricos al efecto de que las encuestas a estudiantes son válidas y fiables (Aleamoni, 1974; Doyle, 1975; Gillmore, 1984; Murray, 1984). En general, se recomienda utilizar los datos de las encuestas a estudiantes conjuntamente con otras fuentes de información para compensar sus limitaciones (Cashin, 1983).

Finalmente, me parece interesante resaltar la puntualización establecida por Gillmore (1984): "para analizar la cuestión de si las encuestas a estudiantes están cualificadas como evidencia para fundamentar las decisiones sobre empleo y de profesorado, debemos referirnos a fiabilidad y validez".

Antecedentes

4. Análisis del CEDA de la UPV

4.1. Introducción

La evaluación del profesorado, como ya hemos comentado, constituye uno de los tópicos de máximo interés en nuestro país, en relación con la evaluación de la calidad de la docencia universitaria, respondiendo tanto a la necesidad de mejorar la calidad de nuestros centros como a las exigencias del marco legislativo universitario. Son muchas las universidades españolas, tanto públicas como privadas, que actualmente han implantado algún sistema de evaluación de su profesorado en el ámbito de la evaluación institucional, siendo el cuestionario de evaluación docente mediante opinión de los alumnos el más utilizado. Las características más relevantes de estos cuestionarios, fundamentalmente las relativas a la dimensionalidad de los mismos, varían de unas Universidades a otras, pudiéndose consultar los de algunas de ellas en la revisión efectuada por Muñoz Cantero et al. (2002) al respecto. En este contexto, en el presente capítulo se presentan las características del modelo de evaluación docente de la Universidad Politécnica de Valencia.

Antes incluso de que la Ley de Reforma Universitaria prescribiese el pase anual de encuestas de alumnado para evaluar la calidad docente del profesorado, éstas ya venían aplicándose en la Universidad Politécnica de Valencia desde el curso 1987-1988, aunque no de forma generalizada ni sistemática. Esta experiencia inicial, permitió detectar los aspectos susceptibles de mejora, que iban desde cuestiones más o menos formales, como pudieran ser la forma en que se procedía al pase de las encuestas, a cuestiones de mayor envergadura, como por ejemplo, la falta de explotación y análisis de los resultados por parte de los Centros y Departamentos, o el que no se hubiesen instrumentado institucional y sistemáticamente planes de mejora y perfeccionamiento que paliaran las deficiencias detectadas por la evaluación.

A partir de 1990, mediante la aprobación de la Resolución 15353 del Consejo de Universidades de 20 de junio de 1990 (BOE 30 de junio), se establecen los criterios generales para la evaluación de la actividad docente del profesorado universitario.

Comienza así, una nueva etapa en la UPV, siendo necesario institucionalizar el proceso de evaluación docente y la aprobación por parte de la Junta de Gobierno de los criterios que regirán el mismo.

A partir de 1995, la evaluación del profesorado de la UPV quedó enmarcada en el contexto más amplio de la Evaluación de la Institución Universitaria, en la que además de la labor docente, se definen otros indicadores que nos permitan obtener una visión más amplia y completa de misma. Cabe resaltar, que dentro del contexto de evaluación institucional, la calidad debe quedar entendida como un sistema de coherencia de los distintos elementos que la componen (necesidades y valores sociales, contexto, input, procesos y productos), siendo necesario definir los indicadores de tipo relacional entre tales componentes (De la Orden, 1987, 1992; Fernández Díaz, 1988 y García Ramos et al., 1995a y 1995b). Sin embargo, este proyecto integrado debe iniciar su puesta en marcha desde abajo, abordando dentro del marco general, cada elemento en particular, para recoger datos de los distintos elementos y poder establecer, posteriormente, las relaciones entre ellos.

En este trabajo, la evaluación se centra especialmente en los aspectos docentes. Es evidente, como ya se ha puesto de manifiesto, que los roles y funciones del profesorado universitario rebasan estas competencias, tal como queda patente por las funciones que legalmente se le asignan.

Dentro de las técnicas descritas de evaluación del profesorado utilizadas actualmente en la UPV, la opinión de los estudiantes continúa siendo una fuente principal de evaluación de la eficacia del profesorado, aunque nadie pone en duda la necesidad de completarlo con otras fuentes e incluso con otras técnicas distintas (Franke-Wikberg, 1990). En este sentido, no puede despreciarse el valor intrínseco que representa la evaluación del CEDA, ya que actualmente sigue representando el principal elemento de apoyo en la evaluación y mejora de la enseñanza, siendo un instrumento de utilización común en la mayoría de las instituciones de educación superior.

Por otro lado, junto a los cuestionarios genéricos que se han venido utilizando en numerosas universidades, se han desarrollado otros que fueran aplicables a universidades

específicas y diferenciadas, como es el caso de la UPV. Por este hecho, nuestro estudio pretende determinar las características técnicas y criterios de bondad del cuestionario utilizado en la UPV a partir de 1995, evaluar su adecuación técnica y desarrollar una metodología para analizar e interpretar la información contenida en el mismo.

4.2. Características del Cuestionario

Disponer de un cuestionario de evaluación adecuado a las características y peculiaridades de una determinada universidad es un proceso sucesivo de selección de ítems, que no puede darse por cerrado, ya que la Institución Universitaria es, como todas las organizaciones, dinámica.

Para establecer los elementos de la docencia que deben estar presentes en la evaluación del profesorado, es necesario partir de la elección de unos criterios referidos, por una parte, al ámbito de la clase y, por otra, a diversos aspectos formales relacionados con la enseñanza que son los que deben servir para el diseño de los instrumentos de valoración y como patrón de medida. En definitiva, se trata de partir de un modelo de competencias docentes que abarque los siguientes ámbitos (ICE, 1994):

- Aspectos formales relacionados con la enseñanza: planificación de la asignatura, programación, secuenciación e interdisciplinaridad; utilización de recursos; cumplimiento de los objetivos didácticos; coherencia entre los objetivos establecidos, la evaluación y el cumplimiento de las obligaciones, como por ejemplo, asistencia a las clases, puntualidad, atención a los alumnos, etc.
- Aspectos referidos al “acto didáctico”: claridad comunicativa en la presentación de los objetivos y contenidos de la asignatura, en las exposiciones, preguntas, etc.; estímulo y conducción de la participación del alumnado en las clases; atención y asistencia de los alumnos y creación de un clima de trabajo propicio al aprendizaje.

El proceso tiene sus inicios en el primer cuestionario seleccionado por los miembros del Seminario, integrado por profesores y directivos de los distintos departamentos, representantes del alumnado y técnicos del ICE. Las fuentes utilizadas en su elaboración fueron las siguientes:

- El cuestionario con el que el ICE venía trabajando desde el año 1980, elaborado a partir de los estudios sobre competencias docentes y que había sido objeto de una investigación a nivel nacional.
- Las encuestas de las Universidades Autónoma de Madrid y Santiago de Compostela, basadas en los trabajos de expertos en el tema, entre los que cabe destacar a Marsh, Hoyt y Cashin entre otros.
- El cuestionario de L'Institut National de Recherche Scientifique de Québec, cuyo autor es el profesor Gagné.

El cuestionario inicial se revisó después de cada una de las aplicaciones realizadas, modificándose a raíz de los análisis llevados a cabo. Así por ejemplo, el número de ítems, pasó de 34 en la primera aplicación del curso 1987-1988 a 26 en el curso 1992-1993.

Las decisiones más importantes que se han ido tomando se refieren a:

- Mantener los ítems que saturan de forma importante y repetida en los factores o dimensiones que se mostraban más estables.
- Eliminar los ítems que quedaban descolgados de la mayoría de las estructuras factoriales estudiadas.
- Mantener la redacción del ítem 14 (el profesor se interesa por los problemas educativos de los alumnos), puesto que en la aplicación del curso 89-90 y, a petición de algunos Jefes de Estudio, se alteró su redacción y los análisis psicométricos mostraron que los alumnos confundían el propósito de esta pregunta, ligándola a aspectos relacionados con el dominio de la asignatura y no con los que se pretendía conocer, que es la interacción del profesor con los alumnos.

- Eliminar los ítems relacionados con la dimensión “asignatura”, al entender que podrían crear confusión al estar integrados en un cuestionario de evaluación de profesores.
- Reordenar algunos de los ítems del cuestionario a fin de favorecer la coherencia interpretativa.
- Mantener el ítem criterio (Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura puede considerarse un buen profesor).

El cuestionario estaba constituido por tres bloques: uno donde se recoge la información de la asignatura y profesor evaluado, otro donde se recoge información relativa al alumno y, un último donde se recogen los 26 ítems propiamente dichos, de respuesta tipo Likert con 5 alternativas de valoración, como se muestran en la tabla 4.1.

Tabla 4.1. Escala de Valoración

VALORACIÓN	SIGNIFICADO
1	Totalmente en desacuerdo
2	Más bien en desacuerdo
3	Indiferente
4	Más bien de acuerdo
5	Totalmente de acuerdo

Las dimensiones del cuestionario utilizado en el curso 1992-1993 eran las que se indican a continuación y comprendían los ítems que se señalan en cada caso:

Tabla 4.2. Dimensiones del cuestionario utilizado en el curso 1992/1993

DIMENSIÓN	ITEMS
Dominio de la asignatura y claridad en la exposición/desarrollo del programa	Del 1 al 11
Interacción profesor con los alumnos	Del 12 al 15
Exámenes	Del 16 al 18
Cumplimiento de las obligaciones formales	Del 19 al 21
Recursos utilizados y prácticas	Del 22 al 24
Valoración global	25 y 26

En julio de 1994, el ICE presentó a la Junta de Gobierno el documento “Evaluación del Profesorado”, aprobándose por unanimidad la encuesta institucional que viene aplicándose desde el curso 1994-1995. En esta propuesta, el número de ítems se había reducido a 19, siendo el último un ítem de valoración global o también denominado ítem criterio. Todos los ítems eran de respuesta tipo Likert, con las 5 alternativas de valoración utilizadas en las aplicaciones anteriores, más dos adicionales: 0, si la respuesta era en blanco y 6 si el alumno no tenía criterios para contestar. La estructura de este cuestionario puede consultarse en el anexo 1.

4.3. Proceso de Evaluación

Para realizar el proceso de evaluación deben tenerse en consideración dos aspectos esenciales: la transparencia o claridad de la información y el sentido común para plantear el proceso en los términos debidos. La experiencia acumulada de los técnicos del ICE ha conseguido que este proceso se haya ido depurando cada vez más, siguiéndose en la actualidad las pautas que se detallan a continuación:

- En primer lugar se efectúa una campaña previa de difusión a Directivos de Centros, profesores y alumnos, informándoles del proceso y solicitando su participación.
- A continuación se procede a actualizar las bases de datos y a preparar el pase de encuestas. La actualización de la base de datos es efectuada por el ICE en colaboración con los distintos Centros. Para poder confeccionar la relación de profesores / asignaturas, cada año el ICE envía a cada centro el listado del curso anterior, para que éste efectúe las revisiones y rectificaciones oportunas. Estos listados son remitidos de nuevo al ICE y con ellos, se elaboran las etiquetas, con sus códigos correspondientes, que son enviadas a los Centros junto con los sobres elaborados específicamente para recoger los cuestionarios.
- Una vez terminado el proceso anterior, es necesario decidir las personas que se harán cargo de realizar la aplicación. Actualmente, el Vicerrectorado de

Coordinación Académica y Alumnado convoca unas becas para seleccionar a determinados alumnos que se encargarán de efectuar el pase de encuestas al profesorado que se le asigne, bajo la supervisión del Responsable de Centro y en coordinación con la Unidad de Evaluación del Instituto de las Ciencias de la Educación.

- El procedimiento de aplicación debe caracterizarse por la individualización, en el sentido de aplicar los cuestionarios a cada profesor en su misma clase, para garantizar en mayor medida, la representatividad de las respuestas recogidas. Por otro lado, el momento de efectuar el pase de encuestas, es establecido por el Centro, tras haberlo notificado previamente a los distintos profesores, para asegurar la seriedad del proceso y evitar que se produzcan situaciones anómalas.
- Finalmente, es necesario proceder a la codificación y grabación de datos para su posterior análisis. Esta etapa es especialmente delicada, ya que puede ser fuente de importantes errores. El proceso comienza en el momento del pase de encuestas, ya que se facilitan los sobres con las etiquetas donde deben introducirse los cuestionarios de cada profesor y cerrarse, figurando la firma del responsable del pase de las encuestas y la del propio profesor. Cualquier anomalía en este sentido, es objeto de estudio por parte del equipo técnico del ICE, advirtiendo de los posibles sesgos que se pueden producir. Una vez que los sobres llegan al ICE, comienza el proceso de grabación de los datos. Para este fin se ha diseñado un sistema de información que se describe en el apartado siguiente y que abarca todo el proceso, desde la entrada de datos hasta la emisión de los distintos informes, garantizando con escasas posibilidades de error, la lectura fiable de la información contenida en los cuestionarios.

Con los datos procesados se efectuarán los análisis necesarios para cubrir los dos frentes propuestos: la elaboración de informes y la investigación sobre el propio proceso de evaluación seguido.

La elaboración de informes, tanto colectivos como individuales, tiene como objetivo ofrecer datos básicos con claridad a los profesores, departamentos y centros, que

son los que reciben los informes de resultados correspondientes. Actualmente, con objeto de contribuir a una mejor y más rápida difusión de los resultados de las encuestas, éstos son colocados anualmente en la página web personal de cada profesor. Los datos recogidos en estos informes son: frecuencia absoluta y relativa de cada una de las alternativas de respuesta de los 19 ítems incluidos actualmente en el cuestionario; la media ponderada y desviación típica de cada ítem; histogramas de frecuencia; número de alumnos que han contestado el cuestionario en cada clase; media ponderada de los resultados de las distintas dimensiones o factores que mide el cuestionario y medias de distintos referentes como pueden ser el curso, departamento, unidad departamental, centro, etc.

En la figura 4.1 se muestra un ejemplo de estos resultados. Comentar, que en la gran mayoría de los casos, las encuestas se pasan antes de que se hayan realizado los exámenes de la asignatura, no apareciendo valorados los ítems referentes a evaluación.

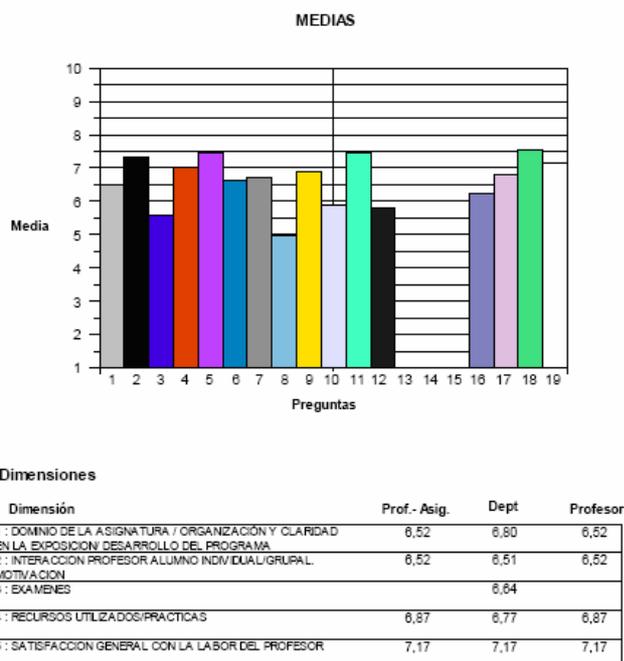


Figura 4.1. Valoraciones medias obtenidas por un profesor

Los informes básicos que el sistema permite elaborar son:

- Individuales: profesor, profesor-asignatura y profesor-asignatura-grupo.
- Colectivos: Centro, Departamento, Unidad Docente, Curso, Asignatura y Universidad.

Para investigar el propio proceso de evaluación efectuado, con la finalidad de ir consiguiendo progresivamente mayores cotas de fiabilidad y validez en nuestro contexto, se efectúa un estudio psicométrico de los ítems incluidos en el cuestionario, que incluye los siguientes análisis: correlación entre ítems; correlación ítems-cuestionario; correlación entre los ítems del cuestionario y el ítem criterio; correlaciones múltiples de cada uno de los ítems con todos los demás; estimación de índices de fiabilidad y validez de cada ítem; contribución de cada ítem a la varianza, fiabilidad y validez del cuestionario; estimación de la fiabilidad y validez global del cuestionario y, finalmente análisis de la estructura factorial del cuestionario.

Actualmente, en este contexto, se está intentando potenciar todos aquellos estudios que sirvan para conocer mejor cada contexto concreto de las diferentes realidades que se pueden producir en la universidad, considerando la influencia que pudiesen tener algunas variables, la posibilidad de elaborar índices correctores de los resultados individuales del informe en cada situación, etc. Fruto de dichas inquietudes, se plantea la necesidad de comenzar el presente trabajo, con la finalidad de conseguir que la evaluación del profesorado se oriente fundamentalmente a la estimación del nivel de calidad de la enseñanza universitaria, a fin de contribuir progresivamente a su mejora.

4.4. Sistema de tratamiento de la información

Para analizar e interpretar la información contenida en las encuestas, se ha diseñado un programa informático denominado *Evaprof (Evaluación del Profesorado)*, que gestiona directamente todos los procesos básicos: control de la lectura de encuestas, almacenamiento de la información, cálculo estadístico primario y salidas de informes a todos los niveles. Esta herramienta ha sido diseñada en varias etapas sucesivas por alumnos

de la Facultad de informática y está considerada actualmente como una de las propuestas más interesantes para el efectivo procesamiento de los datos de evaluación.

El procesamiento informático de las encuestas se basa en la técnica de Reconocimiento de Marcas Ópticas. La mecánica necesaria para la introducción de las encuestas está marcada por las características del sistema físico o lectora de marcas (DATASCAN serie 7) y por el método establecido para la recogida de las encuestas.

Una vez introducidos los datos, el sistema efectúa un proceso automático de depuración de los mismos, que consiste en la eliminación de errores, aunque en realidad éstos están bastante acotados. El principal error que se cometía en los primeros años de efectuar el pase de encuestas era la introducción incorrecta del código del profesor o de la asignatura por parte del alumno, lo que se ha solucionado agrupando las encuestas de un mismo profesor/asignatura en un mismo sobre e introduciendo los códigos únicamente en el mismo por el responsable de efectuar el pase de encuestas.

El resto de posibles errores son propios de la codificación de las preguntas.

Se ha tomado como convenio descartar todos los cuestionarios que se hayan dejado en blanco por completo.

Finalmente, otro tipo de errores que pueden presentarse es que los alumnos hayan marcado dos o más respuestas para una misma pregunta. Dichas encuestas son rechazadas hacia la bandeja de errores de la lectora y automáticamente en la pantalla del ordenador se indica el número de errores detectados en dicho cuestionario y las preguntas a las que corresponden. Dichas preguntas son consideradas nulas y para ello se borran las distintas respuestas de esos ítems.

La interpretación de las encuestas es uno de los principales puntos de acción de lo que constituye la aplicación de *Evaprof*. Para ello, básicamente se aplican una serie de conceptos estadísticos para la generación de los informes básicos asociados a lo que es el análisis primario.

Inicialmente, cada ítem es considerado como una variable de tipo ordinal, de acuerdo con la escala de medida utilizada. A este tipo de variable se le aplica una

transformación, convirtiéndola en una variable de tipo continuo. Los valores empleados, así como las abreviaturas que emplearemos a partir de ahora, se muestran en la tabla 4.3.

Tomando como modelo los informes básicos que se dan a cada profesor, las medidas utilizadas son: las frecuencias absoluta y relativa, así como la representación gráfica de las mismas mediante un histograma de frecuencias; la media y la desviación típica.

Tabla 4.3. Transformación de la escala de valoración

SIGNIFICADO	ABREVIATURA	VARIABLE ORDINAL	VARIABLE DISCRETA
Totalmente en desacuerdo	TED	1	0
Más bien en desacuerdo	MBD	2	2,5
Indiferente	IND	3	5
Más bien de acuerdo	MBA	4	7,5
Totalmente de acuerdo	TDA	5	10
No sabe. No contesta	NSNC	0/6	-

4.5. Descripción del estudio

4.5.1. Base de Datos

El estudio se centra en el análisis del funcionamiento del cuestionario en un periodo que va desde el curso 1995/1996 hasta el curso 2000/2001, ya que es a partir de 1995 cuando comienza a utilizarse el cuestionario objeto de este estudio.

Los datos sobre los que se desarrollarán los distintos estudios han sido extraídos de la Base de Datos del Centro de Cálculo de la Universidad Politécnica de Valencia.

Se nos proporcionaron dos bancos de datos de carácter longitudinal. El primero de ellos estaba constituido por las puntuaciones en la escala tipo Likert con cinco alternativas de respuesta, obtenidas en los 19 ítems incluidos en el cuestionario en cada una de las encuestas realizadas en cada uno de los periodos evaluativos. En estos archivos se incluía

también, el código de la asignatura evaluada, el código del profesor evaluado y una columna auxiliar con información referente al alumno que había respondido la encuesta.

El número de encuestas obtenido en cada periodo evaluativo se muestra en la tabla 4.4.

Tabla 4.4. Relación de encuestas procesadas en cada curso

CURSO	Nº DE ENCUESTAS
1995/1996	117.702
1996/1997	130.756
1997/1998	135.582
1998/1999	140.149
1999/2000	136.267
2000/2001	134.314

El segundo banco de datos, estaba constituido por las puntuaciones medias obtenidas para cada uno de los 19 ítems, en cada posible combinación profesor-asignatura. Estas puntuaciones medias, pueden considerarse ya variables continuas, presentando valores decimales entre 0 y 10 y se obtuvieron multiplicando el valor de la variable creada anteriormente, por el número de encuesta en los que se había señalado dicho valor y promediando en función del número de encuestas procesadas en total para cada combinación profesor-asignatura. En este caso, se incluían las variables disponibles para cada profesor en el archivo de personal y de la asignatura. Hay que tener en cuenta, que al ser archivos dinámicos, sólo se han podido considerar variables bien identificadas, aunque en cada año, independientemente considerado, se pudieran considerar otras distintas. Las variables descriptivas incluidas fueron: categoría del profesor, si era doctor o no, edad, departamento, tiempo de dedicación, titulación, curso y semestre en el que se imparte la asignatura. Es habitual en este tipo de estudios que la información varíe de unas aplicaciones a otras, debido a que en cada aplicación las asignaturas y los profesores que las imparten pueden no ser los mismos. El número de posibles combinaciones profesor-asignatura obtenido en cada periodo evaluativo se muestra en la tabla 4.5.

Tabla 4.5. Posibles combinaciones relación profesor- asignatura

CURSO	COMBINACIONES PROF-ASIG
1995/1996	3.054
1996/1997	3.878
1997/1998	4.384
1998/1999	4.870
1999/2000	5.048
2000/2001	5.760

4.5.2. Análisis de los ítems

Para abordar el presente estudio y posteriormente evaluar la estabilidad, consistencia, discriminación y validez del instrumento de evaluación docente utilizado actualmente en la UPV, se tomaron los datos de los 134.309 cuestionarios, correspondientes a las valoraciones individuales que los alumnos de la Universidad Politécnica de Valencia realizaron de sus profesores en el curso académico 2000/2001, por ser ésta la aplicación más reciente de la que se disponía información.

La aplicación de cualquier técnica multivariante, requiere la realización de un análisis descriptivo de las variables, de manera que se puedan verificar las hipótesis estadísticas en las que se basan la gran mayoría de análisis multivariantes, detectar datos anómalos y, en general, conocer las características de cada uno de los ítems.

En la tabla 4.6, se recogen el número de respuesta obtenidas en cada una de las 19 variables analizadas.

A la vista de los resultados obtenidos y, debido a que las preguntas respondidas como 0 (*Respuesta en blanco*) o 6 (*No tengo criterios para contestar*) podrían afectar a la posterior transformación de la variable, se decidió efectuar un filtrado de los datos, de manera que se excluyesen todas las encuestas en las que alguno de los ítems se había contestado con alguno de estos dos valores. De esta forma se obtuvo una muestra de 97847 cuestionarios, en la que todos los alumnos habían valorado los 19 ítems de 1 a 5.

Tabla 4.6. Estadísticos Descriptivos

	0	1	2	3	4	5	6	Sin responder
R1	341	6184	13914	29475	44767	38341	1289	3
R2	239	4199	9488	24977	46312	47318	1778	3
R3	632	7139	15073	32096	44584	33234	1553	3
R4	564	4570	12025	12025	31773	46873	36435	3
R5	468	5158	10390	29008	47364	38221	3702	3
R6	518	6001	12949	32164	42231	26204	14244	3
R7	423	5302	12805	30193	42470	40447	2671	3
R8	568	8279	22122	45618	35400	18435	3889	3
R9	397	7125	14931	35979	42117	31722	2039	4
R10	430	10247	19466	36041	35747	28856	3523	4
R11	376	6709	11131	29171	42737	41853	2333	4
R12	836	13310	18840	39952	35612	23439	2321	4
R13	102783	2195	3053	6505	9314	8769	2436	4
R14	102783	3009	3739	6839	7503	6287	4151	4
R15	102903	1707	1981	5224	7865	9799	4831	4
R16	2868	6744	14485	36402	41821	26058	5932	4
R17	2867	4437	11423	34714	46605	32232	2031	5
R18	2900	6232	13297	33171	44626	30415	3668	5
R19	2739	5814	10863	28478	43330	41067	2018	5

A su vez, analizando los resultados obtenidos, se puede detectar una elevada presencia de preguntas respondidas en blanco en los ítems codificados como R13, R14 y R15. Analizando con mayor profundidad estos datos, observamos que estos ítems, corresponden a preguntas referentes a los sistemas de evaluación, y la gran mayoría de los alumnos encuestados las ha valorado mediante un cero (*Respuesta en blanco*), lo que indica que en el momento que se realizó la encuesta, todavía no habían sido evaluados.

En consecuencia, es necesario dividir los datos iniciales en dos grupos:

- GRUPO A: Cuestionarios de opinión de aquellos alumnos con juicio para opinar sobre los ítems referentes a evaluación.

- GRUPO B: Cuestionarios de opinión de aquellos alumnos que han valorado los ítems referentes a evaluación como cero. En este caso, eliminaremos del estudio las variables que representan a los ítems codificados como R13, R14 y R15.

En las tablas 4.7 y 4.8 se muestran los principales estadísticos descriptivos de las variables analizadas en cada grupo tras haber eliminado las respuestas con 0 y/o 6 en todos los ítems excepto en los referentes a evaluación. Mediante este análisis podemos detectar que en el grupo de alumnos que sí habían sido evaluados antes de la encuesta, la media de los ítems codificados como R13, R14 y R15 aumenta y la desviación típica disminuye.

Tabla 4.7. Estadísticos Descriptivos Grupo A

	Media	Desviación típica	N del análisis
R1	3,67	1,20	21672
R2	3,87	1,13	21672
R3	3,60	1,19	21672
R4	3,75	1,14	21672
R5	3,73	1,15	21672
R6	3,60	1,16	21672
R7	3,69	1,16	21672
R8	3,25	1,15	21672
R9	3,56	1,18	21672
R10	3,42	1,27	21672
R11	3,70	1,22	21672
R12	3,29	1,27	21672
R13	3,62	1,22	21672
R14	3,35	1,28	21672
R15	3,80	1,18	21672
R16	3,55	1,14	21672
R17	3,70	1,10	21672
R18	3,64	1,14	21672
R19	3,74	1,21	21672

Tabla 4.8. Estadísticos Descriptivos Grupo B

	Media	Desviación típica	N del análisis
R1	3,73	1,11	76175
R2	3,93	1,04	76175
R3	3,62	1,13	76175
R4	3,75	1,05	76175
R5	3,79	1,06	76175
R6	3,56	1,07	76175
R7	3,77	1,10	76175
R8	3,28	1,08	76175
R9	3,59	1,12	76175
R10	3,43	1,19	76175
R11	3,78	1,12	76175
R12	3,30	1,19	76175
R16	3,52	1,09	76175
R17	3,69	1,04	76175
R18	3,61	1,10	76175
R19	3,80	1,10	76175

4.6. Análisis de Fiabilidad y Validez del Cuestionario

El primer criterio sobre el que se analiza la bondad de un instrumento es su fiabilidad, que alude a características intrínsecas del mismo. La fiabilidad o precisión, se puede definir como la ausencia de error aleatorio y representa la influencia del azar en nuestra medida, es decir, es el grado en el que las mediciones están libres de la desviación producida por los errores causales. Además, la precisión de una medida es lo que asegura su repetibilidad (si la repetimos, siempre da el mismo resultado). Existen diversos factores que afectan a la fiabilidad de los cuestionarios. Entre todos ellos cabe destaca, por un lado, el número de preguntas de los cuestionarios (debemos asegurarnos que contamos con un cuestionario que tiene múltiples preguntas que traten cada una de las categorías que estamos midiendo) y, por otro lado, la muestra de usuarios sobre la que se calcula la estimación de la fiabilidad.

En la literatura especializada, encontramos que no existe una metodología clara para estimar la fiabilidad. La posición más aceptada es la que considera el elemento clave para valorar la fiabilidad de este tipo de cuestionarios como correlación *inter-clase*, es decir como correlación entre jueces en la evaluación de un profesor, frente a la fiabilidad *inter-ítems*. El argumento más usual que apoya esta posición es que la consistencia *inter-ítems* es un estimador que sobrevalora la fiabilidad, siendo esta inflación explicable por el hecho de que no tiene en cuenta la parte de error debida a la falta de acuerdo entre individuos (Gilmore et al., 1978; Marsh, 1984 y Marsh y Roche, 1993).

En el presente trabajo, siguiendo lo establecido por Jornet et al. (1995), consideramos que la fiabilidad, desde un punto de vista métrico, alude a la consistencia en la evaluación de la docencia, entendiendo ésta como un constructo teórico que se especifica en el perfil definido por el cuestionario utilizado, por lo que la consistencia *inter-ítems*, estimada para cada evaluación individual del profesor, es un indicador de las características métricas del instrumento, mientras que la correlación *inter-clase* podría considerarse como un indicador para la aplicación específica a cada evaluación individual.

Se utilizará como medida de la fiabilidad del CEDA la medida de consistencia interna α -Cronbach. Este índice fue desarrollado para calcular el grado de consistencia

interna de los ítems, oscilando sus posibles valores entre 0 y 1, siendo mayor la consistencia interna cuanto más próximo esté a este último valor. La siguiente fórmula es aplicable tanto a ítems dicotómicos (sólo admiten dos respuestas: verdadero/falso, sí/no, acuerdo/desacuerdo,..), como a ítems de elección múltiple (con más de una alternativa de respuesta). La expresión formal es:

$$\hat{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum \hat{\sigma}_i^2}{\hat{\sigma}_X^2} \right)$$

donde $\hat{\alpha}$ es el estimador del coeficiente de fiabilidad, k el número de ítems del test, $\hat{\sigma}_i^2$ es la varianza de las respuestas de los sujetos al ítem i , y $\hat{\sigma}_X^2$ la varianza total de las puntuaciones observadas del test.

Los estudios relativos a la fiabilidad de cuestionarios existentes hasta el momento presentan valores del alfa de Cronbach que oscilan entre 0.80 y 0.90 para escalas que contengan ítems generales, mientras que cuando éstas están construidas con ítems específicos sus niveles se sitúan entre 0,5 y 0,8 (Feldman, 1977 y 1978; Marsh y Overall, 1981; Rippley, 1981; Braskamp et al., 1984, entre otros). Asimismo, estos niveles son mayores en los estudios de fiabilidad como consistencia inter-ítems, que en los de consistencia inter-jueces (Marsh y Roche, 1993).

Para realizar el estudio de fiabilidad del cuestionario se utilizaron los datos de los 97847 cuestionarios del curso 2000/2001, ya empleados en el análisis anterior.

Se obtuvo un índice de consistencia interna, α -Cronbach de 0,9561 en el caso de los cuestionarios en los que no se incluían los ítems de evaluación, y de 0,9649 en los cuestionarios en los que habían sido contestados todos los ítems, ambos valores superiores a los obtenidos en los estudios existentes hasta el momento.

El segundo criterio para evaluar la bondad del instrumento de medida es su validez, es decir, hasta que punto la prueba mide el aspecto o rasgo que se pretende medir.

La práctica habitual en los instrumentos de medida supone tomar como punto de partida la suma de las puntuaciones de un conjunto de ítems. Suele hablarse, de que un

determinado individuo obtiene una determinada puntuación en una cierta escala de un instrumento, pero en realidad, esa práctica sólo estaría justificada cuando pudiese comprobarse que los ítems que se suman son parte de un mismo y único rasgo, es decir, cuando se verifique la validez del instrumento.

Tomemos como referencia básica el concepto de validez sostenido en los *Standards for Educational and Psychological Testing* establecidos por las *Normas de la American Psychological Association* (APA), recogidas en una revisión realizada por Tourón en 1989. Dichos estándares definen la validez como un concepto unitario, apoyado en diferentes evidencias (contenido, constructo y criterio), que se refiere al grado en que tales evidencias soportan las inferencias hechas desde las puntuaciones de los instrumentos de medida. Así mismo, en ellos se establece que una validación ideal debe incluir los tres tipos de evidencias tradicionales (contenido, constructo y criterio) y que deberán ser los juicios profesionales los que guiarán las decisiones respecto a las evidencias más importantes a la luz del uso pretendido del test.

Lo que se valida según esa definición no es un cuestionario en sí, ni una escala o un test, sino las inferencias que se hacen para responder a determinados propósitos de investigación (García Jiménez et al., 2000). Los diferentes tipos de inferencia a partir de las puntuaciones requieren de diferentes tipos de evidencias, no de diferentes tipos de validez. El proceso de validación se convierte por tanto en el proceso de recogida de evidencias y pruebas que apoyen los usos que se realizan a partir de las puntuaciones de un test. Las categorías tradicionales de criterio, contenido y constructo no implican separaciones en sentido estricto ni se corresponden con tipos diferentes de inferencias o utilidades del cuestionario. De igual forma, como señala Cronbach (1984), el objetivo final de la validación es la explicación y la comprensión, lo que lleva a considerar que toda validación es validación de constructo o concepto, es decir, se trata de desentrañar el rasgo o construcción que subyace a la varianza de las puntuaciones de un instrumento y para su estudio se requieren pruebas o evidencias experimentales diversas que conlleven a poder admitir grados y no un único índice (Tourón, 1989).

Una de esas evidencias, a partir de las que podría afirmarse que son válidas ciertas inferencias acerca de un concepto científico, al que se le denomina constructo, o sus

relaciones, es la que se apoya en que las conductas observables en un test, las opiniones recogidas en un cuestionario o las valoraciones reflejadas en una escala son indicadores de dicho constructo. La validez de constructo facilita la aproximación a ciertos conceptos hipotéticos que no pueden observarse directamente, utilizando para ello la relación que dichos conceptos guardan con ciertos indicadores con los que están estrechamente relacionados.

Aunque existen multitud de estudios respecto a como evaluar la validez de los instrumentos de medida, la gran mayoría son contradictorios y complejos, no habiendo podido establecerse un procedimiento adecuado para validar estas herramientas.

En el caso de las encuestas de evaluación el concepto de validez hace referencia a si el instrumento es representativo de los aspectos que evidencian el perfil docente que se pretende evaluar (Marsh, 1987a; 1987b). De esta forma, los elementos del instrumento utilizado deben representar el dominio de lo que se define como buena docencia. Así, los factores que contribuyen a conseguir que los instrumentos puedan ser válidos son (Jornet et al., 1993):

- La interpretación de la orientación política de la evaluación a través de su puesta en marcha, a nivel de dimensiones e ítems, del tipo de perfil que se pretende evaluar.
- La participación de los colectivos implicados en la evaluación (institución, agentes que aportan la información y sujetos evaluados).
- La revisión técnica e interpretación del perfil a evaluar desde la investigación disponible acerca de la eficacia docente.
- La revisión del grado de generabilidad/especificidad de los ítems, de forma que se ajuste al propósito general de escala.

En nuestro caso, para evaluar la validez del cuestionario se han realizado tres tipos de análisis tanto en los datos del grupo A como en los del grupo B, ya que no existe un criterio universalmente aceptado de lo que es un profesor efectivo: en primer lugar, se

analiza la validez del cuestionario global considerando todos los ítems menos uno, en segundo lugar, se evalúa la validez de constructo o concepto mediante los valores del α -Cronbach obtenidos para cada dimensión extraída en el análisis factorial exploratorio realizado posteriormente en el capítulo siguiente y, finalmente, se evalúa de nuevo la validez de constructo mediante los valores de consistencia interna obtenidos al considerar el ítem con mayor carga factorial en cada dimensión. Para estos dos últimos estudios de validez será necesario trabajar con los factores obtenidos en el capítulo cinco mediante el análisis de dimensionalidad del cuestionario y con los ítems de mayor carga factorial en cada uno de ellos.

En las tablas siguientes se muestran los índices de validez obtenidos en cada uno de los tres análisis mencionados. Valores de consistencia interna incluyendo todos los ítems del cuestionario menos uno.

Tabla 4.9. Consistencia Interna Grupo A

Item no incluido	α - Cronbach
R19	0,9615
R18	0,9627
R17	0,9631
R16	0,9633
R15	0,9646
R14	0,9642
R13	0,9641
R12	0,9621
R11	0,9626
R10	0,9630
R9	0,9628
R8	0,9637
R7	0,9634
R6	0,9629
R5	0,9627
R4	0,9625
R3	0,9631
R2	0,9629
R1	0,9626

Tabla 4.10. Consistencia Interna Grupo B

Item no incluido	α - Cronbach
R19	0,9510
R18	0,9538
R17	0,9539
R16	0,9545
R12	0,9520
R11	0,9531
R10	0,9536
R9	0,9532
R8	0,9547
R7	0,9543
R6	0,9539
R5	0,9530
R4	0,9529
R3	0,9537
R2	0,9533
R1	0,9525

- Índice de validez de cada una de las dimensiones extraídas mediante el análisis factorial exploratorio tras conocer la base de construcción del cuestionario.

Tabla 4.11. Consistencia Interna Grupo A

Dimensión	α - Cronbach
1	0,9397
2	0,8370
3	0,9139
4	0,8587
5	0,8044

Tabla 4.12. Consistencia Interna Grupo B

Dimensión	α - Cronbach
1	0,9273
2	0,8971
3	0,8268
4	0,7613

- Índices de validez considerando el ítem de mayor carga factorial en cada dimensión.

En los datos del Grupo A, se obtuvo un valor del α - Cronbach de 0,8514 entre los ítems R1, R13, R10, R17 y R8. En los datos del grupo B, se obtuvo un valor del α -Cronbach 0,8095 entre los ítems R1, R10, R16, R8.

A la vista de los resultados, que en casi todos los casos arrojan valores del α -Cronbach superiores a 0,9, podemos concluir que el cuestionario presenta unas evidencias claras en cuanto a su validez de constructo y a su fiabilidad, entendida ésta como consistencia interna. En general, podemos afirmar que el cuestionario globaliza la actuación del profesor, discrimina bien y presenta unos niveles de consistencia interna similares a los cuestionarios estándares de evaluación de la docencia.

Finalmente, comentar que el análisis factorial, en primer lugar exploratorio y posteriormente confirmatorio, que se desarrollan en el capítulo siguiente, se nos revelan también como herramientas que pueden utilizarse para la validación del constructo desde el punto de vista de los indicadores bajo los que se manifiesta el mismo, ya que es una técnica estadística que examina la estructura interna de la unidad de medida, mide si los

Análisis del CEDA

indicadores tienen algo en común, es decir si tienen un común denominador, mide las correlaciones entre los indicadores e intenta descubrir si hay algo subyacente.

5. Dimensionalidad del Cuestionario

5.1. Introducción

Tras efectuar el análisis previo de fiabilidad y validez del cuestionario, el siguiente paso en el estudio del instrumento de medida, es intentar conocer la dimensionalidad del mismo, con la finalidad de analizar e interpretar la información contenida en el mismo.

El Análisis Factorial (AF) constituye una técnica estadística multivariante cuyo principal propósito es definir la estructura subyacente en una matriz de datos. Para ello, se analiza en primer lugar la estructura de interrelaciones entre un gran número de variables y, posteriormente se reducen los datos originales mediante la definición de una serie de dimensiones comunes subyacentes, conocidas como factores. Estas nuevas dimensiones permiten describir los datos originales con un número de conceptos mucho más reducido que las variables individuales originales.

A diferencia de otras técnicas estadísticas multivariantes, el AF es una técnica de interdependencia en la que se consideran todas las variables simultáneamente, cada una relacionada con las demás.

En realidad, el término de AF es un término genérico que agrupa dos tipos de modelos: unos conocidos como Análisis de Componentes Principales (ACP), en los que las variables latentes están determinadas únicamente por las funciones lineales de las variables observadas y, otros conocidos como Análisis de Factor Común (AFC), que van más allá de la explicación de las variables latentes basadas en las funciones lineales de las variables observadas, buscando determinar cuanto covarían entre sí las variables observadas en términos de las variables latentes del factor común, o bien, buscando interpretar las variables observadas como la suma de dos variables latentes, una puntuación verdadera y una puntuación de error (García Jiménez et al., 2000).

Además, con el mismo término (AF) se designan dos metodologías que responden a dos objetivos distintos que las califican como exploratorias y confirmatorias

respectivamente. La primera de ellas se refiere a aquellos procedimientos cuyo objetivo es extraer las variables latentes que subyacen a un conjunto de variables observadas y, la segunda de ellas, se refiere a un método que persigue verificar determinadas hipótesis sobre la estructura de un grupo de variables manifiestas.

En este capítulo, se efectuará en primer lugar un análisis factorial exploratorio para conocer la estructura subyacente del cuestionario y, como ya comentamos al final del capítulo 4, como procedimiento básico para analizar la validez del instrumento y de la fiabilidad de la medida. En segundo lugar, se aplicará una metodología confirmatoria para verificar sobre la base de construcción del cuestionario y sobre los estudios exploratorios realizados, que el instrumento capta las dimensiones básicas, que se consideran “claves” para lograr una enseñanza de calidad.

5.2. Metodología

5.2.1. Análisis Factorial Exploratorio

La técnica multivariante a utilizar para identificar las dimensiones subyacentes existentes en el cuestionario es el Análisis Factorial con fines exploratorios. La metodología del análisis factorial se esquematiza en la figura 5.1.

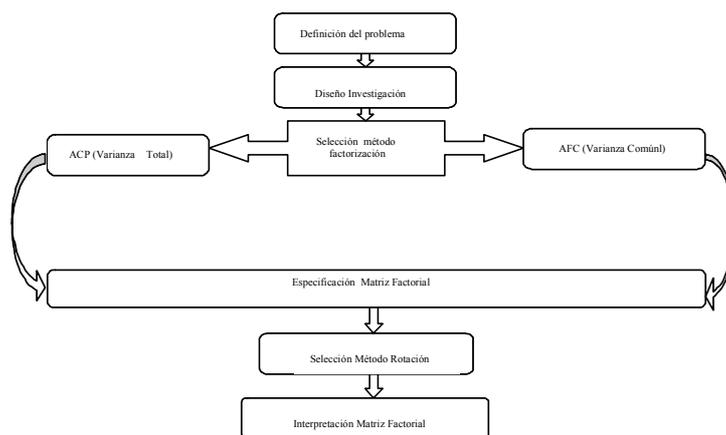


Figura 5.1. Metodología del AF

Esta herramienta, fue utilizada inicialmente por Charles Spearman en 1904, en su estudio efectuado sobre la teorización psicológica de las habilidades intelectuales y que ha sido denominado teoría de los dos factores o teoría del factor común, aplicándose desde ese momento en campos y estudios de muy diversa naturaleza.

5.2.1.1. Diseño del Análisis

El diseño de un Análisis Factorial requiere inicialmente adoptar una serie de decisiones básicas respecto al número de variables a incluir en el estudio, propiedades de medición de las mismas, tipo de variables permisibles, tamaño de muestra necesario, etc. y finalmente estimar los datos de entrada (bien como matriz de correlaciones, bien como matriz de varianzas-covarianzas) a utilizar para alcanzar los objetivos propuestos.

La elección de los coeficientes de correlación dependerá de la escala de medida que se haya utilizado en las variables incluidas en el estudio:

- Si las variables se miden en escala de intervalo y siguen una distribución aproximadamente normal, se puede utilizar el coeficiente de correlación momento-producto de Pearson, típico en el cálculo de la matriz de correlaciones.
- Si las variables son categóricas, puede ser más apropiado utilizar algún coeficiente que se base en el estadístico chi-cuadrado, como el C de contingencia o V de Cramer y, en el caso de variables dicotómicas recurrirse a los coeficientes phi (si la dicotomía es absoluta), biserial puntual (si existen variables continuas y dicotómicas puras), biserial (si existen variables continuas y dicotomizadas) y tetracórico (si sólo existen variables dicotomizadas).
- Si todas las variables son ordinales discretas puede trabajarse con la matriz asintótica de varianzas-covarianzas (Jöreskog y Sörbom, 1989), especialmente indicada para correlaciones policóricas.

En cualquier caso, debe considerarse que la disminución del rango de respuestas en una determinada escala puede afectar al AF, ya que un conjunto de variables con pocas categorías tiende a obtener correlaciones más bajas que si el número fuese más elevado.

En nuestro caso, se trabajará tanto con matriz de correlaciones como con la matriz de varianzas-covarianzas, comparándose los resultados obtenidos en ambos casos.

5.2.1.2 Selección del método de factorización

Como ya se ha comentado, existen dos posibles métodos para efectuar la extracción de los factores: el ACP y el AFC. Para poder comprender las diferencias entre ellos, es necesario definir previamente los tres posibles tipos de varianza que pueden considerarse para los propósitos del análisis factorial: varianza común, varianza específica y varianza de error.

La varianza común es la varianza de una variable que se comparte con todas las otras variables en el análisis. La varianza específica, también denominada única, es la varianza asociada únicamente a la variable considerada. Finalmente, la varianza de error es la varianza debida a la poca fiabilidad en el proceso de recolección de datos, a errores de medición o a un componente aleatorio en el fenómeno medido. En este sentido, se denomina comunalidad a la estimación de la varianza compartida o común entre las variables.

La matriz de correlaciones nos proporciona una información inicial sobre la variabilidad observada en el conjunto de variables del estudio. Así, los valores de la diagonal principal representan la varianza total de cada una de las variables, mientras que la huella de la matriz de correlaciones, es decir, la suma de los elementos de la diagonal principal, nos indica la cantidad de varianza contenida en esa matriz. En función del método de factorización utilizado, en los valores de la diagonal principal, pueden estar incluidas la varianza común o comunalidad, o bien, la varianza total, como se muestra en la figura 5.2.

El objetivo del AF es básicamente encontrar unos factores que expliquen la varianza común de las variables. Para ello, deben estimarse las comunalidades y la estimación de éstas dependerá del método de extracción factorial que estemos utilizando.

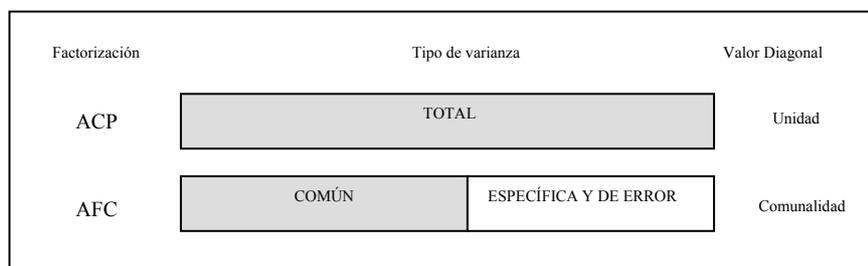


Figura 5.2. Tipos de varianza en el AF

Con el objetivo común de alcanzar la estructura simple de Thurstone (Thurstone, 1935 y 1947), los diferentes métodos de extracción factorial intentan determinar el número mínimo de factores comunes capaces de reproducir la varianza observada en la matriz de correlaciones inicial. La elección de un procedimiento u otro, dependerá, entre otros aspectos, del enfoque utilizado en la extracción, el método de estimación de las comunalidades, la posibilidad de cumplir la propiedad de permanecer invariante ante el cambio de escala, la utilización de estadísticos de ajuste y el que las puntuaciones factoriales se estimen o sean calculadas.

Aunque se tratará con más detalle posteriormente, avanzamos aquí que el ACP analiza la varianza total y las comunalidades iniciales toman el valor de la unidad. En el AFC, se analiza la varianza común y las comunalidades se calculan bien, mediante la estimación de la correlación múltiple de una variable frente a todas las restantes, bien mediante procedimientos iterativos que toman en cuenta el número de factores inicialmente fijados. La ventaja de este último procedimiento es que reduce la cantidad de varianza explicada por la matriz de correlaciones, puesto que la comunalidad de una variable será menor que la unidad, es decir, que con un número menor de factores puede explicarse la varianza de la matriz.

El objetivo principal del ACP es reducir el número de variables originales a unas pocas componentes, de manera que estas nuevas componentes constituyan nuevas variables y el número de componentes retenidas consiga explicar la máxima cantidad de varianza en los datos. El objetivo del AFC es identificar los factores subyacentes, también denominados

constructos o dimensiones latentes, que pueden explicar las intercorrelaciones entre las variables (Sharma, 1996). Por otro lado, el ACP, pone énfasis en explicar la varianza contenida en los datos, mientras el AFC enfatiza el análisis de la correlación entre constructos o variables observables.

En el ACP, las relaciones entre componentes y variables o indicadores, son de la forma:

$$\begin{aligned}\xi_1 &= \omega_{11}\chi_1 + \omega_{12}\chi_2 + \dots + \omega_{1p}\chi_p \\ \xi_2 &= \omega_{21}\chi_1 + \omega_{22}\chi_2 + \dots + \omega_{2p}\chi_p \\ &\dots\dots\dots \\ \xi_m &= \omega_{m1}\chi_1 + \omega_{m2}\chi_2 + \dots + \omega_{mp}\chi_p\end{aligned}$$

donde ξ_1 es el componente formado por las variables χ_1, \dots, χ_p .

En el AFC, las variables o indicadores reflejan la presencia de constructos o factores no observables, siendo las ecuaciones de la forma:

$$\begin{aligned}\chi_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \dots + \lambda_{1m}\xi_m + \varepsilon_1 \\ \chi_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \dots + \lambda_{2m}\xi_m + \varepsilon_2 \\ &\vdots \\ \chi_p &= \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \dots + \lambda_{pm}\xi_m + \varepsilon_p\end{aligned}$$

donde las variables χ_1, \dots, χ_p son función de las variables latentes o constructos ξ_1, \dots, ξ_m y de factores únicos.

En consecuencia, el ACP sigue un principio de extracción que supone maximizar la varianza explicada, es decir, conseguir que la contribución del factor a algunas de las comunalidades de las variables del estudio sea máxima. De ahí que el factor que represente mayor variabilidad se convierta en el primer componente principal, el que representa mayor variabilidad en la matriz de residuales será el segundo componente principal, y así sucesivamente. Este procedimiento analiza la varianza total, de ahí que no se estimen las comunalidades y se tome como punto de partida el valor de la diagonal principal de la matriz de correlaciones. Presenta el inconveniente de no permitir utilizar en un mismo análisis, variables medidas en escalas diferentes si no se estandarizan previamente y tampoco utiliza criterio estadístico alguno para fijar el número de factores significativos.

Frente a ello, presenta las ventajas de calcular directamente las puntuaciones factoriales y, de no necesitar estimar las comunalidades iniciales, por ser éstas, la unidad.

La aplicación del AFC, puede adoptar distintas formas según los criterios utilizados para conseguir al máximo ajuste (o mínima discrepancia) entre el modelo y los datos: método de máxima verosimilitud (MV), método de los mínimos cuadrados, método imagen, factorización alpha, etc. Las ventajas e inconvenientes de estos métodos se muestran la tabla 5.1.

Tabla 5.1. Comparación de procedimientos de extracción directa

TIPO ANÁLISIS	PRINCIPIO DE EXTRACCIÓN	COMUNALIDAD ESTIMADA	ESCALA INVARIANTE	PUNTUACIONES FACTORIALES	TEST PARA EL N° DE FACTORES
COMPONENTES PRINCIPALES	Maximiza la varianza explicada	No es necesaria, por ser la unidad	No	Calculadas	No
FACTOR PRINCIPAL	Maximiza la varianza explicada	Numerosos procedimientos	No	Estimadas	No
RESIDUALES MÍNIMOS	Mínimiza las correlaciones residuales fuera de la diagonal	Repite las estimaciones apropiadas	No	Estimadas	No
IMAGEN	Covarianza de las estimaciones de cada variable de regresión sobre las demás	SMC	No	Calculadas	No
ALFA	Maximiza la generabilidad de los factores	Iterativo (converge hasta el mejor valor)	Sí	Calculadas	No
MÁXIMA VEROSIMILITUD	La mejor estimación de la matriz de los factores reproducida en la población	Iterativo (converge hasta el mejor valor)	Sí	Estimadas	Sí
MÁXIMA VEROSIMILITUD CANÓNICA	Factores que correlacionan en un grado máximo con las variables en tanto que dan la mejor estimación de la matriz de correlaciones reproducida en la población	Iterativo (converge hasta el mejor valor)	Sí	Estimadas con el R^2 más alto posible con los factores	Sí

De todos estos procedimientos cabe destacar aquí el de Máxima Verosimilitud, ya que será el método utilizado para efectuar el Análisis Factorial Confirmatorio.

El principio de extracción de Máxima Verosimilitud (MV), considera la mejor estimación posible de la matriz de correlaciones reproducida en la población como principio de extracción. Utiliza el estadístico chi-cuadrado para determinar el grado de ajuste entre lo observado y lo estimado, lo que permite determinar el número de factores a extraer. La estimación de las comunalidades se realiza por un procedimiento iterativo que

converge en el mejor valor. Permanece invariante al cambio de escala y estima las puntuaciones factoriales.

En un estudio comparativo de los procedimientos de máxima verosimilitud y componentes principales realizado por Pérez Gil y Moreno (1991), se concluye que el método de máxima verosimilitud del análisis del factor común permite delimitar con mayor precisión la significación de la varianza entre factores hipotetizados y variables externas. Sin embargo, en un estudio posterior realizado por Hair et al. (1995), se obtuvo que con ambos procedimientos (Análisis de Componentes Principales y Máxima Verosimilitud) se obtienen resultados prácticamente idénticos, siempre que las varianzas compartidas en la mayoría de las variables exceden de 0.6. A continuación se describen más detalladamente ambos procedimientos, por ser los que aplicaremos en nuestro estudio posteriormente.

a) Análisis de Componentes Principales

Llamaremos X a la matriz de datos que está formada por n observaciones (filas) y p variables (columnas). Cada una de las observaciones está descrita por p variables o características $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$, es decir, cada muestra está representada por un punto en un espacio p -dimensional.

El objetivo del Análisis de Componentes Principales (ACP) es reducir la dimensionalidad de un conjunto de datos formado por un número elevado de variables interrelacionadas, para obtener unas pocas variables nuevas, de tal forma que éstas retengan la mayor parte de la variabilidad presente en el conjunto de datos. Las nuevas variables obtenidas se llaman variables latentes o componentes principales y son combinaciones lineales de las variables originales y ortogonales entre sí. Debido a la ortogonalidad, cada dirección en el nuevo espacio describe una nueva fuente de variación entre las muestras.

En estadística la palabra latente significa que no es manifiesta, es decir, que no puede medirse directamente. Las componentes principales resumen los patrones sistemáticos de variabilidad entre muestras. Se calculan a partir de los datos iniciales con una estructura que se ha denominado en ocasiones el modelo bilineal.

El ACP es un método bilineal que se ajusta a los datos, de acuerdo con el criterio de mínimos cuadrados clásico, con el objetivo de describir la máxima variabilidad contenida en esos datos. Las componentes principales representan los valores propios más grandes de la matriz de varianzas-covarianzas de los datos originales. De esta forma, con unos pocos componentes suele poder describirse la mayor parte de la información de los datos. Así, las redundancias son resumidas y se simplifica la interpretación de los datos.

Este método permite explicar la varianza de una matriz (Jackson, 1991) y opera de la forma siguiente: en primer lugar ACP calcula un vector que indica la dirección sobre la cual los puntos presentan la máxima variabilidad. Esta es la dirección de la primera componente principal. Se calcula como la dirección de una recta que hace que las distancias entre los puntos originales y sus proyecciones sobre la recta sean lo más pequeñas posibles. Puede comprobarse que esta dirección se obtiene calculando la combinación lineal de las columnas de la matriz X dada por el vector propio \bar{p}_1 de la matriz $X'X$ asociado al mayor valor propio. Esta dirección se llama \bar{h}_1 y es:

$$\bar{h}_1 = p_{11}\bar{X}_1 + p_{12}\bar{X}_2 + \dots + p_{1p}\bar{X}_p$$

La forma en que una variable j contribuye a la primera componente principal se llama carga de esa variable (*loading*) p_{1j} . La nueva dirección puede interpretarse como una propiedad principal de las muestras. Esta propiedad no puede ser medida directamente y sin embargo explica la máxima variabilidad entre las muestras.

Después de proyectar los puntos en la primera componente principal, puede calcularse cómo este vector representa los datos. La suma de todas las distancias entre los puntos originales y su proyección es una medida de la variación que no es explicada por la primera componente principal. Se busca entonces una segunda componente principal para describir la variabilidad restante. Esta segunda componente es ortogonal a la primera y explica la máxima variabilidad restante. Se obtiene ajustando una línea a través de los residuos resultantes del ajuste de la primera componente principal, o también como la combinación lineal de las columnas de X dada por el vector propio \bar{p}_2 de la matriz $X'X$ asociado al segundo mayor valor propio.

Este procedimiento puede repetirse hasta que toda la información de los datos es descrita por J componentes principales, donde $J = \text{Min}(n, p)$ (asumiendo que el rango de la matriz X es J). Así, los objetos pueden representarse en un nuevo espacio J-dimensional con ejes ortogonales. Sin embargo, aunque se describe toda la variabilidad, no se reduce la dimensionalidad del problema, ya que sólo se han rotado los ejes de X para obtener una nueva base ortonormal. Pero, con frecuencia, en conjuntos grandes de datos suele ocurrir que, después de extraer las A primeras componentes principales, con $A \ll J$, la mayor parte de la variabilidad de X ha sido explicada. La pequeña cantidad de variabilidad restante se considera ruido y es, por lo tanto, deseable separarlo de la variabilidad total de los datos.

Utilizando estas A primeras componentes principales y resumiendo la información de X mediante nuevas variables \bar{t}_a ($a=1, \dots, A$) definidas como combinaciones lineales de las p variables originales se reduce la dimensión del espacio de p a A. Las nuevas variables se calculan como $\bar{t}_a = X\bar{p}_a$ y con ellas puede descomponerse la matriz X de la forma:

$$X = \bar{t}_1\bar{p}'_1 + \bar{t}_2\bar{p}'_2 + \dots + \bar{t}_A\bar{p}'_A + E$$

donde E es la matriz residual. Si la nueva dimensión A se escoge de forma que E no contenga información relevante de los datos originales, es decir, que sólo represente ruido, entonces:

$$\hat{X} = \bar{t}_1\bar{p}'_1 + \bar{t}_2\bar{p}'_2 + \dots + \bar{t}_A\bar{p}'_A$$

En esta expresión, se llaman cargas (*loadings*) a los vectores unitarios \bar{p}_i asociados a las componentes principales. Estos vectores definen la posición del nuevo espacio en el espacio de las p variables originales. Cada muestra individual se coloca en el nuevo espacio calculando las proyecciones de esa muestra sobre cada una de las A direcciones principales. A esos valores se les denomina puntuaciones (*scores*). Cada puntuación se calcula como el producto escalar entre el vector carga y la expresión de la muestra en el espacio original.

Es decir, ACP obtiene de la matriz X de n filas y p columnas, los A primeros vectores propios, $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_A$, de acuerdo al orden decreciente de sus valores propios

asociados. A partir de ellos se construye la matriz de cargas de p filas y A columnas que denotaremos por $P = [\bar{p}_1, \bar{p}_2, \dots, \bar{p}_A]$ y después se proyecta la matriz X en el nuevo espacio de variables latentes mediante la expresión $T = XP$. En la Figura 5.3 se muestra una representación de los cálculos matriciales de un modelo ACP obtenido de la forma descrita.

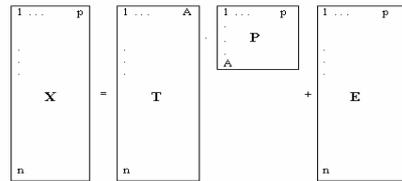


Figura 5.3. Modelo ACP con A componentes. X es la matriz de datos, T la matriz de puntuaciones, P es la matriz de cargas y E la matriz de residuos

b) Máxima Verosimilitud

El objetivo global del procedimiento de máxima verosimilitud (MV), es encontrar la solución factorial que mejor se ajustaría a las correlaciones observadas (Kim y Mueller, 1978). Según estos autores, el método MV (Lawley y Maxwell, 1971; Jöreskog, 1967; Jöreskog y Lawley, 1968), asume que los datos observados constituyen una muestra procedente de una población donde se aplica exactamente un modelo de factor común k , y donde la distribución de las variables es normal multivariante.

Lo que se desconoce, es el valor de las correlaciones entre las variables o indicadores, de manera que este procedimiento pretende estimar los parámetros que tengan la mayor verosimilitud de reproducir poblacionalmente las correlaciones observadas. La solución propuesta por Jöreskog (1967), puede expresarse en función del siguiente determinante:

$$|R_1 - \lambda I|$$

donde λ es un multiplicador de Lagrange y R_1 viene dado por:

$$R_1 = U^{-1}(R - U^{-1})U^{-1} = U^{-1}R U^{-1}$$

donde U^{-1} es la varianza específica en cada etapa y R_1 es la matriz ajustada de covarianzas.

Este procedimiento trabaja con la matriz ajustada en lugar de con la original y la reajusta en cada iteración, de manera que el mayor peso se atribuye a las correlaciones con menos varianza específica. En este procedimiento, la varianza específica se considera como una varianza casi de error, y por ello atribuye mayor peso a las variables con mayor comunalidad.

En este procedimiento suele utilizarse un test de ajuste que se distribuye como una variable chi-cuadrado. Su valor es directamente proporcional al tamaño de la muestra, mientras los grados de libertad, son una función del número de factores y del número de variables. El estadístico chi-cuadrado viene dado por:

$$U_k = N \{ \ln|C| - \ln|R| + \text{tr}(RC^{-1}) - p \}$$

donde,

\ln , es el logaritmo natural;

tr , es la traza de la matriz;

N es el tamaño de la muestra;

p , es el número de variables;

R , es la matriz de covarianzas;

$C = FF' + U^{-1}$, siendo F los pesos factoriales y U^{-1} la varianza específica.

Los grados de libertad asociados vienen dados por la expresión:

$$gl_k = \frac{1}{2} [(p-k)^2 - (p+k)]$$

donde k es el número de factores y p es el número de variables.

Para que el estadístico chi-cuadrado tenga grados de libertad positivos, el número de factores comunes no puede exceder al entero mayor que satisfaga (Norusis, 1986):

$$k < 0,5(2p+1-\sqrt{8+1})$$

Como en la mayoría de los análisis exploratorios se desconoce el número de factores a extraer, este se va aumentando hasta obtener una bondad de ajuste razonable, lo que puede provocar que el número de factores extraídos sea mayor que el necesario.

La ventaja principal de este método es que proporciona un buen nivel de significación para muestras grandes.

El método MV se presenta como un procedimiento adecuado para ser aplicado en el caso de poblaciones sin factores poco representativos, pero no es el más adecuado cuando las desviaciones pequeñas se incluyen en el modelo, dado que si el tamaño muestral es elevado, cualquiera de estas desviaciones puede ser considerada como una dimensión significativa, al margen de la variabilidad muestral.

El procedimiento de MV descrito resulta muy costoso cuando se trabaja con un número elevado de variables, ya que implica comenzar comprobando la adecuación del modelo obtenido con un solo factor común y verificar posteriormente si los datos se desviarían significativamente del modelo cuando consideramos dos factores en lugar de uno, y así sucesivamente. Por ello suele recomendarse que para determinar el número de factores comunes, que además del test de máxima verosimilitud, se utilicen otros procedimientos como la especificación de autovalores mayores que uno, el de la importancia sustantiva de un factor, la prueba de rocalla o *scree-test* desarrollada posteriormente, o bien, los criterios del propio investigador.

5.2.1.3 Criterios de Determinación del Número de Factores a Extraer

Para determinar el número de factores a extraer no se suele utilizar un único criterio. En su lugar, se suele utilizar inicialmente un criterio de extracción predeterminado de los que se definen a continuación, como orientación en un primer intento de interpretación. Después de calcular la solución inicial, conviene calcular otras soluciones adicionales con distinto número de factores, en función de las características del estudio y de los conocimientos previos que se posean del mismo.

Los criterios más utilizados para determinar el número inicial de factores a extraer son:

- *Criterio de Autovalores o Eigenvalue*, que se basa en el criterio de que la varianza extraída por cada componente debería, como mínimo, justificar la varianza de por lo menos una variable. Cada variable contribuye con un valor de uno para el autovalor total, por lo que sólo se considerarán los factores que tienen autovalores mayores que uno. Este criterio presenta el inconveniente de que si el número de variables en el estudio es inferior a 20, extrae un número de factores demasiado pequeño y, por el contrario, si el número de variables es superior a 50 suele extraer demasiados, siendo recomendable utilizarlo junto con otros criterios o heurísticas.
- *Criterio de porcentaje de varianza*, que es una aproximación basada en obtener un porcentaje acumulado especificado de la varianza total extraída. El propósito es asegurar una significación práctica de los componentes o factores derivados, asegurando que expliquen por lo menos una cantidad especificada de la varianza. Sin embargo, no se ha adoptado un umbral absoluto para todas las aplicaciones.
- *Criterio de contraste de caída o scree plot*, propuesto por Cattell (1972). Se utiliza para determinar el número de factores a extraer antes de que la cantidad de varianza única, que es sustancialmente superior en los factores posteriores que en los primeros, empiece a dominar la estructura de la varianza común. El *scree plot*, también denominado gráfico de sedimentación, se obtiene uniendo los autovalores obtenidos en función de su orden de extracción. La forma de la curva obtenida se utiliza para evaluar el número de factores a extraer, deteniendo la factorización cuando comienza el nivel de corte a formar una línea recta, o al menos con un declive horizontal. Por regla general, este criterio suele incluir uno y, a veces, dos o más factores que cuando se utiliza el criterio de autovalores.

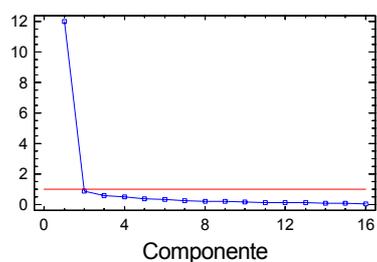


Figura 5.4. Scree plot

5.2.1.4 Criterios de Rotación

La matriz factorial obtenida inicialmente nos proporciona una idea preliminar sobre el número de factores a extraer. En ella se obtiene la mejor combinación lineal de las variables, es decir, aquella combinación particular de las variables originales que explicaría el mayor porcentaje de varianza de los datos.

Las cargas factoriales que aparecen en la matriz, son las correlaciones de cada variable con cada factor. En la matriz inicial, el primer factor suele tener cargas factoriales bastante elevadas de todas las variables, ya que éste se extrae de manera que explique el mayor porcentaje de la varianza de los datos. El segundo factor, será el que mayor porcentaje de varianza residual explique, y así sucesivamente, ya que se debe cumplir el requisito de mantener la ortogonalidad entre los factores, de manera que los últimos factores suelen tener cargas factoriales bastante bajas para todas las variables.

Para solucionar el problema de la situación topológica de los factores, Thurstone (1935) propuso que fueran rotados hasta alcanzar una explicación factorial lo más sencilla posible.

Su principio de estructura simple, definido en términos de la obtención de factores con algunas saturaciones altas y muchas saturaciones bajas, se apoya en cinco criterios establecidos en 1947 (Torrens-Ibern, 1972):

- Todas las filas de la matriz factorial deben tener al menos un cero.

- Si ésta tuviese k factores comunes, todas las columnas de la matriz factorial deberían tener al menos k ceros.
- Para cada par de columnas de la matriz factorial deben anularse muchas variables en una columna, pero no en la otra.
- Para cada par de columnas de la matriz factorial deben anularse una gran proporción de variables en las dos columnas cuando haya cuatro o más factores.
- Para cada par de columnas de la matriz factorial debe haber un pequeño número de variables que se anulen en las dos columnas.

Para alcanzar esta estructura debemos hallar una matriz de transformación T que, tras ser multiplicada por la matriz factorial no rotada, A , sea capaz de generar otra matriz factorial rotada, F , que cumpla el principio de Thurstone. Esta nueva matriz F , deberá representar adecuadamente la información contenida en la matriz de correlaciones originales R , ya que

$$\begin{aligned}AT &= F \\ F.F' &= R\end{aligned}$$

La rotación de la matriz factorial inicial, suele ser un procedimiento adecuado para alcanzar este objetivo. Consiste en girar en el origen los ejes de referencia de los factores hasta alcanzar una determinada posición, de manera que se consiga redistribuir la varianza de los primeros factores a los últimos para lograr un patrón de factores más simple y teóricamente significativo.

Existen dos posibles métodos de rotación: ortogonal y oblicua, dependiendo de que los ejes se mantengan en ángulos rectos en la rotación, o bien, se sitúen en ángulos diferentes.

Existen distintos posibles algoritmos para efectuar la rotación ortogonal:

- *Método Quartimax*, que minimiza el número de factores necesarios para explicar una variable. El objetivo último de este procedimiento es simplificar las filas en

la matriz de factores, es decir, efectúa la rotación de los factores iniciales de manera que una variable cargue alto sobre un factor y tan bajo como sea posible sobre los otros factores. La aplicación del criterio quartimax da lugar a una rotación de ejes de modo que los pesos factoriales maximicen la medida total de la noción de simplicidad factorial (q), que equivale a la maximización de los siguientes términos:

$$Q = \sum_{i=1}^p \sum_{j=1}^k b_{ij}^4$$

donde b_{ij} es el peso factorial de la i -ésima variable en el j -ésimo factor.

El nombre de quartimax refleja el intento de lograr una matriz factorial donde la suma de los pesos factoriales de las variables elevadas a la cuarta potencia represente un máximo.

- *Método Varimax*, cuyo criterio se centra en simplificar las columnas de la matriz factorial, es decir, intenta lograr para cada par de factores, la rotación que haga máxima las varianzas de los cuadrados de las saturaciones ponderadas. En este caso la medición de simplicidad para el conjunto de los factores viene dada por:

$$V = \sum_{j=1}^k v_j = \frac{\sum_{j=1}^k p \sum_{i=1}^p b_{ij}^4 - \left(\sum_{i=1}^p b_{ij}^2 \right)^2}{p^2}$$

donde p es el número de variables.

Para minimizar la posible influencia de pesos iniciales altos en la solución resultante, es habitual utilizar pesos factoriales normalizados.

- *Método Equamax*, que busca maximizar simultáneamente el peso ejercido por las variables y factores, simplificando filas y columnas en la matriz factorial. Para ello aplica los dos criterios anteriores:

$$\alpha Q + \beta V = \text{Máximo}$$

donde Q y V son las expresiones utilizadas para definir los criterios quartimax y varimax respectivamente, y donde α y β están ponderados. Esto puede expresarse (Kim y Mueller, 1978):

$$\sum_{j=1}^k \left(\sum_{i=1}^p b_{ij}^2 \right)^2 - \Gamma \left(\sum_{i=1}^p b_{ij}^2 \right)^2 / p = \text{Máximo}$$

donde $\Gamma = \beta / (\alpha + \beta)$.

La rotación oblicua implica la existencia de correlaciones entre los factores, siendo más general que la rotación ortonormal que impone que los factores tengan correlaciones nulas entre sí. Existen básicamente dos enfoques para la rotación oblicua: uno que utiliza los ejes de referencia y otro que emplea el modelo de matriz primaria o modelo factorial de los factores primarios, que se detallan a continuación.

- *Soluciones basadas en los ejes de referencia*, que se apoyan en la idea de la existencia de grupos definidos de variables que representan dimensiones separadas y correctamente identificadas por factores primarios. Cada grupo de variables tendría, por tanto, una proyección cercana a cero en todos los ejes menos en uno. Este criterio es denominado quartimin y se expresa como:

$$N = \sum_{i=1}^p \sum_{j < h=1}^k a_{ij}^2 a_{ih}^2$$

donde a_{ij} y a_{ih} son proyecciones del j-ésimo y el h-ésimo eje de referencia. Este valor será cero si todas las variables tienen pesos en un solo factor, si bien, esto no suele suceder y lo que se pretende es buscar los pesos factoriales que minimicen N.

Si lo que intentamos es minimizar la covarianza de los elementos al cuadrado de las proyecciones sobre los ejes de referencia, el criterio se denomina covarimin o biquartimin y su expresión quedaría:

$$C = \sum_{j < h=1}^k \left(p \sum_{i=1}^p a_{ij}^2 a_{ih}^2 - \sum_{i=1}^p a_{ij}^2 \sum_{i=1}^p a_{ij}^2 \right)$$

Cuando se aplican a un mismo conjunto de datos el criterio covarimin tiende a producir muy pocos factores oblicuos mientras que el quartimin procura más factores.

Una solución más adecuada puede obtenerse combinando los dos criterios anteriores, que es lo que se conoce como el criterio oblimin general o indirecto:

$$B = \sum_{i=h=1}^k \left(p \sum_{i=1}^p a_{ij}^2 a_{ih}^2 - \Gamma \left(\sum_{i=1}^p a_{ij}^2 \sum_{i=1}^p a_{ij}^2 \right) \right)$$

En general el criterio se reduce a quartimin cuando $\Gamma = 0$ (más oblicuo); bicuartimin cuando $\Gamma = 0,5$ y covarimin cuando $\Gamma = 1$ (menos oblicuo).

- *Soluciones basadas en el modelo de los factores primarios*, que se apoya en la significación de los pesos factoriales sobre los factores primarios. El valor minimizado se corresponde con el establecido en la ecuación del criterio oblimin general, con la diferencia de que se utilizan los pesos factoriales primarios de la matriz factorial, en lugar de los pesos sobre los ejes de referencia. En este tipo de rotación, denominada oblimin directo se busca los pesos factoriales que minimicen B:

$$B = \sum_{i=h=1}^k \left(\sum_{i=1}^p b_{ij}^2 b_{ih}^2 - \delta \left(\sum_{i=1}^p b_{ij}^2 \sum_{i=1}^p b_{ij}^2 \right) / p \right)$$

donde b_{ij} y b_{ih} son los pesos factoriales en el modelo factorial y p es el número de variables. El grado de oblicuidad de la solución final viene representado por

δ . Los valores más grandes de δ generarán soluciones más oblicuas y los valores más pequeños de δ (negativos) generarán soluciones menos oblicuas.

Comentar, que las soluciones ortogonales y oblicuas no modifican las comunalidades de las variables ni determinan cambios en la cantidad de varianza explicada por ellas, sólo realizan una mejor distribución de la cantidad de varianza explicada, es decir, una mejor distribución de los valores de la matriz de correlaciones original. Entre ambos procedimientos los resultados no difieren significativamente y no se producen modificaciones en la interpretación de los factores apoyada en esas rotaciones (García Jiménez et al., 2000).

5.2.1.4. Síntesis de la metodología general del Análisis Factorial

Podemos resumir lo expuesto en este apartado en el esquema general que se muestra en la figura 5.5. La numeración de los apartados que siguen se remite a ese esquema gráfico.

1. Matriz datos originales. En esta matriz, las filas representan a los individuos y las columnas a las variables, y no juegan papeles simétricos: medias y varianzas no tienen sentido en general más que para las columnas.
2. Los datos deben ser centrados y, en su caso, tipificados.
3. En la aplicación del AF, la tabla debe considerarse como una yuxtaposición de columnas.
4. Cada individuo, puede representarse por un punto en el espacio R^p , siendo P el número de variables. En esa nube N_n de individuos, deberemos analizar las distancias entre unos individuos u otros, interpretándolas como semejanzas. El centrado ha hecho coincidir el origen de los ejes y el centro de gravedad de la nube de individuos.

5. Cada variable también puede representarse por un vector de \mathbb{R}^N , siendo N el número de individuos. En la nube N_P de variables, nos interesan fundamentalmente los ángulos que forman las variables entre sí. El coseno del ángulo que forman dos variables se interpreta como coeficiente de correlación entre ambas. La tipificación, hace que todas las variables equidisten del origen y aparezcan situadas sobre una hipersfera de radio 1.
6. El AF de una nube consiste en poner de manifiesto una sucesión de direcciones tales que la inercia con respecto a O , de la proyección de la nube sobre esas direcciones sea máxima. En \mathbb{R}^P donde el origen O coincide con el centro de gravedad G , los ejes factoriales son las direcciones de máximo alargamiento de N_N en \mathbb{R}^N donde la proyección de una variable sobre otra se interpreta como su mutuo coeficiente de correlación, los ejes factoriales son las variables sintéticas más relacionadas con el conjunto de variables iniciales.
7. El plano factorial que cruza dos factores de variables, denominados $G_1(P)$ y $G_2(P)$, aporta una imagen aproximada de N_P en \mathbb{R}^N . Las coordenadas de una variable se interpretan como su coeficiente de correlación con los factores de individuos.
8. El plano factorial que cruza dos factores de individuos, denominados $F_1(N)$ y $F_2(N)$, aporta una imagen aproximada de N_N en \mathbb{R}^I . Las distancias entre dos puntos se interpretan como desemejanzas.
9. Las relaciones de transición expresan los resultados del AF en relación con los resultados del dual.
10. Como consecuencia de las relaciones de transición, las interpretaciones de los ejes factoriales deben efectuarse conjuntamente. Resulta habitualmente cómoda la superposición de ambas representaciones.

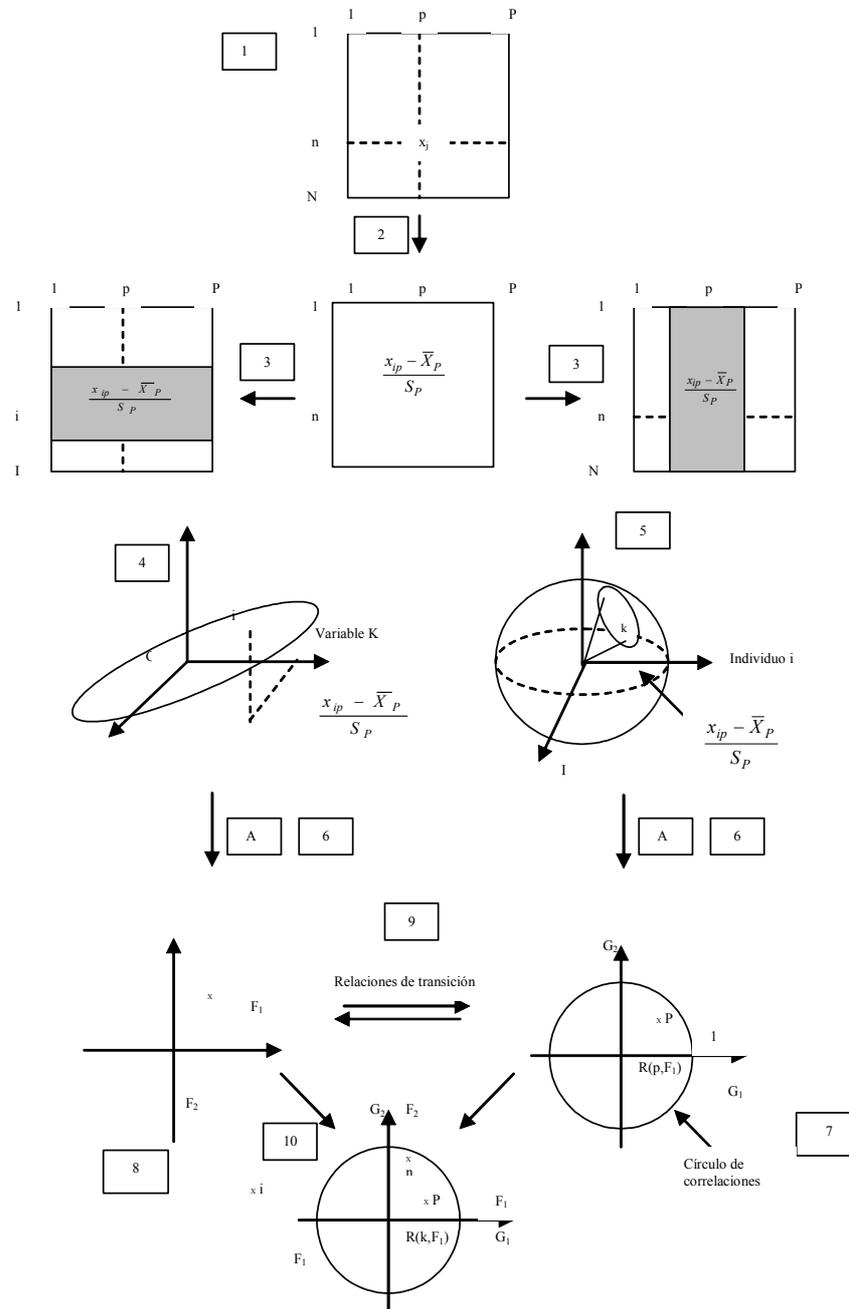


Figura 5.5. Esquema del AF

El análisis factorial desarrollado hasta el momento es denominado análisis factorial exploratorio, ya que a priori, se dispone de poca información acerca de la estructura factorial: número de factores o dimensiones, si los factores son ortogonales u oblicuos, número de variables indicadoras de cada factor, el indicador que está más correlacionado con cada dimensión, etc. Este análisis exploratorio nos permite detectar la estructura factorial que puede explicar las correlaciones entre los indicadores. La siguiente etapa será aplicar un análisis factorial con fines confirmatorios.

5.2.2. Análisis Factorial Confirmatorio: Modelos de Ecuaciones Estructurales

Cuando se asume que la estructura factorial es conocida de antemano, la técnica multivariante a utilizar es el análisis factorial confirmatorio, en la cual debe especificarse a priori la estructura factorial completa, incluyendo los respectivos indicadores de cada dimensión y las correlaciones entre ellos. En el presente trabajo, se utilizará la técnica del Análisis Factorial Confirmatorio (CFA), sobre la base de construcción del cuestionario y sobre una base empírica o experimental obtenida mediante los estudios exploratorios anteriores, como técnica para verificar la dimensionalidad del CEDA (Jöreskog, 1969, 1978), de manera que se verifique que el cuestionario capta las dimensiones básicas, que se consideran “claves” para lograr una enseñanza de calidad.

En este trabajo, se aplicará la técnica de modelización confirmatoria mediante los modelos de ecuaciones estructurales (SEM), que fueron definidos en el campo de la econometría en 1973 por Goldberger y Duncan. En las últimas décadas, otros autores entre los que cabe citar a Jöreskog, Batista –Foguet y Coenders, han continuado desarrollando y extendiendo esta herramienta a otros campos de investigación, debido a que numerosos fenómenos de interés son complejos, engloban muchos aspectos, obedecen a múltiples causas y están frecuentemente medidos con error, siendo necesario la utilización de técnicas multivariantes como los modelos de ecuaciones estructurales, que permiten incorporar el error de medida y considerar relaciones recíprocas entre constructos.

En general, los modelos de ecuaciones estructurales permiten alcanzar los cinco objetivos siguientes:

- Trabajar con conceptos no observables a los que denominaremos constructos, que se miden a través de indicadores, para evaluar posteriormente la calidad de dicha medición.
- Considerar los fenómenos en su verdadera complejidad, desde una perspectiva más realista, incorporando múltiples variables tanto endógenas como exógenas.
- Considerar conjuntamente medida y predicción, es decir, evaluar los efectos de variables latentes entre sí, sin contaminación posible, debido a los errores de medida.
- Introducir la perspectiva confirmatoria en las técnicas estadísticas, ya que en la especificación de los modelos, el investigador debe introducir su conocimiento teórico.
- Descomponer las covarianzas observadas y, no únicamente las varianzas, dentro de una perspectiva del análisis de la interdependencia.

Aunque el resultado último de SEM es siempre la evaluación de una serie de relaciones, esta técnica multivariante constituye una poderosa herramienta apropiada para muchos objetivos de investigación, pudiendo destacar tres estrategias distintas de su aplicación, en función de cómo haya sido definida la investigación: estrategia de modelización confirmatoria, estrategia de modelos rivales y estrategia de desarrollo del modelo (Hair et al., 1995).

En realidad, los modelos de ecuaciones estructurales, abarcan toda una familia de modelos denominados análisis de estructura de la covarianza, análisis de variables latentes, análisis del factor confirmatorio, etc. Sin embargo, todas estas técnicas se distinguen por:

- Estimar relaciones múltiples y cruzadas, es decir, utiliza relaciones distintas para cada conjunto de variables, pero interrelacionadas mediante la especificación de lo que denominaremos modelo estructural. Este modelo estructural expresa las relaciones existentes entre variables dependientes e independientes. Es de naturaleza interdependiente, ya que algunas variables dependientes en una determinada relación, pueden convertirse en independientes en relaciones

ulteriores. Posteriormente, las relaciones propuestas en el modelo estructural serán trasladadas a series de ecuaciones estructurales para cada variable dependiente.

- Representar conceptos no observados, denominados variables latentes o constructos, en estas relaciones y tener en cuenta el error de medida en esa representación, lo que permitirá mejorar la estimación estadística.

Estimación de los coeficientes estructurales

Cada coeficiente de regresión utilizado en los SEM está compuesto, en realidad, por dos elementos: el “verdadero coeficiente estructural” entre la variable dependiente e independiente y la fiabilidad (o medida en que la variable independiente está libre de error) de la variable predictora. El impacto del error de medida puede observarse en la expresión del coeficiente de regresión (Hair et al., 1995):

$$\beta_{x,y} = \beta_s \cdot \rho_x$$

donde $\beta_{x,y}$ es el coeficiente de regresión observado, β_s es el verdadero coeficiente estructural y ρ_x es la fiabilidad de la variable predictora.

Dado que todas las relaciones de dependencia se basan en la correlación observada resultante del coeficiente de regresión entre las variables, se conseguirá mejorar las correlaciones utilizando un modelo de dependencia para realizar estimaciones más precisas de los coeficientes estructurales, teniendo en cuenta la correlación atribuible a cualquiera de los problemas de medida.

Representación de conceptos teóricos

Actualmente, en numerosas investigaciones existe la necesidad de medir conceptos tales como la satisfacción de los usuarios, motivaciones de comportamiento de los clientes etc., es decir, conceptos más bien teóricos o abstractos, que pueden ser una fuente de error,

debido a que los encuestados pueden interpretar las cuestiones de manera diferente a como han sido diseñadas.

Especificación del error de medida

Junto con el tipo de error debido a la medición de conceptos abstractos, podemos encontrarnos con posibles errores de medida, debido a respuestas incorrectas, mediciones mal efectuadas, etc. SEM proporciona lo que se conoce como “modelo de medida”, que especifica las reglas de correspondencia entre las variables latentes y manifiestas. Este modelo de medida, nos va a permitir utilizar una o más variables para un único concepto dependiente o independiente y, estimar su fiabilidad. En el modelo de medida el investigador puede evaluar la contribución de cada ítem de la escala e incorporar la fiabilidad con la que se está midiendo el concepto, en la estimación de las relaciones entre variables dependientes e independientes.

5.2.2.1 Metodología

La metodología aplicada en el CFA es la que se muestra en la figura 5.6, detallándose cada uno de los pasos a continuación.

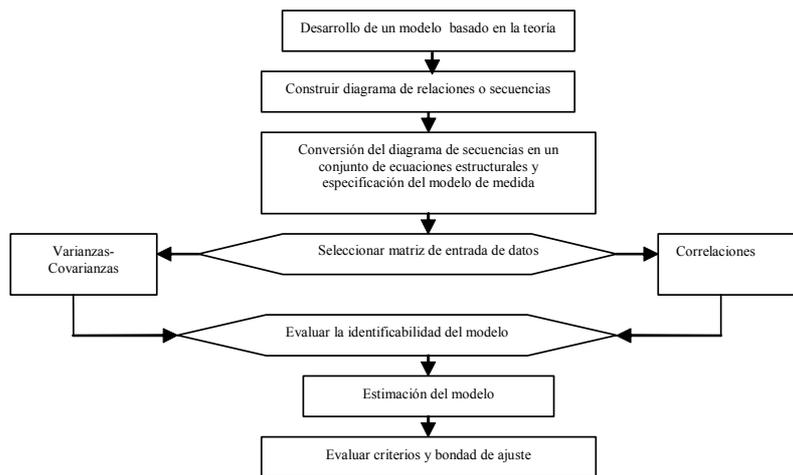


Figura 5.6. Etapas del CFA

El software utilizado en este trabajo es el AMOS 4.0 para SPSS (Arbuckle, 1994b; Wothke, 1993), programa diseñado para realizar análisis de ecuaciones estructurales.

a) Especificar un modelo basado en la teoría

La modelización de ecuaciones estructurales se basa en relaciones causales, en las que el cambio de una variable produce un cambio en otra variable. En consecuencia, el primer paso de esta etapa de especificación, es traducir en un conjunto de ecuaciones las teorías verbales formuladas previamente y que pueden afectar a: variables latentes o dimensiones que deben incluirse; los efectos entre las variables latentes y su tipo, que vienen definidos en el anexo 2; los indicadores asignados a cada dimensión y las covarianzas entre variables latentes exógenas.

La especificación del modelo debe completarse con el cumplimiento de los siguientes supuestos:

- Supuestos estadísticos sobre las fuentes de variación y, en concreto, sobre la forma de su función de distribución conjunta, que se requieren para la operatividad del modelo. Las técnicas de análisis de estructuras de covarianza y, más en concreto el análisis factorial confirmatorio, demandan como punto de arranque una distribución normal multivariada, o por lo menos, que en la práctica todas y cada una de las variables sobre las que se aplican, muestren una distribución continua y normal, así como que resulte satisfecho el supuesto de homocedasticidad. Por otra parte, Muthén y Kaplan (1995), entre otros, proponen que siempre que se utilicen variables ordinales, se puede esperar una distorsión relativamente pequeña en las estimaciones de Máxima Verosimilitud de los parámetros de un modelo de ecuaciones estructurales (SEM), así como en sus errores típicos y en los índices de bondad de ajuste (Mateo y Fernández Sánchez, 1992a y 1992b).
- Supuestos sobre el comportamiento de las variables no consideradas, cuyo efecto se recoge en los términos de error de medida o perturbación. En este

sentido, resulta esencial no omitir variables predictivas claves, que darían lugar a lo que se denomina *errores de especificación*, y cuya consecuencia consiste en sesgar la evaluación de la importancia de otras variables.

- La correlación entre determinados pares de términos de perturbación o términos de error de medida.

b) Construcción de un diagrama de secuencias de relaciones causales

Antes de representar las relaciones causales en términos de ecuaciones, podemos realizarlo también mediante los diagramas causales o *path diagrams* (Duncan, 1975). Estos diagramas no son más que una simple representación visual de las relaciones, tanto relaciones predictivas entre constructos como relaciones asociativas (correlaciones) entre constructos e, incluso, entre indicadores. La técnica utiliza como herramienta unos grafos, que reflejan el proceso causal atendiendo a ciertas convenciones que los hacen acordes con las ecuaciones:

- La relación causal entre variables se indica por una flecha cuyo sentido es desde la variable causa a la variable efecto. Cada flecha está afectada por un parámetro o coeficiente que indica el efecto entre ambas variables. Cuando entre dos variables no hay flecha, se está indicando que el efecto es cero. Las variables a las que llega alguna flecha se denominan endógenas y aquellas a las que no llega ninguna flecha se denominan exógenas.
- La covariación entre dos variables exógenas o dos términos de perturbación se representa por una flecha bidireccional que las une en ambos sentidos. El parámetro asociado a esta flecha es la covarianza.
- Las variables observables se representan por cuadrados y las variables latentes por círculos u óvalos.

- Las relaciones causales y covarianzas también se aplican a la relación entre variables latentes.
- Las denominadas “relaciones de medición” son las que conectan las variables observables con las latentes mediante flechas unidireccionales con origen en las latentes.
- Las variables observables están afectadas por un término de error aleatorio de medición (e_i), que también puede representarse en el grafo mediante una flecha adicional apuntando a dichas variables. La covariación entre dos errores de medición se representa por una flecha bidireccional que los une en ambos sentidos.

En la figura 5.7, se representa un path diagram de un modelo de ecuaciones simultáneas con más de una variable dependiente (X_2 y X_3). Además, la variable X_2 es dependiente y explicativa, el modelo contiene dos efectos directos (β_{32} y β_{21}) y uno indirecto de X_1 sobre X_3 , cuyo valor sería igual al producto de $\beta_{32} \beta_{12}$.

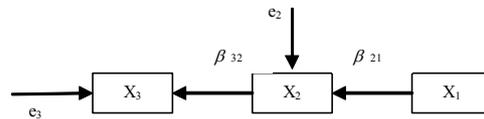


Figura 5.7. Path diagram de un modelo con tres variables y efectos indirectos

En la figura 5.8, se muestra un path diagram donde la variable X_1 se encuentra medida con error (e_1). El valor libre de error de la variable explicativa no es directamente observable, sino latente y se representa por f_1 . Representaremos, siguiendo esta notación, el valor libre de la variable X_2 como f_2 , aunque esta variable no contiene término de error.

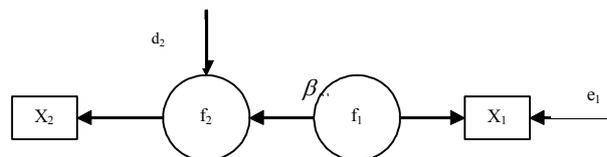


Figura 5.8. Path diagram de un modelo lineal con errores en la variable explicativa

En el path diagram, las variables latentes aparecen dentro de círculos, con una flecha apuntando hacia el indicador o variable observable, a su vez afectada por el error de medición correspondiente, en su caso.

Para representar las teorías causales y de medición en el path diagram de forma equivalente a como lo hacen los sistemas de ecuaciones se debe cumplir una serie de hipótesis básicas:

- Todas las relaciones causales son lineales.
- Todas las relaciones causales están incluidas en el diagrama. Es tan importante justificar porqué no existe una relación causal entre dos constructos como justificar la existencia de otra relación.
- El diagrama es lo más sencillo posible y contiene sólo aquellas relaciones que pueden justificarse con bases teóricas.

c) Conversión del diagrama de secuencias en un conjunto de ecuaciones estructurales y especificación del modelo de medida.

Tras representar el diagrama de secuencias, es necesario especificar el modelo representado en el diagrama en términos más formales. Para ello debe establecerse un sistema de ecuaciones estructurales que expresan cada elemento de la matriz en función de los parámetros del modelo. Este sistema suele representarse como $\Sigma = \Sigma(\vec{\pi})$, donde Σ representa la matriz de varianzas y covarianzas de la población, $\vec{\pi}$ un vector que contiene a todos los parámetros (efectos directos, varianzas y covarianzas de errores, de perturbaciones y de variables exógenas) que imponen una forma o estructura determinada a Σ y $\Sigma(\vec{\pi})$ es el sistema estructural especificado para la población.

Reglas de descomposición del path analysis

Para conseguir el sistema de ecuaciones estructurales que expresan cada elemento de la matriz de covarianzas en función de los parámetros del modelo, es necesario utilizar las reglas de descomposición del path análisis. Estas reglas establecen de forma intuitiva la relación entre los parámetros y las covariaciones entre las variables, lo que nos permitirá, en último término, estimar los parámetros a partir de las relaciones entre variables. La práctica inicial del path analysis consideraba únicamente las variables estandarizadas, los parámetros estandarizados y tomaba el coeficiente de correlación como medida de relación entre variables. Sin embargo, como los modelos de ecuaciones estructurales acostumbran a ajustarse sobre matrices de varianzas-covarianzas, actualmente también es frecuente considerar las variables sólo centradas con respecto a su media y trabajar con parámetros no estandarizados.

Las varianzas y covarianzas que se refieren únicamente a variables exógenas constituyen por sí mismas parámetros del modelo. Las reglas de descomposición para derivar las restantes varianzas y covarianzas son las siguientes:

- La covarianza entre dos variables es igual a la suma del efecto directo, los efectos indirectos, espúreos y conjuntos. Cada uno de estos efectos representa una posible manera de unir ambas variables sobre el path diagram, partiendo de un origen cualquiera y siguiendo las flechas. El origen puede ser una de las dos variables, una tercera variable o una covarianza entre dos variables exógenas. El efecto se calcula como el producto de la varianza de la variable de partida (o covarianza de partida) por todos los parámetros asociados a las flechas recorridas hasta llegar a unir las dos variables de interés. Al calcular un efecto no se puede pasar por una variable más de una vez.
- La varianza de una variable dependiente es igual a la varianza del término de permutación más la varianza explicada por las otras variables del modelo. Esta varianza explicada puede, a su vez, expresarse en función de todas las variables explicativas con efecto directo sobre la dependiente, como suma de todos los

productos entre estos efectos directos y las covarianzas entre la variable dependiente y la variable explicativa, relacionadas por dichos efectos.

Tipos de parámetros

Adicionalmente, en los modelos de ecuaciones estructurales, la especificación puede dar lugar a tres tipos de parámetros:

- *Libres*, si son parámetros desconocidos y no restringidos para ser estimados.
- *Restringidos*, si son parámetros desconocidos que deben tomar el mismo valor al estimarse.
- *Fijos*, si son conocidos por haberseles asignado previamente un valor dado. A este tipo de parámetros se les puede asignar cualquier valor, lo más frecuente es fijarlos a cero. Sin embargo, también podemos encontrarnos con el caso contrario cuando se restringe a la unidad una saturación para fijar la escala de una variable latente.

Tipos de modelos de ecuaciones estructurales

El modelo de ecuaciones estructurales consta de distintos sistemas de ecuaciones: el conjunto de ecuaciones estructurales que vincula a los constructos, el modelo de medida que especifica qué variables miden cada constructo y una serie de matrices que indiquen cualquier correlación supuesta entre constructos o variables.

- *Modelo Estructural. Relaciones entre factores o constructos*

En el modelo estructural, cada constructo endógeno es la variable dependiente en una ecuación separada. Las variables predictoras son todos los constructos o “colas” de las flechas que apuntan a la variable endógena. Para cada efecto supuesto se calcula un

coeficiente estructural y, para cada ecuación se debe incluir un término de error (e_i) que representa la suma de los efectos debidos a la especificación y al error aleatorio de medida.

Así, el sistema de ecuaciones del modelo estructural correspondientes al path diagram representado en la figura 5.7 quedaría:

$$\begin{aligned} X_2 &= \beta_{21}X_1 + e_2 \\ X_3 &= \beta_{32}X_2 + e_3 \end{aligned}$$

Además, deben verificarse una serie de supuestos estadísticos sobre el comportamiento de las fuentes de variación de las variables incluidas en el modelo. Estas fuentes de variación incluyen todos los términos de perturbación de las variables y todas las variables a las que no llega ninguna flecha unidireccional en el path diagram. En nuestro ejemplo, serían e_2 , e_3 y X_1 . Estos supuestos pueden resumirse en la ecuación:

$$\begin{pmatrix} X_1 \\ e_2 \\ e_3 \end{pmatrix} \rightarrow N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix} \right)$$

donde ϕ_{11} es la varianza de X_1 y ψ_{22} , ψ_{33} la varianza de e_2 y e_3 respectivamente, que ponen de manifiesto, por un lado, la distribución normal multivariante de X_1 , e_2 y e_3 , que queda determinada por un vector de esperanzas cuyo valor es cero, reflejando el hecho de que las variables representan desviaciones respecto a su media y, una matriz de varianzas y covarianzas en la que las primeras se disponen en la diagonal principal y las segundas fuera de ella. Por otro lado pone de manifiesto la incorrelación entre términos de error y entre términos de error y variables explicativas, por el que la varianza de X_2 puede analizarse mediante la varianza explicada por X_1 y la varianza no explicada o varianza de perturbación de e_2 . Esta descomposición suele expresarse en términos porcentuales como coeficiente de determinación (R^2), es decir, como el porcentaje de varianza de X_2 explicada por X_1 . Finalmente, pone de manifiesto la distribución idéntica para todas las observaciones, que constituyen realizaciones independientes de la misma, es decir, la homocedasticidad.

Tras haber especificado el modelo, se pueden aplicar las reglas de descomposición del path analysis para derivar el sistema de ecuaciones estructurales $\Sigma = \Sigma \bar{\pi}$, que relacionan las varianzas y covarianzas entre las variables con los parámetros del modelo.

La matriz de covarianzas viene expresada como:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

donde σ_{ii} representa la varianza poblacional de X_i y σ_{ij} la covarianza poblacional entre X_i y X_j . La matriz es simétrica y contiene 6 elementos no duplicados.

El vector de parámetros tiene 5 elementos y podemos expresarlo como $\bar{\pi} (\phi_{11}, \psi_{22}, \psi_{33}, \beta_{21}, \beta_{32})$. La diferencia entre ambos valores es lo que se denomina *grados de libertad del modelo*, que comentaremos más adelante.

Aplicando las reglas de descomposición podemos escribir el sistema de ecuaciones estructurales como:

$$\begin{cases} \sigma_{11} = \phi_{11} \\ \sigma_{21} = \phi_{11}\beta_{21} \\ \sigma_{22} = \psi_{22} + \sigma_{21}\beta_{21} \\ \sigma_{31} = \phi_{11}\beta_{21}\beta_{32} \\ \sigma_{32} = \sigma_{22}\beta_{32} \\ \sigma_{33} = \psi_{33} + \sigma_{32} + \beta_{32} \end{cases}$$

Es decir, la varianza de X_1 constituye directamente un parámetro del modelo. La covarianza entre X_1 y X_2 se calcula a partir de la primera regla de descomposición, comentada anteriormente. La única manera de unir X_1 con X_2 es a partir del efecto directo, que se obtiene multiplicando la varianza de la variable X_1 por el coeficiente que relaciona ambas variables. La varianza de X_2 se calcula a partir de la segunda regla de descomposición, como varianza de la perturbación más el producto entre el efecto directo de X_1 sobre X_2 y la covarianza entre ambas variables. La covarianza entre X_3 y X_1 se calcula mediante la primera regla de descomposición; ambas variables sólo pueden unirse

mediante un efecto indirecto que se estima como producto de la varianza de X_1 por los parámetros asociados a cada una de las flechas que deben recorrerse. La covarianza entre X_3 y X_2 se calcula a partir de la varianza de X_2 multiplicada por el único efecto directo que une ambas variables. La varianza de X_3 se calcula nuevamente a partir de la segunda regla de descomposición como varianza de la perturbación más el producto entre el efecto directo de X_2 sobre X_3 y la covarianza entre ambas variables.

- *Modelo de Medida. Relaciones entre factores e indicadores*

El modelo de medida consiste únicamente en especificar qué variables definen cada constructo o factor. Las variables observadas se denominan en este modelo de medida indicadores, ya que se utilizan para medir o indicar los constructos latentes.

En el caso del path diagram representado en la figura 5.8, con errores de medición en la variable observable, son necesarias especificar tanto las ecuaciones del modelo estructural como las del modelo de medida. Las ecuaciones que relacionan los factores entre sí y, que se interpretan del mismo modo que en el caso anterior, quedarían:

$$f_2 = \beta_{21}f_1 + d_2$$

Las ecuaciones que relacionan los indicadores con sus factores respectivos en este ejemplo, quedarían:

$$\begin{aligned} X_1 &= f_1 + e_1 \\ X_2 &= f_2 \end{aligned}$$

habiendo considerado que los coeficientes que relacionan indicadores y factores son igual a la unidad.

En este caso, la especificación del modelo se completa con el supuesto de incorrelación entre los errores de medición y los factores:

$$\begin{pmatrix} f_1 \\ e_1 \\ d_2 \end{pmatrix} \rightarrow N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \varphi_{11} & 0 & 0 \\ 0 & \theta_{11} & 0 \\ 0 & 0 & \psi_{22} \end{pmatrix} \right)$$

siendo θ_{11} la varianza de e_1 , ϕ_{11} la varianza de X_1 y ψ_{22} la varianza de e_2 .

Mediante estos supuestos, se puede descomponer la varianza de las variables observables entre varianza explicada por los factores y varianza del error de medición. Esta descomposición, puede expresarse en forma de coeficiente de determinación, denominándose calidad de medición y representándose por k_1 :

$$k_1 = 1 - \frac{\theta_{11}}{\sigma_{11}}$$

Interesa un valor del mismo lo más elevado posible.

d) Selección del tipo de matriz de entrada de datos y estimación del modelo

Los modelos de ecuaciones estructurales sólo emplean la matriz de varianzas-covarianzas o la matriz de correlaciones de todos los indicadores utilizados en el modelo como datos de entrada. A partir de ahí, el modelo de medida especifica qué indicadores corresponden a cada constructo y las puntuaciones de los constructos latentes o factores son las empleadas en el modelo estructural.

La interpretación de los resultados obtenidos varía en función de que se utilice un tipo de matriz u otra. Las primeras presentan la ventaja de proporcionar comparaciones válidas entre distintas poblaciones o muestras, característica que no es posible cuando los modelos se estiman con matrices de correlación. Sin embargo, la interpretación de los resultados es más compleja cuando se utilizan covarianzas, dado que los coeficientes deben interpretarse en términos de unidades de medida de los constructos. Además las matrices de correlación tienen un rango común que hace posibles las comparaciones directas de los coeficientes dentro de un modelo, ya que la matriz de correlaciones no es más que la matriz de varianza-covarianza estandarizada en la que la escala de unidad de medida de cada variable se modifica dividiendo las varianzas-covarianzas por el producto de las desviaciones estándar.

Utilizar las correlaciones es adecuado cuando el motivo del investigador es sólo entender el patrón de relaciones entre constructos, pero no cuando se pretende explicar la

varianza total del constructo, o bien, hacer comparaciones entre las diferentes variables, dado que la escala de medida afecta a las covarianzas. Los coeficientes obtenidos de la matriz de correlaciones están siempre en unidades estandarizadas, similares a las ponderaciones beta de la regresión y con un rango entre -1 y 1.

e) Valoración de la identificación del modelo

Consiste en analizar si se posee la suficiente información para proceder a la estimación unívoca de los parámetros del modelo, es decir, comprobar si se pueden estimar los parámetros del modelo a partir de las varianzas y covarianzas de las variables. Se basa en el principio de que se debe tener una ecuación aislada y diferenciada para estimar cada coeficiente.

Las posibles causas de no identificación son: mayor número de coeficientes a estimar que de varianzas-covarianzas o correlaciones, uso de efectos recíprocos y fallo en la fijación de la escala de un constructo.

El estudio de la identificabilidad es complejo, ya que no existe un conjunto simple de condiciones necesarias y suficientes que nos permitan evaluarla.

La condición necesaria indica que los grados de libertad del modelo deben ser mayores o iguales que cero. Los grados de libertad son la diferencia entre el número de correlaciones o covarianzas y el número efectivo de coeficientes en el modelo propuesto y se calcula como:

$$g = 1/2 + [(p + q) * (p + q + 1)] - t$$

donde p es el número de indicadores endógenos, q es el número de indicadores exógenos y t es el número de coeficientes estimados en el modelo propuesto.

En función de las características de los modelos estudiados existen una serie de condiciones suficientes, que se comentan a continuación.

En caso de que el modelo sólo relacione variables observables medidas sin error, se establecen las siguientes condiciones suficientes (Batista Foguet y Coenders, 2000):

- Los modelos de regresión lineal que cumplan la incorrelación entre el término de perturbación y todas las variables explicativas son siempre identificables.
- Se denominan modelos recursivos a aquellos en los que es posible establecer una ordenación de las variables de forma que cada variable afecte solamente a aquellas que tiene a continuación, es decir, el path diagram se puede dibujar con flechas en la misma dirección, como se muestra en la figura 5.9.

Los modelos recursivos que cumplan simultáneamente la incorrelación entre los términos de perturbación y las variables explicativas de cada ecuación y la incorrelación entre todos los términos de perturbación son siempre identificables, como sería el caso de la figura 5.11. Un ejemplo de modelo no identificable sería el que se muestra en la figura 5.9.

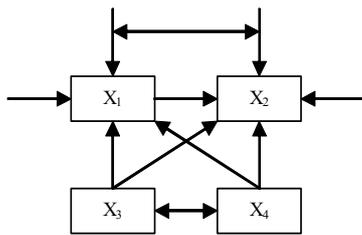


Figura 5.9. Modelo recursivo

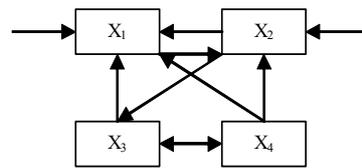


Figura 5.10. Modelo no recursivo

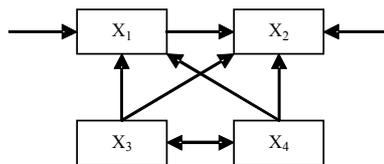


Figura 5.11. Modelo recursivo identificable

- Los modelos recursivos que no contengan ningún efecto directo que relacione las variables endógenas entre sí son identificables aunque las perturbaciones de las distintas variables endógenas estén relacionadas. Un ejemplo se muestra en la figura 5.12.

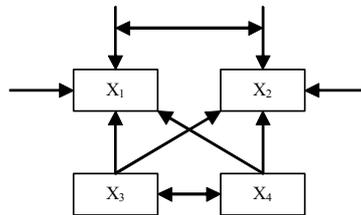


Figura 5.12. Modelo recursivo identificable

- Los modelos no recursivos tienen mayores dificultades respecto a su identificabilidad; utilizan restricciones en los coeficientes, que se reflejan en lo que se denomina condiciones de orden y rango (Goldberger, 1964; Jonhston, 1972, Kmenta, 1971), referentes a los grado de libertad del modelo y determinan si cada parámetro se estima especialmente. El caso más sencillo es aquel en el que dos variables X_1 y X_2 presentan una relación recíproca. En este caso, es necesario que la ecuación de X_1 contenga por lo menos una variable explicativa que no esté incluida en la ecuación de X_2 y que la ecuación de X_2 contenga al menos una variable explicativa que no esté incluida en la ecuación de X_1 . Un ejemplo de modelo no recursivo identificado se muestra en la figura 5.13 y uno de no recursivo no identificado en la 5.14.

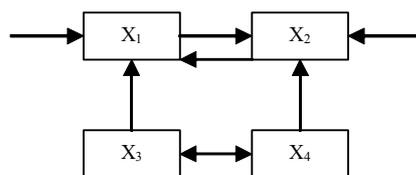


Figura 5.13. Modelo no recursivo identificado

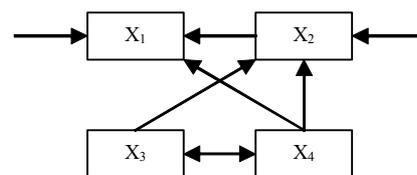


Figura 5.14. Modelo no recursivo no identificado

Cuando el modelo incluye variables latentes, las tres condiciones siguientes son suficientes para garantizar la identificación en modelos con dos o más factores si se cumplen simultáneamente:

- Las relaciones de las variables latentes entre sí están identificadas según las reglas vistas para variables medidas sin error.
- Cada variable latente tiene al menos dos indicadores puros, es decir, que relacionan una única variable latente y sus errores de medida están incorrelacionados con los de cualquier otro indicador.
- Cada variable latente tiene al menos un indicador cuya saturación está restringida a un valor distinto de cero (generalmente la unidad), siendo esto necesario para fijar la escala de la variable latente, de forma que resulte fácil y cómodo de interpretar, lo que implicaría que la variable latente está medida en las mismas unidades que la observable correspondiente.

f) Estimación del modelo

Esta etapa se basa también en la relación entre las varianzas y las covarianzas y los parámetros del modelo. Se trata de estimar el valor de $\vec{\pi}$ a partir de aquellas ecuaciones estructurales. Como la matriz de varianzas-covarianzas poblacionales, Σ , no es conocida, se deberá aproximar mediante la matriz de varianzas y covarianzas muestral, S. Es decir, se trata de estimar aquellos valores, p, de los parámetros $\vec{\pi}$, que ajusten lo mejor posible la matriz S por la que ellos reproducen $\sum(p)$.

La expresión que se minimiza en esta etapa de estimación es del tipo:

$$F = (S - \sum(p))' W (S - \sum(p))$$

en la que $(S - \sum(p))'$ es el vector fila de los residuos, $(S - \sum(p))$ es el vector columna y, W es la llamada matriz de pesos. Como la matriz de los residuos es simétrica, se considera que el vector de los residuos contiene únicamente sus elementos no duplicados.

Las estimaciones obtenidas minimizando esta matriz son consistentes, es decir, convergen a los valores poblacionales de los parámetros para muestras infinitamente grandes para cualquier matriz de W. En el caso más sencillo, W=I, la expresión se reduce a la suma de cuadrados de los residuos. Sin embargo suelen utilizarse matrices W más

complejas, ya que la estimación es asintóticamente eficiente cuando $W = \Gamma^{-1}$, donde Γ es la matriz de varianzas y covarianzas en el muestreo de los elementos de S . Esta matriz relativiza los tamaños de los residuos según la escala de variables observables y la precisión con la que el residuo está medido.

Bajo la hipótesis de normalidad multivariante, $\Gamma = \Sigma \otimes \Sigma$, donde \otimes representa el producto de Kroneker (Bollen y Long, 1993) y Σ la matriz de covarianzas poblacional, que se estima a partir de S o $\Sigma(p)$.

En caso de que no se cumpla la normalidad multivariante $\Gamma = H$, siendo H una función de los momentos de cuarto orden de las variables observables.

Existen distintos procedimientos de estimación del modelo:

- Mínimos cuadrados no ponderados (LS), con $W = I$.
- Mínimos cuadrados ponderados bajo normalidad (NT-WLS), con $W = (S \otimes S)^{-1}$, que lo hace asintóticamente eficiente bajo el supuesto de normalidad multivariante.
- Máxima verosimilitud (MV), con $W = (S(p) \otimes S(p))^{-1}$, que lo hace asintóticamente eficiente bajo el supuesto de normalidad multivariante.
- El método asintóticamente libre de distribución (ADF), con $W = H^{-1}$, que lo hace asintóticamente eficiente para cualquier distribución de las variables observables.

De todos ellos, posiblemente NT-WLS y MV sean los métodos más utilizados en la actualidad, en especial el segundo, debido a su precoz introducción de las restricciones en el proceso de estimación (Jöreskog, 1969) y a que permite aplicar toda la inferencia desarrollada para los estimadores máximo verosímiles, como por ejemplo contrastes de la razón de verosimilitudes y análogos.

En caso de trabajar con la matriz de correlaciones de Pearson en lugar de con la matriz de covarianzas, las estimaciones obtenidas son sólo correctas si se cumple simultáneamente (Cudeck, 1989):

- Las únicas restricciones sobre los parámetros del modelo son las necesarias para fijar la escala de las variables latentes o consisten en fijar parámetros al valor de 0.
- La matriz de residuos contiene ceros en la diagonal principal.

g) Evaluación de los Criterios de Ajuste

Consiste en evaluar la adecuación del modelo especificado, es decir si es correcto y útil para los objetivos especificados. Por ser una de las etapas más importantes ha sido estudiada por numerosos autores (Batista-Foguet y Coenders, 2000; Jöreskog y Sörbom, 1989; etc.).

Un modelo correcto debe incorporar aquellas restricciones y supuestos implícitos que se cumplen en la población y, por lo tanto, especificar correctamente las relaciones entre las variables sin omisión de parámetros. Es decir, un modelo correcto debe predecir adecuadamente la realidad, conduciendo a diferencias reducidas y aleatorias entre las varianzas y covarianzas observadas y las implicadas por el modelo. Esta etapa de diagnóstico nos permitirá distinguir entre modelos que ajusten mal los datos y aquellos que los ajusten razonablemente bien, aunque no nos permitirá fácilmente discernir entre modelos equivalentes que ajusten aproximadamente igual de bien los datos.

Para juzgar la calidad de ajuste entre el modelo y los datos pueden utilizarse distintos indicadores:

- Información sobre las variables y matriz de correlaciones.
- Estimación inicial del modelo, ya que el programa se ejecuta de forma iterativa aproximándose progresivamente a la mejor solución.
- Parámetros estimados por el modelo elegido.

- Coeficientes de determinación para las variables endógenas y exógenas.
- Medidas de bondad de ajuste del modelo completo.
- Información sobre los residuos normalizados.
- Covarianzas.
- Solución estandarizada.

Los requisitos críticos que debe cumplir un buen modelo se detallan a continuación.

- *Estimaciones Infractoras*

Los resultados obtenidos en cada modelo, deben analizarse buscando estimaciones aberrantes o infractoras, es decir, coeficientes estimados tanto en los modelo de medida como en los estructurales que exceden los límites aceptables. Los casos más típicos de estimaciones infractoras son: varianzas de error negativas o no significativas para cualquier constructo; coeficientes estandarizados que sobrepasan o están muy próximos a uno; errores estándar muy elevados asociados con cualquier coeficiente estimado.

- *Ajuste Global del Modelo*

La calidad del ajuste mide la correspondencia entre la matriz de entrada real u observada con la que predice el modelo. En este sentido, interesa tener un elevado número de grados de libertad, ya que de esta manera, el modelo consigue parsimonia, es decir, se obtiene el mejor ajuste posible del modelo para cada coeficiente estimado. Cuanto mejor ajuste pueda conseguirse con pocos coeficientes, mejor será el contraste del modelo y más confianza podemos tener en los resultados.

Las medidas de calidad de ajuste global del modelo son de tres tipos: medidas absolutas del ajuste, medidas del ajuste incremental y medidas del ajuste parsimonia. A continuación se describen brevemente cada una de ellas.

- **Medidas absolutas del ajuste.**

Evalúan el ajuste global del modelo (tanto el del modelo de medida como el del modelo estructural), es decir, el grado en que el modelo conjunto predice la matriz de correlación o covarianza observada. Las medidas de ajuste absoluto utilizadas en este trabajo son: el estadístico de verosimilitud chi-cuadrado, el índice de bondad de ajuste, el residuo cuadrático medio y el error de aproximación cuadrático medio.

- *Estadístico de verosimilitud chi-cuadrado.* Todo método de estimación conduce, como se ha comentado a los p valores de los parámetros minimizando una determinada función de ajuste que se considera la discrepancia entre S y $\Sigma(p)$. Valores reducidos representan menor discrepancia entre lo observado y lo predecido por el modelo y, por tanto, un mejor ajuste. Existe una transformación del mínimo de la función de ajuste denominada estadístico χ^2 de bondad de ajuste, que si se cumplen las hipótesis de que los datos siguen una distribución normal multivariante, se emplea el estimador de Máxima Verosimilitud (MV), el modelo es correcto y la muestra suficientemente grande, dicho estadístico sigue una distribución χ^2 con los mismos grados de libertad que el modelo. Dicho estadístico permite contrastar la hipótesis nula de que el modelo es correcto respecto al grado de especificación, es decir, que el modelo no presenta especificaciones incorrectas o no existe necesidad de introducir parámetros adicionales. Este enfoque es completamente contrario al empleado en otras técnicas, como por ejemplo regresión, donde mediante la prueba F de significación del modelo se valora la hipótesis nula de ausencia de relación entre la variable dependiente y el conjunto de variables explicativas. Este diferente enfoque de la prueba χ^2 implica que la potencia de la prueba deba considerarse como la probabilidad de detectar errores

de especificación, es decir, restricciones incorrectas o necesidad de introducir parámetros adicionales. Un valor elevado del estadístico χ^2 con respecto a los grados de libertad significa que las matrices observadas y estimadas difieren considerablemente. Los niveles de significación estadística indican la probabilidad de que estas diferencias se deban solamente a variaciones de la muestra.

Debemos considerar, que como los modelos de ecuaciones estructurales suelen desarrollarse sobre muestras relativamente grandes, la potencia es a menudo alta. Además, como la finalidad de SEM suele ser más bien obtener modelos aproximados que obtener modelos exactamente correctos, a menudo suele resultar más relevante evaluar la importancia de los errores de especificación que la significación estadística, lo que implica que el estadístico χ^2 sea considerado más como una medida de bondad de ajuste que como una prueba de significación. Por ello, generalmente el diagnóstico del modelo suele basarse muchas veces en otras medidas descriptivas de bondad de ajuste, como por ejemplo, medidas sintéticas del tamaño de los residuos, o incluso transformaciones del propio estadístico χ^2 , que se detallan a continuación.

- *Índice de Bondad de Ajuste (GFI: Godness of Fit Index)*. Es una medida no estadística que fué propuesta para modelos ajustados con Máxima Verosimilitud (Joreskog y Sorbön, 1981 y 1986). Representa el grado de ajuste conjunto (residuos al cuadrado de la predicción comparados con los datos efectivos), pero no está ajustada por los grados de libertad. Puede tomar valores entre 0 (mal ajuste) y 1 (ajuste perfecto).
- *Residuo cuadrático medio (RMSR: Root Mean Square Residual)*. Considera el tamaño global de los residuos de la matriz $S-\Sigma(p)$, pero trabaja con residuos normalizados ya que por lo general son más sencillos de interpretar. Los residuos estandarizados se definen como la discrepancia entre las correlaciones observadas y las predichas por el modelo. Es la raíz cuadrada de la media de los residuos al cuadrado. Su utilización fue propuesta sobre la base del cálculo de los residuos normalizados (Joreskog y Sorbön, 1981 y 1986). Es más útil cuando se utiliza

como matriz de entrada de datos la matriz de correlaciones, ya que están todas en la misma escala. Su valor ideal es próximo a 0.

- *Error de Aproximación Cuadrático Medio (RMSEA: Approximation Root Mean Square Residual)*. Intenta corregir la tendencia del estadístico chi-cuadrado para rechazar cualquier modelo especificado con una muestra suficientemente grande. Se puede interpretar, al igual que el índice anterior, como la discrepancia media por grado de libertad, pero ahora esta discrepancia se mide en términos de población, en lugar de sólo con la muestra utilizada. Se define como (Brown y Cudeck, 1993):

$$RMSEA = \sqrt{\frac{\chi^2 - g}{N * g}}$$

donde, χ^2 es el estadístico chi-cuadrado y g son los grados de libertad. Suelen considerarse aceptables valores entre 0,05 y 0,08.

- **Medidas del ajuste incremental.**

Comparan el modelo propuesto con otro de referencia, conocido como modelo nulo, que es un modelo de constructo único, con todos los indicadores midiendo perfectamente a ese constructo. Los estadísticos más utilizados son: el índice ajustado de bondad de ajuste, el índice de ajuste normal y el índice de Turker-Lewis.

- *Índice Ajustado de Bondad de Ajuste (AGFI: Adjusted Goodness Fit Index)*. Fue propuesto para modelos ajustados con máxima verosimilitud (Joreskog y Sorbön, 1986 y 1989). Es una extensión del GFI ajustado por los grados de libertad del modelo propuesto y del modelo nulo. Puede tomar valores entre 0 (mal ajuste) y 1 (ajuste perfecto). Es un índice de ajuste y parsimonia del modelo, penalizando los modelos con muchos parámetros.
- *Índice de Ajuste Normado (NFI: Normal Fit Index)* de Bentler y Bonet (1980). Evalúa la disminución del estadístico χ^2 de nuestro modelo con respecto al modelo base. Se expresa como:

$$NFI = \frac{\chi_B^2 - \chi^2}{\chi_B^2}$$

donde χ^2 es el estadístico chi- cuadrado del modelo en cuestión y χ_B^2 el estadístico chi-cuadrado del modelo base. El valor recomendado es de 0,90 o superior (Hair et al., 1995).

Este índice no suele ser recomendable, ya que al no considerar los grados de libertad del modelo, aumenta siempre al añadir parámetros al modelo, favoreciendo la adopción de modelos sobreparametrizados.

- *Índice de Ajuste No Normado (NNFI:No Normal Fit Index)* de Turker-Lewis (1973). No introduce directamente los estadísticos χ^2 , sino que los compara previamente con su esperanza, los grados de libertad del modelo base (g_B) y del modelo en cuestión (g). Este índice tiene en cuenta la parquedad del modelo, es decir, si se añaden parámetros al modelo los índices sólo aumentan si el estadístico χ^2 disminuye en mayor medida que los grados de libertad.

$$NNFI = \frac{(\chi_B^2 / g_B) - (\chi^2 / g)}{(\chi_B^2 / g_B) - 1}$$

Los valores recomendados son superiores a 0,90 (Hair et al., 1995).

- **Medidas del ajuste de parsimonia**

Permiten comparar entre modelos con distinto número de coeficientes para determinar la cantidad de ajuste conseguido por cada coeficiente. Es decir, relacionan la calidad de ajuste del modelo con el número de coeficientes estimados exigidos para conseguir este nivel de ajuste. Pretende diagnosticar si el ajuste del modelo se ha conseguido debido a un sobreajuste de los datos por tener demasiados coeficientes. Es un procedimiento similar al ajuste del R^2 en regresión múltiple. El más utilizado es el índice de ajuste normado parsimonia.

- *Índice de ajuste normado parsimonia (PNFI: Parsimonia Godness Fit Index)*. Modifica al NFI, considerando los grados de libertad del modelo utilizados para conseguir un buen nivel de ajuste. La parsimonia se define en la consecución de mayores niveles de ajuste por grado de libertad utilizado, es decir un grado de libertad por coeficiente estimado. Por tanto es deseable una mayor parsimonia. Se define como:

$$PNFI = \frac{gl_{propuesto}}{gl_{nulo}} * NFI$$

Puede tomar valores entre 0 y 1, cuanto más próximos a 1, mayor parsimonia del modelo (Hair et al., 1995).

- *Índice de calidad de ajuste parsimonia (PGFI)*, que modifica al GFI pero de distinta forma que el AGFI, basándose en la parsimonia del modelo estimado.

$$PGFI = \frac{gl_{propuesto}}{1/2(N^{\circ} \text{ variables manifiestas})(N^{\circ} \text{ variables manifiestas} + 1)} * GFI$$

Puede tomar valores entre 0 y 1, siendo los valores elevados los que indican mayor parsimonia.

h) Interpretación de los resultados

Una vez que el modelo se considera aceptable es necesario examinar los resultados y su correspondencia con la teoría propuesta. A la hora de interpretar los resultados, podemos analizar la solución estandarizada o la no estandarizada. En la primera, los coeficientes estandarizados tienen todos igual varianza, un valor máximo de 1, como en las ponderaciones beta en regresión. Los coeficientes cercanos a cero tienen poco efecto sustantivo y un aumento en el valor corresponde a un aumento de la importancia en las relaciones causales. Los coeficientes estandarizados son útiles en la determinación de la

importancia relativa, pero son específicos de una muestra y no son comparables entre muestras.

Los coeficientes sin estandarizar corresponden a las ponderaciones de regresión en una regresión múltiple en la que se expresan en términos de la escala del constructo, en este caso su varianza. Esto hace comparables los coeficientes para las muestras y retiene los efectos escala. Sin embargo, la comparación entre coeficientes es más difícil que con los coeficientes estandarizados.

Finalmente, una vez que se ha completado la interpretación del modelo, puede interesar mejorar el ajuste del modelo y su correspondencia con la teoría subyacente, que es lo que se denomina “*reespecificación del modelo*” y que consiste en añadir o quitar parámetros del modelo original, de manera que se obtengan un conjunto de modelos rivales para ser comparados. Esta práctica, sin embargo, no suele ser aconsejable sino existe una justificación teórica para lo que se considera empíricamente aceptable.

5.3. Resultados y Discusión

A continuación se muestran los resultados obtenidos mediante la aplicación del Análisis Factorial, tanto con fines exploratorios como con fines confirmatorios, comentándose la relación existente entre las variables o ítems del cuestionario y que constituyen lo que denominamos factores o componentes, identificándose la variable indicadora de cada dimensión subyacente extraída y finalmente validándose la escala de medida de los constructos especificados y las correlaciones existentes entre los mismos.

5.3.1. Análisis Factorial Exploratorio

Para determinar la dimensionalidad del cuestionario, completar y confirmar los estudios de validez del mismo, se realizó un análisis factorial exploratorio, utilizando como método de extracción de las componentes el análisis de componentes principales y aplicando, en aquellos casos que fuese necesario, rotación oblicua de los factores extraídos

para facilitar la interpretación de los mismos. Se eligió este tipo de rotación debido a que se suponía que los componentes estaban correlacionados y la rotación ortogonal parte de la idea de que los componentes son independientes entre sí.

En este apartado se describen los resultados obtenidos trabajando en distintas situaciones. En primer lugar, se realizará el análisis factorial exploratorio sobre las puntuaciones individuales obtenidas en los cuestionarios del curso académico 2000/2001, por ser ésta la aplicación más reciente de la que se disponía información. A continuación, para comprobar la estabilidad temporal de las dimensiones medidas por el cuestionario, se efectuó un análisis factorial exploratorio de las puntuaciones individuales de los cuestionarios de la aplicación 1999/2000. Finalmente, se compararon los resultados obtenidos en ambas aplicaciones al trabajar con las puntuaciones medias obtenidas en cada uno de los ítems del cuestionario para cada combinación profesor-asignatura, en lugar de con las puntuaciones individuales.

5.3.1.1. Análisis de los datos del curso 2000/2001

Para abordar el presente estudio se partió de los 97.847 cuestionarios de evaluación del profesorado, correspondientes a las valoraciones individuales que los alumnos de la Universidad Politécnica de Valencia, realizaron de sus profesores en el curso académico 2000/2001. De acuerdo con los grupos definidos anteriormente, fue necesario estudiar, por un lado, los 76175 cuestionarios de los que los alumnos no tenían juicio suficiente para responder a los ítems codificados como R13, R14 y R15 referentes a evaluación y, por otro lado, los 21672 cuestionarios, en los que los estudiantes si poseían juicio necesario para responder a todos los ítems del cuestionario, incluidos aquellos referentes al sistema de evaluación del profesor.

En cada uno de estos grupos se realizó el análisis factorial utilizando dos posibles criterios para determinar el número de factores a extraer (el criterio de eigenvalue y un criterio basado en las características de construcción del cuestionario) y dos posibles matrices de entradas de datos (matriz de correlaciones y matriz de varianzas covarianzas).

A continuación se muestran los resultados obtenidos en cada caso.

5.3.1.1.1 Análisis factorial utilizando como criterio de extracción el criterio “eigenvalue” y como matriz de entrada de datos la matriz de correlaciones

En este caso, se decidió trabajar con el criterio de eigenvalue, también llamado autovalores mayores que uno o criterio de Kaiser, para determinar el número de factores a extraer. Con este criterio, el número de factores a extraer depende de que la varianza única domine a la varianza común, aunque se debe tener en cuenta que al trabajar con un número de variables inferior a 20, este procedimiento tiende a extraer un número de factores demasiado pequeño.

Como matriz de entrada de datos se utilizó la matriz de correlaciones ya que todas las variables están medidas en la misma escala.

a) Análisis de los datos del grupo B

Se comenzó evaluando si nuestros datos presentaban características apropiadas para la aplicación del análisis factorial. Para ello se estudiaron las correlaciones entre variables, detectándose que todas ellas presentan valores elevados, superiores a 0.30, con un nivel de significación inferior 0.001, lo que nos indica la conveniencia de proseguir con el análisis factorial.

También podemos contrastar la idea anterior con otro indicador, como la medida de adecuación de muestreo de Kaiser-Meyer-Olkin (KMO), que es un índice que compara las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parciales (García Jiménez et al., 2000):

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum a_{ij}^2}$$

donde r_{ij} es el coeficiente de correlación simple entre la variable i y la variable j y a_{ij} es el coeficiente de correlación parcial entre esas mismas variables.

Si la suma de los coeficientes de correlación parciales al cuadrado entre todos los pares de variables es pequeña comparada con la suma de los coeficientes de correlación al

cuadrado, la medida KMO se aproxima a la unidad, lo que indica la adecuación de utilizar el modelo factorial, ya que las correlaciones entre pares de variables pueden explicarse por otras variables. En nuestro caso el valor de KMO obtenido es de 0,970 que según la relación establecida por Kaiser y Caffrey (1965), indica la existencia de una relación “muy buena”.

En la tabla 5.2 se muestran las comunalidades (o proporción de varianza explicada por los componentes extraídos) obtenidas.

Tabla 5.2. Comunalidades

	Inicial	Extracción
R1	1,000	,672
R2	1,000	,612
R3	1,000	,582
R4	1,000	,640
R5	1,000	,636
R6	1,000	,563
R7	1,000	,527
R8	1,000	,500
R9	1,000	,609
R10	1,000	,587
R11	1,000	,621
R12	1,000	,703
R16	1,000	,516
R17	1,000	,558
R18	1,000	,568
R19	1,000	,790

En ella aparecen tanto la comunalidad inicial que en este caso es la unidad al trabajar con el método de componentes principales como la comunalidad cuando se han extraído el número de componentes deseado. Cuanto mayor sea el valor de estas comunalidades, más variabilidad de la variable es explicada por las componentes extraídas. En nuestro caso todos los valores son superiores a 0.5, lo que nos indica que todos los ítems son explicados por los componentes.

Se obtuvo sólo un componente extraído, con autovalor superior a uno, que explica el 60,532% de la varianza total, como se muestra en la tabla 5.3.

Tabla 5.3. Porcentaje de varianza explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	9,685	60,532	60,532	9,685	60,532	60,532
2	,882	5,510	66,042			
3	,750	4,686	70,729			
4	,630	3,937	74,665			
5	,504	3,148	77,813			
6	,449	2,807	80,619			
7	,393	2,453	83,073			
8	,383	2,393	85,466			
9	,377	2,354	87,820			
10	,350	2,187	90,007			
11	,328	2,050	92,057			
12	,321	2,004	94,062			
13	,289	1,808	95,870			
14	,253	1,583	97,453			
15	,222	1,387	98,839			
16	,186	1,161	100,000			

Comentar, que al estar utilizando como método de extracción el análisis de componentes principales y no haber sido necesaria en este caso ninguna rotación para facilitar la interpretación de los factores, los valores obtenidos para las sumas de las saturaciones al cuadrado de la extracción, coinciden con el de los autovalores iniciales, mientras que con otros procedimientos obtendríamos valores ligeramente inferiores, debido fundamentalmente a errores de medición.

En la tabla 5.4 se muestran las cargas factoriales obtenidas, que nos indican el grado de correlación entre las variables y los componentes. El cuadrado de estas cargas factoriales nos indica la proporción de varianza compartida por la variable y el componente.

Tabla 5.4. Cargas Factoriales

Componente	
1	
R1	,820
R2	,782
R3	,763
R4	,800
R5	,797
R6	,751
R7	,726
R8	,707
R9	,780
R10	,766
R11	,788
R12	,839
R16	,718
R17	,747
R18	,754
R19	,889

En ella podemos observar que el único componente está relacionado con todos los ítems (los doce primeros ítems, referentes a la interrelación profesor-alumno y al programa de la asignatura y los ítems codificados como R16, R17 y R18, referentes a los recursos utilizados). El ítem global R19 o ítem criterio el que presenta mayor correlación, seguido del ítem R12 (Consigue que los alumnos estén motivados por la asignatura) y R1 (explica en clase con orden y claridad).

A la vista de los resultados, se puede sugerir que los alumnos que en el momento de la encuesta no han efectuado ningún examen, responden a los ítems del cuestionario atendiendo a una sola dimensión de evaluación, presumiblemente actuación docente del profesor, incluyéndose en ésta la interrelación profesor-alumno.

b) Análisis de los datos del grupo A

En este caso se obtuvo un valor de la medida de adecuación muestral KMO de 0,974 indicándonos también que es adecuado proceder a la factorización. Las comunalidades obtenidas en la extracción se muestran en la tabla 5.5.

Tabla 5.5. Comunalidades

	Inicial	Extracción
R1	1,000	,731
R2	1,000	,677
R3	1,000	,646
R4	1,000	,701
R5	1,000	,697
R6	1,000	,637
R7	1,000	,596
R8	1,000	,570
R9	1,000	,656
R10	1,000	,627
R11	1,000	,668
R12	1,000	,731
R13	1,000	,739
R14	1,000	,739
R15	1,000	,686
R16	1,000	,609
R17	1,000	,613
R18	1,000	,675
R19	1,000	,801

En este caso se observa que los valores obtenidos son ligeramente superiores en casi todos los ítems respecto a las del análisis anterior, lo que indica que aumenta el porcentaje de varianza de estas variables que es explicada por las componentes.

En esta ocasión se obtuvieron dos componentes que explican un porcentaje de varianza total del 67,349%. El primero de ellos, como se muestra en la tabla 5.6, explica un 61,664%, siendo esta cifra ligeramente superior al obtenido en el análisis de los datos del grupo anterior.

Tabla 5.6. Porcentaje de varianza explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	11,716	61,664	61,664	11,716	61,664	61,664	11,382	61,664	61,664
2	1,080	5,685	67,349	1,080	5,685	67,349	7,909	5,685	67,349
3	,822	4,327	71,676						
4	,659	3,469	75,144						
5	,534	2,809	77,953						
6	,468	2,462	80,415						
7	,390	2,054	82,469						
8	,366	1,925	84,394						
9	,341	1,793	86,188						
10	,327	1,720	87,907						
11	,324	1,704	89,611						
12	,307	1,615	91,226						
13	,284	1,494	92,720						
14	,282	1,485	94,205						
15	,276	1,451	95,656						
16	,249	1,312	96,968						
17	,210	1,106	98,074						
18	,190	,999	99,073						
19	,176	,927	100,000						

La matriz de estructura se muestra en la tabla 5.7. En este caso, vemos que el componente 1 está relacionado con los doce primeros ítems y con los cuatro últimos, mientras que el componente dos está relacionado únicamente con las variables R13, R14 y R15, referentes a aspectos del sistema de evaluación.

Esto confirma que en el caso de que los alumnos que responden al cuestionario, hubiesen efectuado algún tipo de examen, responden atendiendo a una bimimensionalidad, por un lado, la labor docente del profesor y por otro, la evaluación. Ambas dimensiones están bastante correlacionadas ya que se obtuvo una correlación entre componentes de 0,68.

En este caso, se detecta que la variable R19, está fuertemente correlacionada con la primera componente, seguida de nuevo por R12 y R1, es decir, que “la consideración global

de ser un buen profesor”, se debe mayoritariamente a la labor docente que realice y a la interrelación profesor alumno.

Tabla 5.7. Matriz de Estructuras

	Componente	
	1	2
R1	,849	,506
R2	,821	,520
R3	,803	,523
R4	,837	,576
R5	,833	,530
R6	,777	,664
R7	,772	,518
R8	,754	,488
R9	,810	,567
R10	,790	,582
R11	,812	,620
R12	,853	,623
R13	,602	,859
R14	,602	,859
R15	,567	,828
R16	,719	,712
R17	,751	,673
R18	,769	,736
R19	,890	,674

5.3.1.1.2. Análisis factorial utilizando como criterio de extracción las características de elaboración del mismo utilizando como matriz de entrada la matriz de correlaciones

Tras analizar la información disponible acerca de la elaboración del CEDA de la UPV y, tomando como referencia las distintas fuentes bibliográficas existentes respecto al análisis de fiabilidad y consistencia de los cuestionarios de Evaluación Docente de otras universidades (Fernández y Mateo, 1991; Mateo y Fernández, 1992a y 1992b; De Miguel, 1991; Escudero, 1999), se plantea la necesidad de realizar de nuevo un análisis factorial con el objetivo de es validar el constructo de evaluación docente, de manera que se pueda

llegar a obtener un sistema de indicadores válido y fiable que permita diseñar un buen Sistema de Evaluación de la Calidad de la Docencia Universitaria.

Se pretendía identificar las 5 dimensiones siguientes:

- Cumplimiento formal de obligaciones docentes
- Estructuración y desarrollo del programa
- Dominio de las habilidades docentes
- Manejo de los recursos didácticos
- Justicia e imparcialidad en calificaciones (evaluación)

En nuestro caso, por las características de elaboración del cuestionario y por las referencias bibliográficas revisadas, se fijó el número de factores a extraer en 3 más que los que se estimaron inicialmente. En función de la información obtenida en estos análisis y tras interpretar las matrices factoriales, se valoró el sentido práctico de los factores extraídos en cada caso, escogiendo la solución que mejor representa los datos.

a) Análisis factorial de los datos del grupo A

En la tabla 5.8 se muestra el porcentaje de varianza explicada mediante los componentes extraídos. En este caso, se obtuvieron cinco factores que explicaban un porcentaje de varianza total del 77,953%.

Mediante el primer factor se explica el 61,664% de la variabilidad, porcentaje similar al obtenido con el criterio eigenvalue, mientras que el porcentaje de varianza explicada por los otros cuatro componentes es relativamente pequeño, oscilando entre un 5,685% en el caso del componente dos a un 2,809% en el caso del quinto componente.

Tabla 5.8. Porcentaje de varianza explicada

Comp	Autovalores iniciales		Sumas de las saturaciones al cuadrado de la extracción		Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total
1	11,716	61,664	61,664	11,716	61,664	61,664	9,647
2	1,080	5,685	67,349	1,080	5,685	67,349	6,677
3	,822	4,327	71,676	,822	4,327	71,676	7,540
4	,659	3,469	75,144	,659	3,469	75,144	7,732
5	,534	2,809	77,953	,534	2,809	77,953	6,164
6	,468	2,462	80,415				
7	,390	2,054	82,469				
8	,366	1,925	84,394				
9	,341	1,793	86,188				
10	,327	1,720	87,907				
11	,324	1,704	89,611				
12	,307	1,615	91,226				
13	,284	1,494	92,720				
14	,282	1,485	94,205				
15	,276	1,451	95,656				
16	,249	1,312	96,968				
17	,210	1,106	98,074				
18	,190	,999	99,073				
19	,176	,927	100,000				

La matriz de estructuras quedaría como se muestra en la tabla 5.9. En ella aparecen en **negrita** las cargas factoriales de los distintos ítems con cada uno de los factores con los que presentan mayor correlación.

En ella podemos detectar que:

- El componente 1 está correlacionado con los seis primeros ítems y con R19. Las variables indicadoras serían R1 (“*El profesor explica con orden y claridad*”) y R2 (“*Da la sensación de tener bien preparadas las clases*”) con lo que podríamos identificar esta dimensión “**Dominio, organización y claridad de la asignatura y del programa**”.

Dimensionalidad del Cuestionario

- El componente 2 está correlacionado con los ítems R13, R14 y 15 siendo la variable indicadora R13, con lo que podríamos identificarlo como “**Evaluación**”.
- El componente 3 está correlacionado con las variables R9, R10, R11 y R12, siendo las variables indicadoras R10 (“Dialoga con los alumnos sobre la marcha de las clases, tomando en cuenta sus opiniones”) y R11 (“Tiene una actitud receptiva hacia los alumnos”) por lo que podríamos identificarla como “**Interrelación profesor-alumno**”.
- El componente 4 está correlacionado con R16, R17 y R18, pudiendo identificarlo como “**Recursos**”.
- El componente 5 está correlacionado con R7 y R8, pudiendo identificarlo como “**Interrelación con otras asignaturas**”.

Tabla 5.9. Matriz de Estructuras

	Componente				
	1	2	3	4	5
R1	,916	,459	,558	,602	,494
R2	,881	,463	,523	,628	,475
R3	,835	,503	,546	,525	,548
R4	,838	,545	,589	,579	,603
R5	,828	,475	,649	,593	,528
R6	,755	,642	,482	,605	,635
R7	,637	,449	,533	,599	,862
R8	,578	,428	,603	,531	,885
R9	,656	,488	,843	,584	,598
R10	,613	,510	,902	,552	,556
R11	,676	,551	,893	,585	,491
R12	,752	,567	,773	,606	,632
R13	,550	,885	,413	,549	,440
R14	,517	,883	,535	,517	,389
R15	,470	,790	,536	,629	,265
R16	,613	,560	,510	,893	,485
R17	,642	,510	,541	,902	,519
R18	,706	,651	,538	,770	,532
R19	,855	,610	,711	,694	,562

Es decir, que se verifican las cinco dimensiones con las que se había construido el cuestionario inicialmente. Además, todas las componentes presentan correlaciones bastante elevadas con los otros factores, como se muestra en la tabla siguiente. El componente 1 presenta en todos los casos correlaciones superiores a 0,5, siendo el componente 4 con el que más fuertemente está correlacionado.

Tabla 5.10. Matriz de Correlación de los Componentes

Componente	1	2	3	4	5
1	1,000	,529	,602	,640	,575
2	,529	1,000	,483	,567	,380
3	,602	,483	1,000	,533	,490
4	,640	,567	,533	1,000	,474
5	,575	,380	,490	,474	1,000

b) Análisis factorial de los datos del grupo B

En este caso, al haber eliminado los ítems R13, R14 y R15, referentes a evaluación, las dimensiones que deberían identificarse eran 4. El porcentaje de varianza total explicada por las 4 componentes extraídas fue de 74,665%, aunque con el primer factor se consigue explicar un 60,532% de la misma, como se muestra en la tabla 5.11. Se extraían un total de 14 componentes que explicaban el 100% de la varianza.

Tabla 5.11. Porcentaje de varianza explicada por los cinco primeros componentes

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	9,685	60,532	60,532	9,685	60,532	60,532	8,254		
2	,882	5,510	66,042	,882	5,510	66,042	6,963		
3	,750	4,686	70,729	,750	4,686	70,729	6,993		
4	,630	3,937	74,665	,630	3,937	74,665	6,025		
5	,504	3,148	77,813						

Dimensionalidad del Cuestionario

En la tabla 5.12 se muestra la matriz de estructuras. En ella se observa que las 4 componentes están correlacionadas con las mismas variables que en el caso anterior, siendo la diferencia de dimensionalidad debida al factor “evaluación”. Además podemos observar que en este caso el ítem que presenta mayor correlación con el componente 1 pasa a ser el ítem R1, seguido del R19 y R2, mientras que el ítem R12, no presenta valores tan elevados como en el caso anterior.

Tabla 5.12. Matriz de Estructuras

	Componente			
	1	2	3	4
R1	,902	,570	,618	,529
R2	,864	,519	,601	,520
R3	,796	,579	,576	,521
R4	,808	,611	,604	,608
R5	,810	,652	,582	,555
R6	,723	,505	,700	,579
R7	,608	,527	,563	,886
R8	,530	,597	,536	,899
R9	,622	,853	,566	,631
R10	,579	,905	,563	,580
R11	,652	,890	,579	,523
R12	,735	,785	,671	,633
R16	,570	,510	,891	,502
R17	,611	,543	,861	,549
R18	,663	,552	,812	,510
R19	,867	,730	,711	,613

La matriz de correlaciones entre los factores quedaría como se muestra en la tabla 5.13. En este caso todos los componentes presentan correlaciones con el resto de factores superiores a 0,5, siendo los componentes más correlacionados el 1, denominado “**Dominio, organización y claridad de la asignatura y del programa**” y el 3, denominado “**Recursos**”, coincidiendo con lo obtenido en el grupo A.

Tabla 5.13. Matriz de Correlación entre los Factores

Componente	1	2	3	4
1	1,000	,645	,687	,603
2	,645	1,000	,588	,597
3	,687	,588	1,000	,584
4	,603	,597	,584	1,000

5.3.1.1.3 Análisis factorial exploratorio utilizando como criterio de extracción las características de elaboración del mismo y como matriz de entrada la matriz de varianzas-covarianzas

El objetivo de este nuevo estudio es analizar si existen diferencias en los resultados obtenidos al aplicar el Análisis Factorial utilizando como matriz de entrada de datos la matriz de correlaciones o la matriz de varianzas-covarianzas. Esta última matriz nos proporciona una información inicial sobre la variabilidad observada en el conjunto de variables de nuestro estudio: los valores de la diagonal principal representan la varianza total de cada una de las variables, mientras que la suma de los elementos de la diagonal principal, nos indica la cantidad de varianza contenida en esa matriz. Los índices de covariación fuera de la diagonal principal representan la varianza común entre dos variables. La utilización de esta matriz presenta la ventaja de que nos permite efectuar comparaciones válidas entre diferentes poblaciones o muestras, cosa que no es posible al utilizar la matriz de correlaciones. Sin embargo, presenta el inconveniente de que la interpretación de los resultados es más compleja, ya que los coeficientes deben interpretarse en términos de las unidades de medida de los constructos.

a) Análisis factorial de los datos del grupo A

Se procedió de manera similar a los casos anteriores, realizando un AF utilizando como principio de extracción el análisis de componentes principales rectificándose luego con rotación oblicua. Los cinco factores extraídos explicaban un porcentaje de varianza bruta total del 78,100%, como puede verse en la tabla 5.14.

Tabla 5.14. Porcentaje de varianza bruta explicada

	Comp	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación
		Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total
Bruta	1	16,462	61,679	61,679	16,462	61,679	61,679	13,463
	2	1,586	5,942	67,620	1,586	5,942	67,620	9,229
	3	1,186	4,443	72,063	1,186	4,443	72,063	11,452
	4	,890	3,334	75,397	,890	3,334	75,397	7,825
	5	,721	2,703	78,100	,721	2,703	78,100	9,883
	6	,650	2,434	80,533				
	7	,550	2,061	82,594				
	8	,519	1,943	84,537				
	9	,464	1,740	86,277				
	10	,461	1,729	88,006				
	11	,440	1,648	89,654				
	12	,412	1,543	91,198				
	13	,407	1,526	92,724				
	14	,381	1,426	94,149				
	15	,368	1,377	95,526				
	16	,354	1,328	96,854				
	17	,318	1,191	98,045				
	18	,271	1,015	99,060				
	19	,251	,940	100,000				

En la tabla 5.15 se presentan los valores de varianza reescalada. El reescalado únicamente afecta a los valores obtenidos tras la rotación, ya que podemos observar que el valor de los eigenvalores iniciales son los mismos en ambos casos. En este caso, al trabajar con el principio de extracción de componentes principales, el valor de la columna total de las sumas de las saturaciones al cuadrado de la extracción coincide con el valor de los autovalores iniciales. Los valores reescalados son, en realidad, estandarizados para lograr explicar la varianza en los autovalores iniciales.

Tabla 5.15. Porcentaje de varianza reescalada explicada

Comp	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación	
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	
Reescalada	1	16,462	61,679	61,679	11,712	61,640	61,640	9,639
	2	1,586	5,942	67,620	1,076	5,665	67,305	6,442
	3	1,186	4,443	72,063	,820	4,316	71,621	8,006
	4	,890	3,334	75,397	,660	3,472	75,093	5,631
	5	,721	2,703	78,100	,533	2,803	77,896	7,169
	6	,650	2,434	80,533				
	7	,550	2,061	82,594				
	8	,519	1,943	84,537				
	9	,464	1,740	86,277				
	10	,461	1,729	88,006				
	11	,440	1,648	89,654				
	12	,412	1,543	91,198				
	13	,407	1,526	92,724				
	14	,381	1,426	94,149				
	15	,368	1,377	95,526				
	16	,354	1,328	96,854				
	17	,318	1,191	98,045				
	18	,271	1,015	99,060				
	19	,251	,940	100,000				

Las matrices de estructuras y de configuración se muestran en las tablas 5.16 y 5.17 respectivamente. Comentar, que en los casos en los que se aplica rotación ortogonal, los valores de ambas matrices coinciden, proporcionándonos los valores de las cargas factoriales de cada variable en cada factor tras efectuar la rotación. Sin embargo, al aplicar rotación oblicua los valores de ambas matrices son distintos ya que las correlaciones parciales y las correlaciones simples no son iguales. En ellas encontramos también los valores brutos y reescalados. Los valores brutos nos proporcionan el valor de la covarianza entre las variables y los factores y están en la misma escala que la varianza de los ítems. Los valores reescalados han sido estandarizados para que todos los valores estén medidos en la misma escala, ajustados a las varianzas de los ítems y se mueven en un rango entre 0 y

1. El hecho de que la matriz de estructuras tenga valores superiores que los de la matriz de configuración para la variable que no se “resume” en cada componente confirma la relación entre componentes, lo que justificaría la utilización de la rotación oblicua.

Tabla 5.16. Matriz de Estructuras

	Bruta					Reescalada				
	Componente					Componente				
	1	2	3	4	5	1	2	3	4	5
R1	1,109	,543	-,697	,566	,675	,921	,451	-,579	,470	,561
R2	,991	,510	-,620	,508	,683	,876	,450	-,548	,449	,604
R3	,997	,595	-,675	,627	,562	,840	,501	-,568	,528	,473
R4	,948	,609	-,699	,654	,623	,833	,535	-,614	,575	,547
R5	,953	,533	-,762	,575	,657	,825	,462	-,660	,498	,569
R6	,870	,725	-,605	,704	,676	,751	,625	-,522	,608	,584
R7	,736	,511	-,650	,978	,665	,634	,440	-,560	,843	,573
R8	,665	,492	-,714	,995	,566	,578	,427	-,621	,864	,492
R9	,777	,562	-,998	,678	,656	,656	,474	-,842	,573	,554
R10	,774	,630	-1,163	,668	,650	,611	,497	-,917	,527	,513
R11	,827	,653	-1,101	,552	,681	,676	,534	-,901	,451	,557
R12	,963	,718	-1,008	,784	,698	,757	,564	-,792	,616	,548
R13	,666	1,076	-,543	,491	,663	,547	,884	-,446	,403	,545
R14	,657	1,160	-,716	,434	,642	,513	,906	-,559	,339	,502
R15	,551	,884	-,657	,223	,825	,465	,747	-,555	,188	,697
R16	,701	,614	-,606	,539	,991	,617	,540	-,533	,474	,872
R17	,712	,543	-,619	,559	,956	,647	,493	-,562	,508	,868
R18	,804	,715	-,647	,580	,857	,706	,628	-,568	,509	,752
R19	1,034	,721	-,881	,640	,797	,858	,598	-,731	,531	,661

En ellas podemos detectar que:

- El componente 1 está correlacionado con los seis primeros ítems y con R19. Las variable indicadoras serían de nuevo R1 y R2, con lo que podríamos identificar esta dimensión como “**Dominio, organización y claridad de la asignatura y del programa**”.

- El componente 2 está correlacionado con los ítems R13, R14 y 15 siendo la variable indicadora R13, con lo que podríamos identificarlo como “**Evaluación**”.
- El componente 3 está correlacionado con las variables R9, R10, R11 y R12, siendo las variables indicadoras R10 y R11 por lo que podríamos identificarla como “**Interrelación profesor-alumno**”.
- El componente 5 está correlacionado con R16, R17 y R18, pudiendo identificarlo como “**Recursos**”.
- El componente 4 está correlacionado con R7 y R8, pudiendo identificarlo como “**Interrelación con otras asignaturas**”.

Tabla 5.17. Matriz de configuración

	Bruta					Reescalada				
	Componente					Componente				
	1	2	3	4	5	1	2	3	4	5
R1	1,136	-,050	-,020	-,069	,040	,943	-,041	-,016	-,057	,034
R2	,961	-,046	,020	-,064	,169	,849	-,040	,017	-,057	,150
R3	,915	,134	-,047	,106	-,125	,770	,113	-,039	,089	-,105
R4	,731	,133	-,094	,163	,000	,642	,117	-,083	,143	,000
R5	,747	-,031	-,254	,017	,088	,647	-,027	-,220	,015	,076
R6	,521	,349	,102	,310	,104	,450	,301	,088	,268	,090
R7	,101	,039	-,066	,784	,238	,087	,034	-,057	,675	,205
R8	-,048	,059	-,264	,838	,092	-,041	,051	-,230	,729	,080
R9	,124	-,013	-,776	,201	,103	,105	-,011	-,655	,170	,087
R10	-,004	,043	-1,071	,137	,013	-,004	,034	-,845	,108	,010
R11	,197	,068	-,939	-,052	,054	,161	,056	-,768	-,042	,044
R12	,404	,158	-,570	,246	-,026	,317	,124	-,448	,193	-,021
R13	,127	1,003	,135	,134	,058	,104	,824	,111	,110	,048
R14	,010	1,089	-,194	,002	-,060	,008	,851	-,152	,001	-,047
R15	-,092	,601	-,266	-,255	,515	-,078	,507	-,225	-,215	,435
R16	,080	,057	-,018	,130	,851	,071	,050	-,016	,114	,749
R17	,128	-,047	-,064	,152	,811	,116	-,043	-,058	,138	,736
R18	,310	,245	-,015	,132	,478	,272	,215	-,013	,116	,419
R19	,672	,132	-,300	,019	,158	,557	,110	-,249	,016	,131

Las componentes presentan correlaciones bastante elevadas, como se muestra en la tabla 5.18.

Tabla 5.18. Matriz de Correlación entre los Componentes

Componente	1	2	3	4	5
1	1,000	,514	-,629	,551	,598
2	,514	1,000	-,502	,328	,545
3	-,629	-,502	1,000	-,480	-,524
4	,551	,328	-,480	1,000	,397
5	,598	,545	-,524	,397	1,000

b) Análisis factorial de los datos del grupo B

En este caso, al haber eliminado los ítems R13, R14 y R15, referentes a evaluación, las dimensiones que se identificaron fueron 4. El porcentaje de varianza bruta total y reescalada explicada por las 4 componentes extraídas fue de 74,832%, como se muestra en la tabla 5.19. Se extrajeron un total de 16 componentes que explicaban el 100% de la varianza.

Tabla 5.19. Porcentaje de varianza explicada por las cinco componentes extraídas

	Compp	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las sat. al cuadrado de la rotación
		Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total
Bruta	1	11,717	60,692	60,692	11,717	60,692	60,692	9,935
	2	1,099	5,694	66,386	1,099	5,694	66,386	8,604
	3	,875	4,532	70,918	,875	4,532	70,918	8,429
	4	,756	3,914	74,832	,756	3,914	74,832	7,263
	5	,584	3,023	77,855				
Reesc	1	11,717	60,692	60,692	9,681	60,507	60,507	8,266
	2	1,099	5,694	66,386	,880	5,498	66,005	6,964
	3	,875	4,532	70,918	,748	4,674	70,678	7,029
	4	,756	3,914	74,832	,632	3,949	74,627	6,024
	5	,584	3,023	77,855				

Las matriz de estructuras se muestra en la tabla 5.20, resaltando en negrita los ítems de mayor correlación con cada factor.

Tabla 5.20. Matriz de Estructuras

	Bruta				Reescalada			
	1	2	3	4	1	2	3	4
R1	1,005	-,631	,687	,588	,906	-,569	,620	,530
R2	,890	-,545	,630	,547	,854	-,523	,605	,525
R3	,914	-,648	,645	,582	,811	-,575	,572	,516
R4	,843	-,639	,639	,636	,804	-,610	,609	,606
R5	,851	-,681	,621	,593	,805	-,644	,587	,561
R6	,769	-,550	,758	,613	,717	-,513	,707	,572
R7	,667	-,579	,617	,982	,608	-,528	,562	,895
R8	,573	-,645	,580	,957	,533	-,599	,539	,889
R9	,698	-,946	,633	,709	,624	-,847	,566	,635
R10	,686	-1,087	,668	,682	,578	-,917	,563	,576
R11	,733	-,986	,647	,584	,657	-,883	,579	,523
R12	,884	-,940	,805	,752	,742	-,789	,676	,632
R16	,621	-,557	,976	,548	,568	-,510	,894	,502
R17	,635	-,563	,873	,574	,613	-,544	,842	,554
R18	,724	-,605	,904	,555	,660	-,552	,824	,506
R19	,954	-,798	,784	,675	,869	-,727	,714	,615

En la tabla 5.21 se muestra la matriz de configuración. En ellas se muestra de nuevo en negrita la correlación de cada ítem con cada factor. En este caso se detecta de nuevo que el componente 1 está correlacionado con los 6 primeros ítems y con el ítem R19, el segundo componente con los ítems R9, R10, R11 y R12, el tercero con los ítems R16, R17 y R18 y el cuarto con los ítems R7 y R8.

Es decir, la única diferencia de dimensionalidad se debe únicamente de nuevo al factor evaluación, cuyos ítems no se consideran en este grupo de datos.

Finalmente, en la tabla 5.22 se muestra la correlación obtenida entre los distintos componentes. En ella podemos observar que el componente dos presenta correlaciones negativas con el resto de factores.

Tabla 5.21. Matriz de configuración

	Bruta				Reescalada			
	1	2	3	4	1	2	3	4
R1	1,034	,026	,004	-,024	,932	,024	,004	-,022
R2	,892	,074	,043	,026	,856	,071	,042	,025
R3	,847	-,092	-,003	,016	,752	-,082	-,003	,014
R4	,662	-,096	,033	,159	,631	-,092	,031	,152
R5	,690	-,205	-,010	,059	,653	-,194	-,010	,056
R6	,442	,064	,405	,146	,412	,060	,378	,137
R7	,141	,086	,024	,934	,129	,078	,022	,851
R8	-,084	-,142	,022	,910	-,078	-,132	,020	,846
R9	,066	-,779	,012	,196	,059	-,697	,010	,176
R10	-,082	-1,073	,056	,058	-,070	-,905	,048	,049
R11	,169	-,898	,048	-,083	,151	-,804	,043	-,075
R12	,332	-,546	,185	,117	,279	-,458	,155	,098
R16	-,102	-,010	1,045	-,007	-,093	-,009	,957	-,007
R17	,018	-,036	,794	,077	,017	-,035	,766	,075
R18	,181	-,056	,767	-,036	,165	-,051	,700	-,033
R19	,658	-,262	,160	,027	,600	-,238	,146	,024

Tabla 5.22. Matriz de correlación de los componentes

Componente	1	2	3	4
1	1,000	-,647	,690	,605
2	-,647	1,000	-,591	-,598
3	,690	-,591	1,000	,585
4	,605	-,598	,585	1,000

En consecuencia, podemos concluir que no existen diferencias significativas al utilizar como matriz de entrada de datos la matriz de correlaciones o la matriz de varianza-covarianza. En ambos casos se han obtenido el mismo número de factores extraídos y prácticamente la misma proporción de varianza explicada. En este sentido y de acuerdo con Hair et al. (1995) entre otros autores, se propone la utilización de la matriz de correlaciones, ya que simplifica, en cierta manera, la interpretación de los resultados al tener un rango

común que hace directas las comparaciones de los coeficientes dentro del modelo, aunque no permite efectuar comparaciones válidas entre distintas poblaciones o muestras.

5.3.2. Análisis Factorial Exploratorio trabajando con Valores Medios

Siguiendo las líneas de trabajo de diversos autores (Escudero, 1987, 1988, 1989a y 1989b; Jornet et al., 1989; Jornet et al., 1995 y 1996; Tejedor, 1991, Tejedor et al., 1988, etc.) se decidió efectuar finalmente, un análisis factorial exploratorio trabajando con las puntuaciones medias obtenidas para cada posible combinación asignatura-profesor, de manera que pudiésemos eliminar la posible influencia de observaciones extremas (encuestas en las que se había puntuado extremadamente alto o extremadamente bajo a los profesores correspondientes). Además de esta manera pueden compararse resultados por grupos (combinación profesor asignatura).

En este caso, el análisis se efectuó sin considerar los ítems de evaluación, y se efectuó tanto para el curso académico 2000/2001 como para el 1999/2000, debido a que por los problemas que se han comentado sobre las bases de datos, éstas estaban incompletas, y al realizar el cruce de las mismas únicamente pudimos obtener 515 observaciones en el primer caso y 2147 en el segundo. En la aplicación 2000/2001 detectamos que la mayoría de las observaciones correspondían a unas determinadas titulaciones y asignaturas, no pudiendo considerar los datos como una muestra aleatoria y representativa, siendo necesario utilizar la base del curso anterior para el resto de análisis efectuados en los capítulos posteriores.

Se utilizó como criterio de extracción las características de construcción del CEDA y como matriz de entrada de datos la matriz de correlaciones.

5.3.2.1 Análisis de los datos del curso académico 2000/2001

Como en los casos anteriores se realizó un análisis factorial utilizando como método de extracción el análisis de componentes principales y aplicando una rotación oblicua de los ejes para facilitar la interpretación de los factores extraídos.

Como se muestra en la tabla 5.23, se obtuvieron un total de 4 factores extraídos, coincidiendo con lo obtenido en los apartados anteriores, que explicaban un total de varianza del 87,517%, porcentaje ligeramente superior al 77,953% obtenido trabajando con observaciones individuales. El primer factor es el más que contribuye a la varianza total explicando un 76,369%, porcentaje también superior al 61,664% que obteníamos en el caso anterior.

Tabla 5.23. Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación(a)
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total
	1	12,219	76,369	76,369	12,219	76,369	76,369
2	,831	5,193	81,562	,831	5,193	81,562	8,655
3	,491	3,071	84,633	,491	3,071	84,633	8,709
4	,461	2,884	87,517	,461	2,884	87,517	6,796
5	,345	2,157	89,674				
6	,300	1,876	91,550				
7	,232	1,447	92,997				
8	,210	1,315	94,312				
9	,186	1,162	95,474				
10	,171	1,069	96,543				
11	,141	,878	97,421				
12	,123	,769	98,190				
13	,112	,697	98,887				
14	,076	,476	99,363				
15	,059	,371	99,734				
16	,043	,266	100,000				

La matriz de estructura se muestra en la tabla 5.24. Las dimensiones extraídas pueden considerarse las mismas que trabajando con los valores individuales. Se puede apreciar que cada uno de los ítems está relacionado con el mismo factor que en el caso anterior, excepto el ítem R12, “*Consigue que los alumnos estén motivados por la asignatura*”, que en este caso aparece bastante correlacionado tanto con el componente 1

como con el 2, siendo la correlación con el primero de ellos ligeramente superior. Por otro lado, las cargas de los ítems con el componente 4 presentan signo negativo, es decir, que la proyección de las observaciones sobre este eje factorial va en sentido inverso, cosa que no sucedía trabajando con valores individuales. También se detecta que los valores de las correlaciones entre cada variable y el factor correspondiente son ligeramente superiores a los obtenidos con valores individuales. De nuevo se obtuvo que el ítem de mayor correlación con el componente 1 es el R1, “*El profesor explica en clase con orden y claridad*”, seguido del ítem R19 o ítem criterio y del ítem R2, “*Da la sensación de tener bien preparadas las clases*”.

Tabla 5.24. Matriz de estructura

	Componente			
	1	2	3	4
MEDIA 1	,963	,617	,730	-,581
MEDIA 2	,924	,547	,744	-,617
MEDIA 3	,872	,716	,651	-,551
MEDIA 4	,917	,745	,683	-,619
MEDIA 5	,908	,723	,731	-,656
MEDIA 6	,889	,627	,767	-,554
MEDIA 7	,706	,625	,662	-,923
MEDIA 8	,709	,713	,607	-,873
MEDIA 9	,737	,917	,644	-,681
MEDIA 10	,722	,959	,657	-,624
MEDIA 11	,738	,943	,657	-,604
MEDIA 12	,876	,824	,732	-,723
MEDIA 16	,761	,619	,928	-,558
MEDIA 17	,739	,600	,931	-,594
MEDIA 18	,815	,644	,805	-,415
MEDIA 19	,951	,775	,782	-,680

A la vista de los resultados obtenidos, podemos concluir que si consideramos los ítems del cuestionario como variables independientes entre sí, por la aplicación del Teorema Central del Límite, la media de las observaciones tenderá a seguir una distribución normal y podremos evitar las consecuencias de posibles valores extremos, mejorando los

resultados obtenidos mediante la aplicación del Análisis Factorial, al incrementarse el porcentaje de varianza explicada.

5.3.2.2. Análisis de los datos del curso académico 1999/2000

Con la finalidad de comprobar si existen variaciones en cuanto a la estructura dimensional del CEDA en distintos periodos evaluativos, se repitió la aplicación del análisis factorial a los datos del curso académico 1999/2000, ya que, por los motivos comentados anteriormente, va a constituir la base de trabajo para los capítulos posteriores.

Como en los casos anteriores se realizó un análisis factorial utilizando como método de extracción el análisis de componentes principales y aplicando una rotación oblicua de los ejes para facilitar la interpretación de los factores extraídos. El criterio para determinar el número de factores a extraer

En la tabla 5.25, se muestra el porcentaje de varianza extraída por los de 4 factores extraídos, coincidiendo con lo obtenido en los apartados anteriores, que explicaban un total de varianza del 87,161% muy similar al obtenido en el curso académico 2000/2001, siendo el primer factor el que más contribuía a la misma, explicando un 74,9%, porcentaje también ligeramente superior al obtenido en el caso anterior. Se obtenían un total de 16 factores que explicaban el 100% de la varianza.

Tabla 5.25. Varianza total explicada por las 5 primeras componentes

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación(a)
	Total	% de la varianza	% Acumulado	Total	% de la Varianza	% acumulado	Total
1	11,984	74,900	74,900	11,984	74,900	74,900	10,449
2	,884	5,526	80,426	,884	5,526	80,426	8,533
3	,568	3,547	83,973	,568	3,547	83,973	8,742
4	,510	3,188	87,161	,510	3,188	87,161	6,561
5	,364	2,274	89,435				

La matriz de estructura se muestra en la tabla 5.26, pudiendo apreciar como en el caso anterior que cada uno de los ítems está relacionado con el mismo factor que trabajando con los valores individuales excepto el ítem R12, que vuelve a aparecer con mayor carga en el componente 1, con lo que las dimensiones extraídas pueden considerarse las mismas. De nuevo las cargas con el componente 4 aparecen con signo negativo.

Tabla 5.26. Matriz de estructura

	Componente			
	1	2	3	4
MEDIA R1	,969	,605	,723	-,544
MEDIA R2	,929	,533	,734	-,540
MEDIA R3	,826	,727	,623	-,566
MEDIA R4	,904	,759	,738	-,615
MEDIA R5	,906	,732	,716	-,615
MEDIA R6	,858	,625	,818	-,529
MEDIA R7	,648	,616	,601	-,934
MEDIA R8	,656	,681	,622	-,887
MEDIA R9	,728	,925	,663	-,648
MEDIA R10	,660	,951	,645	-,649
MEDIA R11	,743	,935	,667	-,615
MEDIA R12	,857	,834	,725	-,686
MEDIA R16	,716	,591	,923	-,567
MEDIA R17	,712	,576	,898	-,639
MEDIA R18	,750	,646	,855	-,371
MEDIA R19	,954	,763	,776	-,634

De esta manera podemos confirmar que la dimensionalidad del cuestionario se mantiene en los distintos periodos evaluativos y que los resultados obtenidos son muy similares trabajando con observaciones individuales y con valores medios, aunque en el segundo caso se mejora ligeramente el porcentaje de varianza explicado por las componentes extraídas. Únicamente el ítem R12, “*Consigue que los alumnos estén motivados por la asignatura*”, aparece bastante correlacionado tanto con el componente que habíamos denominado dominio, organización y claridad de la asignatura y del programa como con el componente interrelación profesor-alumno. Finalmente, los ítems R7, “*Da a la*

asignatura un enfoque aplicado” y R8, “*Establece conexiones con otras asignaturas*” presentan un valor de carga negativo con la componente a la que habíamos denominado “interrelación con otras asignaturas”, de manera que la proyección factorial de los objetos queda en sentido inverso.

5.3.3. Análisis Factorial Confirmatorio

En los cuestionarios de evaluación docente mediante opinión de los estudiantes, las variables de interés son frecuentemente de naturaleza no observable, siendo necesario hacerlo de forma indirecta recurriendo a determinadas variables o indicadores no observables. Estos indicadores intentan medir sus conceptos de referencia lo más significativamente posible y una forma de incrementar esta significación es a través de la combinación de indicadores entre sí, de forma que surjan sistemas a través de los cuales delimitar empíricamente lo indicado. Cada constructo teórico ha de estar representado por unos indicadores, de manera que éste quedará entonces delimitado por lo que tienen en común todos los indicadores. La bondad de la medición depende de la relación entre esos indicadores observables y los constructos subyacentes. En este sentido, el análisis factorial confirmatorio se presenta como una técnica multivariante para comprobar la fiabilidad y validez del instrumento de medida y confirmar las estructuras de ítems propuestas. De acuerdo con esta metodología las estructuras que se propongan pueden considerarse modelos de ecuaciones estructurales en los que se especifican las variables para cada factor latente (Hair et al., 1995). Las etapas necesarias para su aplicación se detallan a continuación.

5.3.3.1. Desarrollo de los modelos teóricos y construcción de los diagramas de secuencias

Siguiendo la corriente definida por Feldman (1977, 1978, 1979, 1983 y 1984) y Marsh (1980, 1982a, 1982b, 1984, 1987a y 1987b), que centran la calidad de la docencia universitaria en el propio docente y parten de la hipótesis de que la enseñanza de calidad es un constructo multidimensional cuyas principales características pueden valorarse

aplicando un cuestionario a los alumnos, han surgido multitud de trabajos en los que se afirma que cuando la docencia del profesorado se evalúa por los alumnos, se constatan unas dimensiones asociadas con la calidad de la enseñanza universitaria y se encuentra una estructura factorial bastante bien definida (Tejedor, 1991 y Marsh, 1987a, 1987b).

Sobre la base de configuración del CEDA de la UPV (Informe sobre la evaluación del profesorado del Instituto de las Ciencias de la Educación de la *Universidad Politécnica de Valencia*, 1994), de los análisis exploratorios realizados y de la extensa bibliografía revisada al respecto (Ibáñez- Martín, 1990; Cajide, 1994, De la Orden, 1987 y 1988; Escudero, 1991; González Such, 1997; De Miguel, 1998; Mora, 1998; Luxán, 1998; etc.), se formularon los modelos que definen las 5 estructuras factoriales alternativas hipotéticas subyacentes en los datos del grupo A y las 4 estructuras factoriales alternativas hipotéticas subyacentes en los datos del grupo B. Las características de estos modelos se detallan a continuación.

▪ *Datos del grupo A*

En el caso de los cuestionarios que habían sido respondidos por alumnos ya evaluados, definiremos 5 posibles modelos a evaluar, como se muestran en la figura 5.15. En estos diagramas, aparecen representadas todas las posibles relaciones causales y se asume que la naturaleza de todas ellas es lineal.

- MODELO 1: Se parte de la hipótesis de dos factores correlacionados (solución oblicua).
- MODELO 2: Se parte de la hipótesis de dos factores independientes (solución ortogonal).
- MODELO 3: Se definen dos factores, correlacionados con 18 de las variables del cuestionario, que están a su vez correlacionados con la otra variable exógena, correspondiente al ítem criterio del cuestionario.
- MODELO 4: Se definen 5 factores correlacionados (solución oblicua).

- MODELO 5: Se definen 5 factores correlacionados con 18 de las variables del cuestionario, que están a su vez todos ellos correlacionados con el ítem criterio del cuestionario.

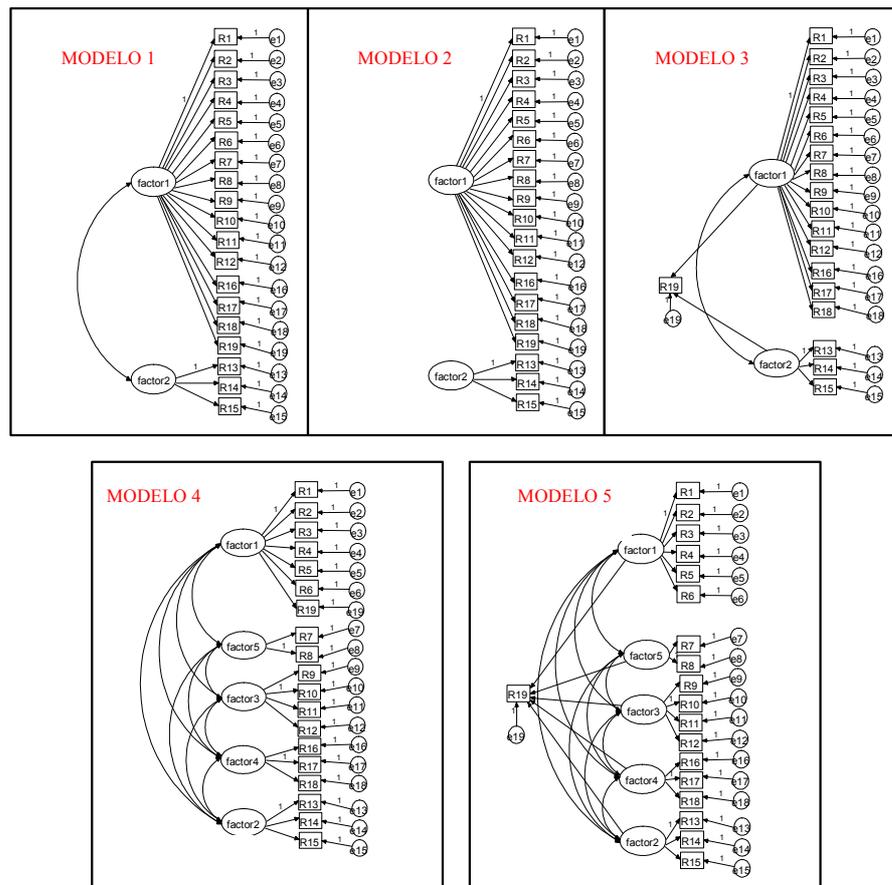


Figura 5.15. Modelos del grupo A

- *Datos del grupo B*

En el caso de los cuestionarios que habían sido respondidos por alumnos aún no evaluados, definiremos 4 modelos como se muestran en la figura 5.16, ya que no se

plantarán modelos incorrelaciones al haber resultado el modelo de peor ajuste entre todos los plantados en el grupo anterior:

- MODELO 1: Unidimensionalidad de la calidad de la enseñanza: Todos los elementos del cuestionario están más o menos igualmente relacionados entre sí.
- MODELO 2: Aparecen dos dimensiones latentes correlacionadas entre sí.
- MODELO 3. Se definen 4 dimensiones latentes relacionadas entre sí.
- MODELO 4: Se definen 4 dimensiones latentes correlacionadas entre sí, relacionados con 18 de las variables del cuestionario, que están a su vez relacionados con el ítem R19.

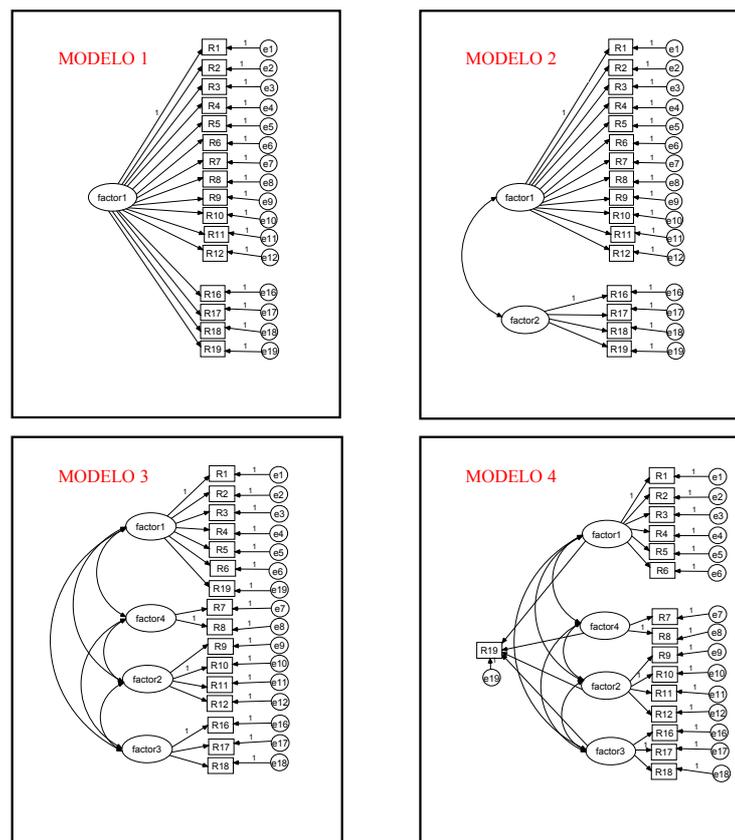


Figura 5.16. Modelos propuestos para los datos del grupo B

De todos los modelos anteriores cabe pensar que se obtendría un mejor ajuste en los modelos 4 y 5, dentro de los definidos en el grupo A y en los modelos 3 y 4 en los definidos en el grupo B, por ser los que presentan una estructura subyacente con un número de dimensiones o factores correlacionados similares a los obtenidos en el análisis factorial exploratorio efectuado en el aparto anterior y desarrollados en base a la construcción del cuestionario.

En nuestro caso, las variables de medida son los 19 ítems del cuestionario, correlacionados en los distintos modelos con las variables latentes correspondientes. La asignación de indicadores específicos (variables observables) a dimensiones concretas (constructos) es una de las mayores aportaciones a la perspectiva confirmatoria. Como se ha comentado, esta etapa de especificación concierne más con el conocimiento teórico que se tenga que con el instrumental estadístico. Sin embargo, en nuestro caso la realización del análisis factorial exploratorio nos ha permitido descubrir y detectar hechos característicos y relaciones interesantes entre las variables o indicadores, sin imponer ningún modelo a los datos, facilitándonos esta etapa de especificación. Asimismo, se procuró que todas las variables manifiestas estuvieran incluidas en el modelo, de manera que se evitara cometer errores de especificación (omisión de alguna variable predictiva clave o suposición incorrecta de implicaciones causales).

5.3.2.2 Valoración de la identificación del modelo

Un problema de identificación es la incapacidad del modelo propuesto para generar estimaciones aisladas de los parámetros. El software utilizado, analiza la identificación de los parámetros del modelo, indicando los posibles parámetros que presentan problemas y, en ese caso, no procediendo a la estimación de los mismos.

Todos los modelos planteados incluyen 19 variables iniciales diseñadas para la medida de las distintas dimensiones docentes mediante el cuestionario. En ninguno de nuestros modelos existen problemas de identificación, ya que el número de grados de libertad es muy superior a 0, siendo del tipo de modelos que habitualmente se conocen

como “posiblemente sobreidentificados”, es decir, que incluyen menos parámetros que varianzas y covarianzas.

Se fijaron a uno las saturaciones correspondientes a una variable en cada uno de los factores, siguiendo como criterio el Índice de Homogeneidad mayor respecto a la Escala Total. El análisis factorial confirmatorio, presenta la ventaja frente a la perspectiva exploratoria de permitir restricciones en algunas saturaciones, es decir, permite suponer que cada ítem satura en un único factor, de manera que se consigue delimitar el concepto de factor común a aquel que subyace únicamente a sus indicadores concretos, evitando introducir factores ad hoc que puedan dificultar posteriormente la interpretación.

Además, en nuestro caso, como los modelos incluyen variables latentes se les ha obligado a cumplir las tres condiciones, comentadas anteriormente, propuestas por Batista Foguet y Coenders (2000) necesarias para evitar problemas de identificabilidad.

5.3.2.3 Estimación del modelo propuesto y evaluación de los criterios de calidad de ajuste

El las tablas siguientes se muestran los principales indicadores de ajuste, de los distintos modelos del grupo A y B. El significado e interpretación de cada uno de ellos ha sido comentado anteriormente.

Tabla 5.27. Parámetros de los modelos del grupo A

MODELO	Chi-Cuadrado	G.I	G.F.I	A.G.F.I	N.F.I	C.F.I	RMSEA
1	30737,774	151	0,884	0,804	0,912	0,912	0,097
2	46687,093	152	0,813	0,766	0,866	0,866	0,119
3	30712,179	150	0,844	0,803	0,912	0,912	0,097
4	14523,641	142	0,932	0,909	0,958	0,959	0,068
5	12942,710	138	0,938	0,681	0,963	0,963	0,065

Tabla 5.28. Parámetros de los modelos del grupo B

MODELO	Chi-Cuadrado	G.I	G.F.I	A.G.F.I	N.F.I	C.F.I	RMSEA
1	79944	104	0,863	0,821	0,913	0,913	0,100
2	78299,556	103	0,865	0,655	0,915	0,915	0,100
3	36100	98	0,943	0,921	0,961	0,961	0,069
4	33211.142	95	0,946	0,923	0,964	0,964	0,068

En todos los modelos se obtuvo un elevado valor del estadístico Chi-cuadrado. Sin embargo, y dado que para tamaños de muestra elevados esta medida tiene tendencia a indicar diferencias significativas para modelos equivalentes, se debe considerar este dato meramente con carácter orientativo.

Los índices G.F.I y AGFI (Joreskog y Sorbön, 1981) que son independientes del tamaño muestral y menos sensibles a desviaciones de la normalidad de las variables, muestran en los modelos cuatro y cinco dentro de los del grupo A y en los modelos tres y cuatro dentro de los del grupo B, valores muy próximos a 1, que es el valor indicativo de que el ajuste es aceptable.

Asimismo, el valor de RMSEA (error cuadrático medio de aproximación), toma valores muy pequeños, próximos a cero, en los modelos 4 y 5 del grupo A y en los modelos 3 y 4 del grupo B.

A la vista de los resultados y evaluando los restantes índices de bondad de ajuste y los residuos normalizados, en los que la proporción con valores superiores a cinco y dos respectivamente, es menor en los modelos 4 y 5, en el caso de los cuestionarios en los que se habían contestado los ítems relativos a evaluación y, en los modelos del 3 y 4 en el caso de los cuestionarios en los que no se habían incluido, parece plausible proponer estos modelos con cinco y cuatro dimensiones latentes respectivamente, representativos de la estructura subyacente del cuestionario.

Por otro lado, el análisis de los modelos propuesto pone de manifiesto la existencia de correlaciones entre los factores, independientemente del número de éstos.

5.3.2.4. Interpretación de los modelos

Tras estimar los modelos y evaluar la calidad de ajuste, es necesario interpretar los resultados verificando la estructura de los factores propuestos y analizar si son necesarias modificaciones significativas, examinando los residuos normalizados y los índices de modificación.

a) Modelos del Grupo A

A continuación se presentan las estimaciones de los parámetros del modelo 5, en la solución estandarizada, pudiendo consultarse los parámetros no estandarizados en el anexo 4. Como puede apreciarse en la tabla 5.29, el número de parámetros a estimar es mucho menor que el número de varianzas y covarianzas originales, siendo la diferencia entre ambos valores lo que constituye los grados de libertad del modelo.

Tabla 5.29. Descripción del modelo 5

Nº de variables en el modelo	43
Nº de variables exógenas	24
Nº de variables endógenas	19
Nº de varianzas y covarianzas de origen	190
Nº de parámetros a estimar	52
Grados de Libertad	138

En el esquema o “path diagram” 5.17, se muestra la solución estandarizada obtenida para el modelo 5. Esta solución estandarizada presenta las saturaciones y los errores de medición de los indicadores ($1 - R^2$). En este diagrama se pueden observar las relaciones entre los constructos, entre las variables observables o indicadores y entre constructos e indicadores. Cada variable observable viene afectada por un término de error de medida.

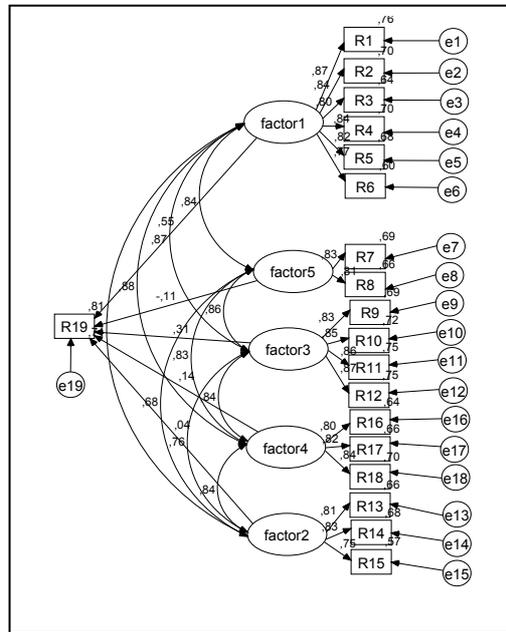


Figura 5.17. Solución estandarizada del modelo 5

De ella resultan interesantes los valores de correlaciones entre factores, más fáciles de interpretar que las covarianzas originales. Dos correlaciones próximas a la unidad, indicarían que dos o más dimensiones pueden confundirse en una sola. En nuestro caso, todas son menores que 0,9. Para facilitar la comprensión del mismo, se muestran en las tablas siguientes todos los valores de los parámetros representados en este path diagram.

Los coeficientes de regresión estandarizados obtenidos, que aparecen en el path diagram sobre las flechas que unen cada ítem con el factor no observado, se muestran en la tabla 5.30. En ella podemos observar que los efectos estandarizados entre los distintos ítems y factores son bastante elevados, positivos y significativos en todos los casos, excepto en el valor obtenido entre el ítem R19 y los factores 2, 3, 4 y 5, es decir que la relación de este ítem criterio con estos factores es relativamente pequeña. En el caso particular del valor obtenido entre el ítem R19 y el factor 5, destaca además el hecho de que sea negativo, esto nos indicaría que la relación entre ellos se produce en sentido inverso.

Tabla 5.30. Coeficientes de regresión estandarizados para el modelo 5

Coef. Regresión	Estimado	Coef. Regresión	Estimado
R14← factor2	0,825	R12← factor3	0,867
R15← factor2	0,753	R7← factor5	0,828
R13← factor2	0,809	R8← factor5	0,812
R6← factor1	0,774	R16← factor4	0,801
R5← factor1	0,824	R18← factor4	0,836
R4← factor1	0,835	R17← factor4	0,815
R3← factor1	0,799	R19← factor5	-0,114
R2← factor1	0,839	R19← factor3	0,311
R1← factor1	0,869	R19← factor4	0,144
R9← factor3	0,831	R19← factor2	0,04
R10← factor3	0,847	R19← factor1	0,548
R11← factor3	0,864		

Las correlaciones entre los factores se muestran en la tabla 5.31. Las covarianzas obtenidas pueden consultarse en el anexo 5.

Tabla 5.31. Correlación estimada entre los factores del modelo 5

CORRELACIÓN	ESTIMADO
factor 2 ↔ factor 1	0,740
factor 1 ↔ factor 5	0,837
factor 1 ↔ factor 3	0,870
factor 3 ↔ factor 5	0,860
factor 1 ↔ factor 4	0,878
factor 3 ↔ factor 4	0,842
factor 5 ↔ factor 4	0,832
factor 2 ↔ factor 4	0,835
factor 2 ↔ factor 3	0,759
factor 2 ↔ factor 5	0,676

En ellas se puede observar que entre todos los constructos existen correlaciones bastante elevadas, siendo el factor 1 el que presenta mayores correlaciones con los otros

factores y el factor 2 el que presenta correlaciones más bajas, aunque en todos los casos son superiores a 0,50.

Por último, en la tabla 5.32, se muestran los valores obtenidos para las correlaciones al cuadrado, mostrándose las varianzas en el anexo 5.

Tabla 5.32. Correlaciones al cuadrado estimadas para el modelo 5

SQUARED MÚLTIPLE CORRELATIONS	ESTIMADO	SQUARED MÚLTIPLE CORRELATIONS	ESTIMADO
R19	0,809	R9	0,690
R13	0,655	R8	0,660
R15	0,567	R7	0,686
R14	0,681	R6	0,600
R18	0,699	R5	0,680
R17	0,665	R4	0,698
R16	0,641	R3	0,639
R12	0,751	R2	0,704
R11	0,747	R1	0,755
R10	0,717		

El coeficiente de correlación múltiple al cuadrado para las variables observables en los modelos con variables latentes es una estimación de la fiabilidad de cada variable, es decir, una estimación de la estabilidad de la medida. Se puede observar que en este caso las de mayor valor, es decir, las más fiables son R19, R1 y R12.

En consecuencia, las ecuaciones del modelo, tanto las que relacionan factores como las que relacionan factores e indicadores, quedarían como se muestran a continuación.

Las ecuaciones que relacionan a los factores entre sí, muestran que los efectos de todos ellos son positivos y significativos. El efecto de los cinco factores sobre la “consideración general de buen profesor” (ítem R19), explica un 80,9 % de la varianza del mismo.

$$\begin{aligned}
 & \text{factor } 1 = 0,740 \quad \text{factor } 2 + 0,870 \quad \text{factor } 3 + 0,878 \quad \text{factor } 4 + 0,837 \quad \text{factor } 5 + d_1 \\
 & \text{factor } 2 = 0,740 \quad \text{factor } 1 + 0,759 \quad \text{factor } 3 + 0,835 \quad \text{factor } 4 + 0,832 \quad \text{factor } 5 + d_2 \\
 & \text{factor } 3 = 0,870 \quad \text{factor } 1 + 0,759 \quad \text{factor } 2 + 0,842 \quad \text{factor } 4 + 0,860 \quad \text{factor } 5 + d_3 \\
 & \text{factor } 4 = 0,878 \quad \text{factor } 1 + 0,835 \quad \text{factor } 2 + 0,842 \quad \text{factor } 3 + 0,835 \quad \text{factor } 5 + d_4 \\
 & \text{factor } 5 = 0,837 \quad \text{factor } 1 + 0,676 \quad \text{factor } 2 + 0,860 \quad \text{factor } 3 + 0,832 \quad \text{factor } 4 + d_5 \\
 \\
 & R_1 = 0,869 \quad \text{factor } 1 + e_1; R^2 = 0,755 \\
 & R_2 = 0,839 \quad \text{factor } 1 + e_2; R^2 = 0,704 \\
 & R_3 = 0,799 \quad \text{factor } 1 + e_3; R^2 = 0,639 \\
 & R_4 = 0,835 \quad \text{factor } 1 + e_4; R^2 = 0,698 \\
 & R_5 = 0,824 \quad \text{factor } 1 + e_5; R^2 = 0,680 \\
 & R_6 = 0,774 \quad \text{factor } 1 + e_6; R^2 = 0,600 \\
 & R_7 = 0,548 \quad \text{factor } 1 + e_7; R^2 = 0,686 \\
 & R_8 = 0,548 \quad \text{factor } 1 + e_8; R^2 = 0,660 \\
 & R_9 = 0,831 \quad \text{factor } 1 + e_9; R^2 = 0,690 \\
 & R_{10} = 0,847 \quad \text{factor } 1 + e_{10}; R^2 = 0,717 \\
 & R_{11} = 0,864 \quad \text{factor } 1 + e_{11}; R^2 = 0,747 \\
 & R_{12} = 0,867 \quad \text{factor } 1 + e_{12}; R^2 = 0,751 \\
 & R_{13} = 0,809 \quad \text{factor } 1 + e_{13}; R^2 = 0,655 \\
 & R_{14} = 0,825 \quad \text{factor } 1 + e_{14}; R^2 = 0,681 \\
 & R_{15} = 0,753 \quad \text{factor } 1 + e_{15}; R^2 = 0,567 \\
 & R_{16} = 0,801 \quad \text{factor } 1 + e_{16}; R^2 = 0,641 \\
 & R_{17} = 0,815 \quad \text{factor } 1 + e_{17}; R^2 = 0,665 \\
 & R_{18} = 0,836 \quad \text{factor } 1 + e_{18}; R^2 = 0,699 \\
 & R_{19} = 0,548 \quad \text{factor } 1 + 0,04 \quad \text{factor } 2 + 0,311 \quad \text{factor } 3 + 0,144 \quad \text{factor } 4 - 0,114 \quad \text{factor } 5 + e_{19}; R^2 = 0,809
 \end{aligned}$$

Figura 5.19. Ecuaciones modelo 5 datos del grupo A

A continuación, se muestran los parámetros estimados en el caso del modelo 4, pudiéndose comprobar que no presenta diferencias apreciables con el modelo 5.

Tabla 5.33. Descripción del modelo 4

Nº de variables en el modelo	43
Nº de variables exógenas	24
Nº de variables endógenas	19
Nº de varianzas y covarianzas de origen	190
Nº de parámetros a estimar	48
Grados de Libertad	142

En este caso el haber eliminado los efectos que relacionan el ítem R19 con los factores 2, 3, 4 y 5, el número de parámetros a estimar disminuye de 52 a 48 y los grados de libertad aumentan de 138 a 142.

La solución estandarizada se muestra en la figura o path diagram 5.20, pudiéndose consultar la solución no estandarizada en el anexo 5.

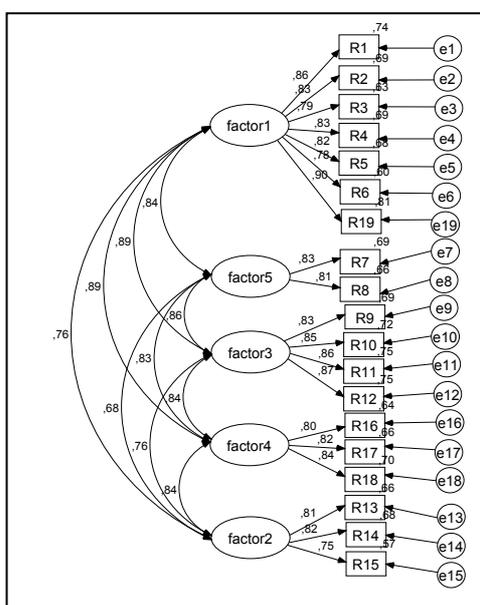


Figura 5.20. Solución estandarizada del modelo 4

Los coeficientes de regresión estandarizados, se muestran en la tabla 5.34. En este caso, todos los efectos estandarizados obtenidos tienen valores positivos y superiores a 0,7, obteniéndose incluso un valor de 0,9 entre R19 y el factor 1. El valor obtenido para las correlaciones entre los factores y las correlaciones al cuadrado se muestran en las tablas 5.35 y 5.36, pudiendo observarse de nuevo que todos los factores están altamente correlacionados entre sí. Además se detecta que el valor de los coeficientes obtenido entre el factor uno y el resto de los factores son ligeramente superiores que los obtenidos en el modelo 5.

Tabla 5.34. Coeficientes de regresión estandarizados del modelo 4

Coef. Regresión	Estimado	Coef. Regresión	Estimado
R14← factor2	0,824	R10← factor3	0,848
R15← factor2	0,753	R11← factor3	0,864
R13← factor2	0,811	R12← factor3	0,865
R6← factor1	0,777	R7← factor5	0,829
R5← factor1	0,822	R8← factor5	0,812
R4← factor1	0,832	R19← factor1	0,900
R3← factor1	0,794	R16← factor4	0,801
R2← factor1	0,830	R18← factor4	0,836
R1← factor1	0,859	R17← factor4	0,815
R9← factor3	0,833		

Tabla 5.35. Correlaciones obtenidas en el modelo 4

CORRELACIÓN	ESTIMADO	CORRELACIÓN	ESTIMADO
factor 2 ↔ factor 1	0,762	factor 3 ↔ factor 4	0,842
factor 1 ↔ factor 5	0,844	factor 5 ↔ factor 4	0,832
factor 1 ↔ factor 3	0,891	factor 2 ↔ factor 4	0,835
factor 3 ↔ factor 5	0,860	factor 2 ↔ factor 3	0,759
factor 1 ↔ factor 4	0,893	factor 2 ↔ factor 5	0,676

Tabla 5.36. Correlaciones al cuadrado obtenidas en el modelo 4

SQUARED MÚLTIPLE CORRELATIONS	ESTIMADO	SQUARED MÚLTIPLE CORRELATIONS	ESTIMADO
R19	0,809	R9	0,693
R13	0,657	R8	0,659
R15	0,567	R7	0,687
R14	0,679	R6	0,603
R18	0,698	R5	0,676
R17	0,665	R4	0,692
R16	0,642	R3	0,631
R12	0,749	R2	0,689
R11	0,746	R1	0,738
R10	0,719		

Es decir, que la expresión del modelo quedaría como se muestra a continuación.

$$\begin{aligned} \text{factor } 1 &= 0,762 * \text{factor } 2 + 0,891 * \text{factor } 3 + 0,893 * \text{factor } 4 + 0,844 * \text{factor } 5 + d_1 \\ \text{factor } 2 &= 0,762 * \text{factor } 1 + 0,759 * \text{factor } 3 + 0,835 * \text{factor } 4 + 0,676 * \text{factor } 5 + d_2 \\ \text{factor } 3 &= 0,891 * \text{factor } 1 + 0,759 * \text{factor } 2 + 0,842 * \text{factor } 4 + 0,860 * \text{factor } 5 + d_3 \\ \text{factor } 4 &= 0,893 * \text{factor } 1 + 0,835 * \text{factor } 2 + 0,842 * \text{factor } 3 + 0,832 * \text{factor } 5 + d_4 \\ \text{factor } 5 &= 0,844 * \text{factor } 1 + 0,676 * \text{factor } 2 + 0,860 * \text{factor } 3 + 0,832 * \text{factor } 4 + d_5 \end{aligned}$$

$$\begin{aligned} R_1 &= 0,859 \text{ factor } 1 + e_1; R^2 = 0,738 \\ R_2 &= 0,830 \text{ factor } 1 + e_2; R^2 = 0,689 \\ R_3 &= 0,794 \text{ factor } 1 + e_3; R^2 = 0,631 \\ R_4 &= 0,832 \text{ factor } 1 + e_4; R^2 = 0,692 \\ R_5 &= 0,822 \text{ factor } 1 + e_5; R^2 = 0,676 \\ R_6 &= 0,777 \text{ factor } 1 + e_6; R^2 = 0,603 \\ R_7 &= 0,829 \text{ factor } 5 + e_7; R^2 = 0,687 \\ R_8 &= 0,812 \text{ factor } 5 + e_8; R^2 = 0,659 \\ R_9 &= 0,833 \text{ factor } 3 + e_9; R^2 = 0,693 \\ R_{10} &= 0,848 \text{ factor } 3 + e_{10}; R^2 = 0,719 \\ R_{11} &= 0,864 \text{ factor } 3 + e_{11}; R^2 = 0,746 \\ R_{12} &= 0,865 \text{ factor } 3 + e_{12}; R^2 = 0,749 \\ R_{13} &= 0,811 \text{ factor } 2 + e_{13}; R^2 = 0,657 \\ R_{14} &= 0,824 \text{ factor } 2 + e_{14}; R^2 = 0,679 \\ R_{15} &= 0,753 \text{ factor } 2 + e_{15}; R^2 = 0,567 \\ R_{16} &= 0,801 \text{ factor } 4 + e_{16}; R^2 = 0,642 \\ R_{17} &= 0,815 \text{ factor } 4 + e_{17}; R^2 = 0,665 \\ R_{18} &= 0,836 \text{ factor } 4 + e_{18}; R^2 = 0,698 \\ R_{19} &= 0,900 \text{ factor } 1 + e_{19}; R^2 = 0,809 \end{aligned}$$

Figura 5.21. Ecuaciones modelo 4 de los datos del grupo A

Las ecuaciones que relacionan a los factores entre sí, muestran que los efectos de todos ellos son positivos y significativos. El efecto del factor 1 sobre la “consideración general de buen profesor” (ítem R19), explica un 80,9 % de la varianza del mismo, porcentaje similar al modelo anterior, lo que nos indicaría que el efecto del resto de factores sobre este ítem es prácticamente despreciable.

En resumen podemos concluir que los modelo 4 y 5, con cinco dimensiones latentes correlacionadas entre sí, son los que mejor representaría la estructura subyacente

del cuestionario, coincidiendo con los resultados obtenidos en el análisis factorial exploratorio.

- *Modelos del Grupo B*

En el caso de que los alumnos no hubiesen sido evaluados antes de contestar al cuestionario de evaluación docente, los modelos que mejores ajustes presentan son el 3 y 4, que parten igual que en el caso anterior, del supuesto de correlación entre las variables latentes o factores. A continuación se muestran las ecuaciones obtenidas para cada uno de ellos. Los valores completos de los parámetros estimados y los path diagrams de cada uno de ellos pueden consultarse en el anexo 5.

En el caso del modelo 4, las ecuaciones que relacionan a los factores entre sí, muestran que los efectos de todos ellos son positivos y significativos. El efecto del factor 1 sobre la “*consideración general de buen profesor*” (ítem R19 o ítem criterio), tiene mayor correlación que cualquiera de los otros, siendo la de los factores 3 y 4 prácticamente despreciable.

En el caso del modelo 3, el efecto del factor 1 sobre la “*consideración general de buen profesor*” (ítem R19), explica un 81,2 % de la varianza del mismo, pudiendo concluir que el modelo 3, con cuatro dimensiones latentes correlacionadas entre sí, es el que mejor representaría la estructura subyacente del cuestionario, coincidiendo con los resultados obtenidos en el análisis factorial exploratorio.

$$\begin{aligned} \text{factor } 1 &= 0,863 * \text{factor } 2 + 0,866 * \text{factor } 3 + 0,818 * \text{factor } 4 + d_1 \\ \text{factor } 2 &= 0,863 * \text{factor } 1 + 0,821 * \text{factor } 3 + 0,850 * \text{factor } 4 + d_2 \\ \text{factor } 3 &= 0,866 * \text{factor } 1 + 0,821 * \text{factor } 2 + 0,803 * \text{factor } 4 + d_3 \\ \text{factor } 4 &= 0,818 * \text{factor } 1 + 0,850 * \text{factor } 2 + 0,803 * \text{factor } 3 + d_4 \end{aligned}$$

$$\begin{aligned} R_1 &= 0,857 \text{ factor } 1 + e_1; R^2 = 0,738 \\ R_2 &= 0,810 \text{ factor } 1 + e_2; R^2 = 0,689 \\ R_3 &= 0,760 \text{ factor } 1 + e_3; R^2 = 0,631 \\ R_4 &= 0,796 \text{ factor } 1 + e_4; R^2 = 0,634 \\ R_5 &= 0,797 \text{ factor } 1 + e_5; R^2 = 0,635 \\ R_6 &= 0,729 \text{ factor } 1 + e_6; R^2 = 0,531 \\ R_7 &= 0,790 \text{ factor } 4 + e_7; R^2 = 0,624 \\ R_8 &= 0,778 \text{ factor } 4 + e_8; R^2 = 0,606 \\ R_9 &= 0,810 \text{ factor } 2 + e_9; R^2 = 0,656 \\ R_{10} &= 0,820 \text{ factor } 2 + e_{10}; R^2 = 0,672 \\ R_{11} &= 0,833 \text{ factor } 2 + e_{11}; R^2 = 0,693 \\ R_{12} &= 0,847 \text{ factor } 2 + e_{12}; R^2 = 0,717 \\ R_{16} &= 0,774 \text{ factor } 3 + e_{16}; R^2 = 0,600 \\ R_{17} &= 0,801 \text{ factor } 3 + e_{17}; R^2 = 0,641 \\ R_{18} &= 0,779 \text{ factor } 3 + e_{18}; R^2 = 0,606 \\ R_{19} &= 0,646 \text{ factor } 1 + 0,293 \text{ factor } 2 + 0,068 \text{ factor } 3 - 0,086 \text{ factor } 4 + e_{19}; R^2 = 0,807 \end{aligned}$$

Figura 5.21. Ecuaciones modelo 4 datos del grupo B

$$\begin{aligned} \text{factor } 1 &= 0,885 * \text{factor } 2 + 0,876 * \text{factor } 3 + 0,825 * \text{factor } 4 + d_1 \\ \text{factor } 2 &= 0,885 * \text{factor } 1 + 0,821 * \text{factor } 3 + 0,850 * \text{factor } 4 + d_2 \\ \text{factor } 3 &= 0,876 * \text{factor } 1 + 0,821 * \text{factor } 2 + 0,803 * \text{factor } 4 + d_3 \\ \text{factor } 4 &= 0,825 * \text{factor } 1 + 0,850 * \text{factor } 2 + 0,803 * \text{factor } 3 + d_4 \end{aligned}$$

$$\begin{aligned} R_1 &= 0,849 \text{ factor } 1 + e_1; R^2 = 0,721 \\ R_2 &= 0,802 \text{ factor } 1 + e_2; R^2 = 0,644 \\ R_3 &= 0,756 \text{ factor } 1 + e_3; R^2 = 0,572 \\ R_4 &= 0,796 \text{ factor } 1 + e_4; R^2 = 0,629 \\ R_5 &= 0,793 \text{ factor } 1 + e_5; R^2 = 0,634 \\ R_6 &= 0,796 \text{ factor } 1 + e_6; R^2 = 0,529 \\ R_7 &= 0,790 \text{ factor } 4 + e_7; R^2 = 0,624 \\ R_8 &= 0,778 \text{ factor } 4 + e_8; R^2 = 0,606 \\ R_9 &= 0,812 \text{ factor } 2 + e_9; R^2 = 0,659 \\ R_{10} &= 0,821 \text{ factor } 2 + e_{10}; R^2 = 0,674 \\ R_{11} &= 0,832 \text{ factor } 2 + e_{11}; R^2 = 0,692 \\ R_{12} &= 0,845 \text{ factor } 2 + e_{12}; R^2 = 0,714 \\ R_{16} &= 0,774 \text{ factor } 3 + e_{16}; R^2 = 0,600 \\ R_{17} &= 0,801 \text{ factor } 3 + e_{17}; R^2 = 0,641 \\ R_{18} &= 0,779 \text{ factor } 3 + e_{18}; R^2 = 0,606 \\ R_{19} &= 0,901 \text{ factor } 1 + E_{19}; R^2 = 0,812 \end{aligned}$$

Figura 5.22. Ecuaciones modelo 3 datos del grupo B

5.4. Conclusiones

Evaluando los resultados obtenidos en cada uno de los análisis anteriores y teniendo en cuenta que la finalidad de este trabajo es proponer una metodología para el análisis de los cuestionarios de evaluación docente lo más sencilla y eficaz posible, podemos concluir que:

- En primer lugar, respecto los estudios de fiabilidad y validez, la aplicación de la metodología del análisis factorial comprueba y supera satisfactoriamente todos los tipos de análisis sobre la pertinencia, fiabilidad y validez de la matriz de datos
- No existen diferencias significativas al utilizar como matriz de entrada de datos la matriz de correlaciones o la matriz de varianza-covarianza. En ambos casos se han obtenido el mismo número de factores extraídos y prácticamente la misma proporción de varianza explicada. En este sentido y de acuerdo con Hair et al. (1995) entre otros autores, se propone la utilización de la matriz de correlaciones, ya que, simplifica en cierta manera la interpretación de los resultados al tener un rango común que hace directas las comparaciones de los coeficientes dentro del modelo.
- La utilización del Análisis Factorial mediante el procedimiento de Componentes Principales ha resultado ser una técnica adecuada para evaluar la dimensionalidad del Cuestionario de Evaluación Docente del Profesorado mediante Opinión de los Alumnos de la UPV. Los 19 ítems incluidos en el cuestionario son adecuados para evaluar las dimensiones latentes del mismo. Todas las dimensiones parecen ser relevantes para evaluar la calidad de la docencia universitaria y se hallan altamente correlacionadas y son, en definitiva, complementarias entre sí en la medición del constructo.
- Para facilitar la interpretación de los factores se propone utilizar la rotación oblicua, conocida como *oblimin directo*, con el parámetro $\delta = 0$, ya que se ha comprobado que los factores están correlacionados entre sí.

- Examinando las matrices factoriales obtenidas utilizando los distintos criterios para calcular el número de factores a ser extraídos y, debido a las características de elaboración del cuestionario, el número de factores que mejor representa los datos iniciales sería 5 en los datos del grupo A (Dominio, organización y claridad de la asignatura y del programa; Evaluación; Interrelación profesor-alumno; recursos; interrelación con otras asignaturas) y 4 en el caso de los datos del grupo B (Dominio, organización y claridad de la asignatura y del programa; Interrelación profesor-alumno; recursos; interrelación con otras asignaturas).
- La aplicación de los modelos de ecuaciones estructurales con fines confirmatorios, nos ha permitido comparar distintos posibles modelos de la estructura del CEDA. Los modelos que parten del supuesto de correlación entre los factores o dimensiones latentes, son más plausibles que los que parten del supuesto de incorrelación. Por tanto, se confirma así la multidimensionalidad del constructo medido, cuyas dimensiones se hallan fuertemente relacionadas entre sí.
- En el caso de los datos del grupo A, los modelos 4 y 5, ambos con cinco variables latentes correlacionadas, son los que obtienen mejores valores de ajuste, proponiéndose como los modelos teóricos más plausibles en la explicación de la estructura del constructo medido a través del Cuestionario de Evaluación docente de la UPV. En el caso de los datos del grupo B, los modelos 3 y 4, ambos con cuatro variables latentes correlacionadas, son los que obtienen mejores valores de ajuste. En consecuencia, se cuestiona la existencia de un solo factor que de explicación a la composición del cuestionario, es decir, se cuestiona la conveniencia de obtener una única puntuación o índice general de la calidad docente.

6. Obtención de Grupos Homogéneos

6.1. Introducción

Una comprobación básica que afecta a la utilización del CEDA es determinar si permite diferenciar tipos de profesores en función de la satisfacción que muestran los estudiantes. En este sentido, con la finalidad de poder establecer perfiles o tipologías de calidad docente en la UPV, a partir de los resultados obtenidos mediante el cuestionario de evaluación docente, se decidió realizar un análisis de conglomerados, a partir de las medias de los ítems incluidos en el cuestionario, estimadas para cada posible combinación profesor-asignatura. Una vez identificados los perfiles, se intentó evaluar la relación existente entre los conglomerados definidos y la estructura subyacente existente en los datos. Para ello, se analizó, en primer lugar, un conjunto de características descriptivas adicionales referentes al profesor y asignatura, de manera que podamos determinar qué factores contribuyen a caracterizarlos y, en consecuencia, obtener referencias sobre las cualidades de los profesores asignados a cada uno de los perfiles.

Finalmente, se pretende efectuar un análisis discriminante (LDA) para identificar los ítems del cuestionario que mejor nos permiten discernir entre los distintos grupos obtenidos, de manera que podamos interpretar mejor las diferencias existentes entre los distintos perfiles y determinar una función para poder pronosticar la futura clasificación de nuevos individuos.

6.2. Metodología

Para lograr la consecución de los objetivos comentados se utilizó el análisis clúster, también denominado de conglomerados, construcción de tipologías, análisis de clasificación o taxonomía numérica. Este análisis es un grupo de técnicas multivariantes cuyo principal objeto es agrupar observaciones basándose en las características que poseen,

de manera que cada objeto u observación es muy parecido a los que hay en su conglomerado con respecto a algún criterio de selección predeterminado.

Una vez determinados los grupos homogéneos y la clasificación de las observaciones, se intentó caracterizar a los mismos en función de una serie de variables descriptivas adicionales referentes a la asignatura y al profesor.

Finalmente, se efectuó un análisis discriminante con la finalidad de identificar qué variables o ítems incluidos en el cuestionario discriminan mejor entre los grupos definidos a priori y establecer el número y la composición de las dimensiones de discriminación entre los grupos formados a partir del conjunto de variables independientes.

6.2.1.- Definición y diseño de la investigación

a) Definición de la Investigación

Con la finalidad de determinar los perfiles o tipologías del profesorado a partir de la calidad percibida por los estudiantes mediante el CEDA, se partió de los valores medios obtenidos para cada posible combinación profesor asignatura, seleccionándose los 16 ítems incluidos en el cuestionario, sin considerar los referentes a evaluación, como las variables a utilizar para caracterizar los objetos a agrupar. De esta manera, como ya se ha comentado, realizando el análisis clúster sobre estos valores medios, podremos obtener como resultados grupos homogéneos de calidad docente, en lugar de las tipologías en las respuestas de los alumnos, que obtendríamos si trabajásemos con los datos individuales.

En principio, todas las variables se consideraron relevantes para abordar el estudio. Tras la realización del análisis clúster se evaluó la relevancia o irrelevancia de cada una de ellas, procediendo a la eliminación de aquellas que se consideró conveniente, ya que la inclusión de alguna variable no relevante aumentaría la posibilidad de crear atípicos, cosa que podría afectar considerablemente los resultados obtenidos.

Posteriormente se procedió a incluir información relativa al profesor (edad, categoría docente, tiempo de dedicación, si es doctor o no) y de la asignatura (departamento, titulación, curso, semestre) para cada una de las observaciones. Esta información es necesaria para poder caracterizar cada uno de los clusters extraídos. La creación de una base de datos de carácter longitudinal, que abarcara diversos periodos evaluativos fue un proceso complejo, debido al elevado dinamismo que se produce cada año, tanto por la modificación del personal docente como por las variaciones en las asignaturas que implanta cada profesor. Además, nos encontramos con el inconveniente de que las bases de datos existentes en la universidad no siguen los mismos criterios al introducir nuevos datos o están incompletas y, en ocasiones, no resulta posible cruzarlas. En consecuencia, únicamente se pudo obtener información complementaria para 541 de las posibles combinaciones profesor-asignatura de las 5760 correspondientes al curso académico 2000-2001 y 2142 de las 5048 observaciones correspondientes al curso 1999-2000. Por ello, se decidió comenzar a trabajar con los datos del curso 1999-2000 y utilizar posteriormente los del 2000-2001 para verificar que las tipologías de calidad docente determinadas se mantienen a lo largo del tiempo.

Como ya se ha comentado, con este estudio no se pretende obtener resultados definitivos en la caracterización de perfiles de calidad docente, sino definir una metodología estadística que permita, en un futuro, trabajando con bases de datos completas, optimizar la información contenida en los cuestionarios de opinión.

b) Tratamientos previos

Esta etapa consiste, por un lado, en analizar si existen atípicos en las observaciones y eliminarlos si es posible y, por otro lado, en definir si se deben o no estandarizar los datos.

Los atípicos son aquellos objetos u observaciones que son muy diferentes al resto, que pueden ser tanto observaciones no representativas de la población en general, como un conjunto reducido de la población que produce una mala representación del grupo de la muestra. Los atípicos pueden clasificarse en datos aberrantes y datos extremos. El análisis

clúster es muy sensible a los datos extremos en el caso de trabajar con datos individuales y tamaños de muestra muy pequeños, ya que pueden provocar que los conglomerados obtenidos no sean representativos de la verdadera estructura de la población, por ello es necesario un análisis preliminar para detectarlos.

La manera más sencilla de efectuar un estudio preliminar de los atípicos es mediante un diagrama de perfil gráfico, en el que se representan todas las variables en el eje horizontal y los valores de las variables en el eje vertical. Cada punto del gráfico representa el valor de la correspondiente variable y los puntos están conectados mediante líneas para facilitar la interpretación. Sin embargo, debido a que en nuestro estudio el número de casos es muy elevado (superior a 500 observaciones), la interpretación de estos perfiles resulta compleja. Por otro lado, en nuestro estudio trabajaremos con valores medios, con lo que los valores extremos pierden relevancia y pueden ser detectados en el propio proceso de clasificación.

Respecto a los datos aberrantes, hasta el momento no se realizado ningún proceso de detección, ya que resulta muy complejo definir que observaciones son aberrantes y en función de qué criterios, quedando esta actuación propuesta como una línea futura de investigación.

Por otro lado la utilización de datos no estandarizados, puede conllevar inconsistencia entre las soluciones clúster cuando la escala de las distintas variables es diferente. En nuestro estudio, la estandarización de las variables, obtenida restando la media y dividiendo por la desviación típica de cada variable, no fue necesaria en este estudio, al estar todas las variables medidas en la misma escala tipo Likert. Finalmente, también puede ser necesaria la estandarización de encuestados u observaciones, en caso de que se desee analizar la importancia relativa de una variable u otra en los conglomerados obtenidos, eliminándose así los efectos de tipo respuesta o pautas sistemáticas de respuesta, tales como los que siempre dan respuestas favorables en todas las cuestiones o los que siempre dan respuestas desfavorables. La estandarización por encuestado, también denominada tipificación centrada por filas, estandarizaría cada cuestión no por la media de la muestra, sino por la puntuación media del encuestado y en nuestro caso tampoco ha sido efectuada, ya que en el proceso de introducción de los datos de las encuestas, son detectadas y

eliminadas todas aquellas que pudieran considerarse mal cumplimentadas, no pudiendo considerar como pauta sistemática de respuesta el que todos los alumnos puntúen alto o bajo en la gran mayoría de los ítems.

6.2.2.- Seleccionar medidas de similitud

La similitud es un concepto fundamental en el análisis clúster. La similitud entre objetos es una medida de correspondencia o parecido entre objetos que van a ser agrupados. En este caso, una vez especificadas las características que van a definir la similitud, se combinan todas ellas en una medida de similitud calculada para todos los pares de objetos, al igual que utilizábamos las correlaciones entre variables en el análisis factorial. De esta forma cualquier objeto puede ser comparado con cualquier otro a través de esa medida de similitud.

Existen distintas maneras de medir la similitud entre objetos en el análisis clúster: medidas de correlación, medidas de similitud de distancia y medidas de asociación.

Las medidas de correlación representan la similitud mediante la correspondencia de patrones en las características. Son poco utilizadas en el análisis clúster, ya que por lo general el interés de este análisis suele radicar en las magnitudes de los objetos y no en los patrones de valores.

Las medidas de asociación comparan objetos cuyas características se suelen medir en términos no métricos (nominal u ordinal). Sin embargo los programas informáticos dan poco apoyo a las medidas de asociación, no estando aún bien defendido cuando tenemos variables con varias categorías.

Las medidas de similitud de distancia representan la similitud como proximidad de cada observación respecto a las otras para las variables del análisis clúster. Son las más utilizadas en esta técnica multivariante. Existen varias tipos de medidas de distancia siendo la más utilizada la distancia euclídea simple o la distancia euclídea al cuadrado. De las posibles formas de medir la similitud entre objetos, en nuestro estudio se han utilizado las medidas de distancia, que representan la similitud como proximidad entre observaciones.

Estas medidas de distancia son en realidad medidas de diferencia, ya que valores elevados indican una menor similitud. Dados dos objetos i y j , en los que se han medido p variables, existen distintos posibles tipos de distancia que definimos a continuación:

- *Euclídea simple* o al cuadrado que presenta la ventaja frente a la primera de no tener que estimar la raíz, acelerando los cálculos. Es la distancia recomendada para los métodos de análisis clúster del centroide y Ward (Hair et al., 1995). La distancia euclídea simple viene definida por:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

- *Del máximo*: es la distancia máxima entre dos componentes de x e y . Su utilización requiere que las variables estén medidas en las mismas unidades.

$$d_{ij} = \max_k |x_{ik} - x_{jk}|$$

donde $k = 1, \dots, p$.

- *De Manhattan*: es la distancia absoluta entre los dos vectores. Presenta problemas si existe colinealidad entre las variables.

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|$$

donde $k = 1, \dots, p$.

- *De Canberra*:

$$d_{ij} = \sum_{k=1}^p \left| \frac{x_{ik} - x_{jk}}{x_{ik} + x_{jk}} \right|$$

En este caso, los términos con cero en el numerador y en el denominador se omiten de la suma y se tratan como valores perdidos.

- *Distancia de Mahalanobis* (D^2), que puede expresarse como:

$$d_{ij} = (\bar{x}_i - \bar{x}_j)' \hat{\Sigma}^{-1} (\bar{x}_i - \bar{x}_j)$$

donde \bar{x}_i y \bar{x}_j son vectores de p observaciones para cada sujeto y $\hat{\Sigma}$ es la matriz de varianzas-covarianzas de las variables consideradas.

6.2.3. Supuestos del análisis clúster

El análisis clúster, no es una técnica de inferencia estadística en la que se analizan los parámetros de una muestra en la medida en la que pueden ser representativos de una población. Por el contrario, el análisis clúster, es una metodología objetiva de cuantificación de las características estructurales de un conjunto de observaciones, y en consecuencia, hipótesis como la normalidad, linealidad y homocedasticidad son poco relevantes en este caso, siendo los aspectos más críticos a considerar la representatividad de la muestra y la existencia de multicolinealidad (Hair et al., 1995).

a) Representatividad de la muestra

El análisis clúster es sólo tan bueno como lo sea la representatividad de la muestra. Por consiguiente, es esencial trabajar con una muestra representativa para asegurar que los resultados son generalizables para la población a estudiar.

b) Impacto de la multicolinealidad

Como ya se ha comentado en el análisis factorial, el efecto de la multicolinealidad es enmascarar el impacto verdadero de las variables. Sin embargo, en el análisis clúster el efecto es ligeramente diferente, ya que las variables multicolineales están implícitamente ponderadas con más fuerza. La multicolinealidad actúa en este caso como proceso de ponderación no aparente para el observador, pero que sin embargo, afecta al análisis. Por

ello, puede ser adecuado si existe una multicolinealidad substancial entre ciertas variables, reducir el número de las mismas. En este sentido, puede plantearse la posibilidad de trabajar con las puntuaciones factoriales obtenidas en el capítulo anterior, sin embargo, como establece Groth, (1997) las variables que verdaderamente discriminan entre grupos subyacentes no están bien representadas en la mayoría de las soluciones factoriales, por tanto, al trabajar con ellas es bastante posible que se obtenga una mala representación de la verdadera estructura de los datos. Es necesario considerar tanto la multicolinealidad entre variables como la discriminabilidad de las mismas para alcanzar la mejor representación de la estructura.

6.2.4. Selección del algoritmo para la obtención de conglomerados

Existen distintos procedimientos para obtener los conglomerados, aunque todos ellos se basan en maximizar las diferencias entre los conglomerados relativas a la variación dentro de los conglomerados. La razón entre la variación entre conglomerados y la media de variación dentro de los conglomerados es estimada mediante una razón F aproximada a la del análisis de la varianza.

Los procedimientos de obtención de los conglomerados pueden clasificarse en jerárquicos, no jerárquicos y el denominado two-step. Los primeros consisten en la construcción de una estructura en forma de árbol, pudiendo actuar de manera aglomerativa o divisiva, mientras que los procedimientos no jerárquicos asignan los objetos a los conglomerados una vez que se ha especificado el número a formar de éstos. El procedimiento two-step consiste en dos etapas consecutivas: en la primera de ellas, denominada pre-cluster, se agrupan los datos originales en una serie de subclusters; en la segunda etapa, denominada clúster, los subclusters formados durante la primera fase se agrupan mediante un procedimiento jerárquico.

En nuestro caso, se aplicó en primer lugar, un análisis jerárquico en una fase exploratoria para determinar el número de clusters a formar, ya que este procedimiento presenta la ventaja de no requerir un conocimiento previo de los mismos. A continuación la

solución obtenida ha sido sometida a un análisis no jerárquico con la finalidad de refinarla, ya que el procedimiento jerárquico presenta el inconveniente de que una vez que las observaciones son asignadas a un clúster determinado no pueden ser reasignadas a otro.

La solución obtenida utilizando los procedimientos jerárquicos y no jerárquicos como complementarios, ha sido comparada con la proporcionada mediante el procedimiento two-step.

6.2.4.1. Metodología del análisis clúster jerárquico.

Los procedimientos jerárquicos consisten en la construcción de una estructura en forma de árbol. Existen dos tipos básicos para la obtención de los conglomerados:

- **Procedimientos de aglomeración**, en los que cada observación empieza dentro de su propio conglomerado, combinándose los conglomerados más cercanos para crear uno nuevo y reduciéndose así el número de conglomerados en cada paso. Son los más utilizados.
- **Procedimientos de división**, en los que se procede en dirección opuesta a los procedimientos de aglomeración. Se parte de un gran conglomerado que contiene todas las observaciones y éste se irá dividiendo en conglomerados más pequeños, hasta que cada observación constituye un conglomerado por sí misma.

Dentro de los métodos aglomerativos existen cinco posibles algoritmos para desarrollar los conglomerados: método de encadenamiento simple, método de encadenamiento completo, método de encadenamiento medio, método de Ward y método del centroide.

El método de encadenamiento simple, también conocido como el del vecino más cercano, se basa en la distancia mínima, colocando en el primer conglomerado los dos objetos separados por la distancia más corta y repitiendo lo mismo hasta que todos los objetos están en un único conglomerado. Presenta el inconveniente de que si los

conglomerados están mal definidos, se forman largas cadenas y todos los individuos pueden situarse en una cadena, pudiendo ser las observaciones extremas muy diferentes.

El procedimiento de encadenamiento completo es similar al anterior pero el criterio de aglomeración se basa en la distancia máxima. Esta distancia máxima entre individuos de cada conglomerado representa la esfera de diámetro mínimo que puede incluir todas las observaciones de ambos conglomerados. Todos los objetos de un conglomerado quedan vinculados con el resto a alguna distancia máxima o por la mínima similitud.

El método de encadenamiento medio comienza de manera similar a los anteriores, pero el criterio de aglomeración es la distancia media de todos los individuos de un conglomerado con todos los de otro. No depende de valores extremos como sucede en los procedimientos anteriores y la partición se basa en todos los miembros de los conglomerados. Presenta el inconveniente de tender a combinar conglomerados con variaciones reducidas dentro de ellos y además tiende a producir conglomerados con aproximadamente la misma varianza.

El método de Ward utiliza como distancia entre conglomerados la suma de los cuadrados entre dos conglomerados sumados para todas las variables. La distancia entre dos grupos es la suma de cuadrados *ANOVA* entre los mismos (acumulada sobre todas las variables). En cada etapa, la suma de cuadrados dentro de cada grupo se minimiza sobre todas las particiones obtenibles juntando pares de grupos de la etapa previa. Las sumas de cuadrados divididas por la suma de cuadrados total para obtener proporciones de varianza son fácilmente interpretables. El método de Ward junta grupos maximizando la verosimilitud bajo las siguientes hipótesis: distancia normal multivariante; matrices de covarianza esféricas iguales; probabilidades de muestreo iguales. El método de Ward tiende a juntar grupos con pocas observaciones, de forma que produce grupos de tamaño parecido. Es muy sensible a los valores atípicos.

El método del centroide utiliza la distancia (normalmente euclídea simple o al cuadrado) entre sus centroides. Los centroides de grupo son los valores medios de las variables en el valor teórico del conglomerado. Los centroides de grupo van cambiando a

medida que se fusionan conglomerados. Es más robusto a los atípicos que la mayoría de los otros métodos.

Por consideraciones que se comentarán posteriormente, en nuestro caso el procedimiento jerárquico utilizado ha sido el método de Ward, por lo que lo desarrollamos más extensamente a continuación. Este procedimiento utiliza una estrategia de agregación basada en la varianza. Puede ser presentado como una generalización multidimensional del modelo del Análisis de la Varianza (Roux, 1993):

$$SCD_{tot} = SCD_{res} + SCD_{fac} \quad (1)$$

donde,

- SCD_{tot} : suma del cuadrado de las diferencias de las observaciones y la media general.
- SCD_{res} : suma del cuadrado de las diferencias de las observaciones en cada grupo, con respecto a la media del grupo para todos los grupos.
- SCD_{fac} : suma del cuadrado de los diferencias de las observaciones en cada grupo, con respecto a la media general.

Dada una matriz de n observaciones y una sola variable, si se distinguen k grupos en las observaciones, podemos escribir la expresión (1) como:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} (x_i - \bar{x}_k)^2 + \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^2 \quad (2)$$

Es decir, que la inercia total es descompuesta en una suma de inercia intraclases y de la inercia interclases.

Si se reemplazan en (2) las diferencias con respecto a la media por el cuadrado de las distancias euclídeas con respecto al centro de gravedad, G, obtendremos:

$$\sum_{i=1}^n d^2(i, G) = \sum_{k=1}^K \sum_{i=1}^{n_k} d^2(i, G) + \sum_{k=1}^K n_k d^2(G_k, G) \quad (3)$$

siendo :

- Coordenadas de $G: (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)$, donde el término general es:

$$\bar{x}_p = \frac{1}{n} \sum_{i=1}^n x_{ip}$$

- Coordenadas de $G_k: (\bar{X}_{k1}, \bar{X}_{k2}, \dots, \bar{X}_{kp})$, con el término general: $\bar{x}_{kp} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ip}$

- Coordenadas de $i: (\bar{X}_{i1}, \bar{X}_{i2}, \dots, \bar{X}_{ip})$

- Distancias euclídeas:

$$d^2(i, G) = \sum_{p=1}^P (x_{ip} - \bar{x}_p)^2 \forall i = 1, \dots, i; \forall p = 1, \dots, p$$

$$d^2(i, G_k) = \sum_{p=1}^P (x_{ip}^k - \bar{x}_{kp})^2 \forall i = 1, \dots, i; \forall p = 1, \dots, p; \forall k = 1, \dots, k$$

$$d^2(G_k, G) = \sum_{p=1}^P (x_{kp}^k - \bar{x}_p)^2; \forall p = 1, \dots, p; \forall k = 1, \dots, k$$

Reemplazando estos valores en (3) obtenemos la expresión siguiente (4):

$$\sum_{i=1}^n \sum_{p=1}^P (X_{ip} - \bar{x}_p)^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{p=1}^P (x_{ip}^k - \bar{x}_{kp})^2 + \sum_{k=1}^K \sum_{p=1}^P n_k (\bar{x}_{kp} - \bar{x}_p)^2 \quad (4)$$

Si las observaciones se presentan en grupos bien diferenciados, las inercias intragrupos deben ser bajas y las inercias intergrupos deben ser elevadas.

6.2.4.2. Metodología del análisis clúster no jerárquico.

Los procedimientos no jerárquicos trabajan asignando los objetos a los conglomerados una vez que el número de conglomerados a formar ha sido especificado, en lugar de construir estructuras en forma de árbol como trabajaban los jerárquicos. Para ello, en primer lugar, debe seleccionarse una semilla de conglomerado como centro inicial del mismo y todas las observaciones dentro de una distancia umbral previamente especificada

se incluyen dentro del conglomerado resultante. A continuación se escoge otra semilla de conglomerado y la asignación continúa hasta que todos los objetos están asignados. Finalmente, los objetos pueden reasignarse a otros conglomerados más cercanos distintos del inicial.

El método no jerárquico más difundido es el denominado k-means propuesto por McQueen (1967), en el que se divide el conjunto de individuos en k conglomerados, de forma que al final del proceso cada individuo pertenece al clúster cuyo centro está más cercano a él. La distancia euclídea es la más utilizada para establecer la proximidad entre cada individuo y cada clúster. Se pueden utilizar distintos procedimientos para asignar las observaciones a los distintos conglomerados:

- **Procedimiento de umbral secuencial**, que empieza seleccionando una semilla de conglomerado e incluye todos los objetos que caen dentro de una distancia previamente especificada. Cuando todos los objetos dentro de la distancia están incluidos se selecciona una segunda semilla de conglomerado y se incluyen todos los objetos dentro de la distancia previamente especificada y, así el proceso continúa hasta asignar todas las observaciones. Cuando un objeto se incluye en un conglomerado con una semilla, no se considerará a efectos de semillas posteriores
- **Procedimiento de umbral paralelo**, que selecciona varias semillas de conglomerado simultáneamente, y va asignando los objetos dentro de la distancia umbral hasta la semilla más cercana. Conforme se van asignando observaciones, pueden modificarse la distancia umbral para incluir más o menos objetos en los conglomerados. Algunas observaciones pueden quedar sin asignarse a ningún conglomerado si están fuera de la distancia umbral especificada.
- **Procedimiento de optimización**, que es similar a los otros anteriores pero permite reasignar las observaciones, si se acercan más a otro conglomerado distinto del que tiene asignado en ese momento.

En nuestro caso el procedimiento utilizado ha sido el del umbral paralelo, en el que tras especificar el número de conglomerados deseado, estimado previamente tras efectuar el análisis jerárquico, se seleccionan los puntos semilla aleatoriamente de las observaciones originales.

El algoritmo general utilizado en el procedimiento de k-means se describe a continuación.

1. Se seleccionan k centroides o semillas iniciales de clusters, siendo k el número de clusters deseado.
2. Se calculan las distancias euclídeas de cada elemento al centroide de cada grupo, asignándose al más próximo.
3. La asignación se repite secuencialmente y al introducir un nuevo elemento en un grupo se recalculan las coordenadas de la nueva medida de grupo.
4. Cada observación debe ser reasignada al grupo más próximo.
5. El proceso se repite hasta que no se produce ninguna reasignación más de las observaciones.

Como inconvenientes, hay que destacar que el resultado final depende del valor de k , aunque el mayor problema está relacionado con la inicialización de los centros, ya que es fácil comprobar que el resultado está sesgado por éstos. Aunque los centros iniciales pueden seleccionarse aleatoriamente, si éstos se seleccionan de manera que se dispersen uniformemente sobre el espacio de representación, es más fácil que el resultado final sea mejor, además de acelerar la convergencia.

Los procedimientos no jerárquicos presentan la ventaja frente a los jerárquicos de ser menos susceptibles frente a los datos atípicos, a la medida de distancia utilizada y a la inclusión de variables irrelevantes o inapropiadas. Sin embargo, conviene trabajar con puntos de semilla especificados y no aleatorios ya que sino no se garantiza una formación de conglomerados óptima de las observaciones.

6.2.4.3. Metodología del análisis clúster “two-step”

Frente a los métodos tradicionales de análisis clúster, jerárquicos y no jerárquicos, ha surgido recientemente un nuevo algoritmo denominado two-step. Este procedimiento cluster, fue propuesto inicialmente por Zhang et al. (1996) y denominado BIRCH. Posteriormente fue mejorado por Chiu et al. (2001), que incluyó un procedimiento que permite al algoritmo determinar automáticamente el número de conglomerados más adecuado y, además, un pretratamiento de los datos para detectar y eliminar del estudio las observaciones atípicas. La ventaja de esta técnica frente al análisis clúster jerárquico y no jerárquico es que nos permite manejar bases de datos de elevado tamaño y trabajar simultáneamente con variables categóricas y numéricas.

Respecto a la medida de distancia utilizada por esta técnica, no podía utilizarse la distancia euclídea ya que presenta el inconveniente de que no puede emplearse con variables continuas y categóricas simultáneamente (Guha et al., 1999).

En este sentido, Chiu et al. (2001) propusieron una medida de distancia derivada del modelo probabilístico y que se basa en la disminución del valor del logaritmo de la función de máxima verosimilitud que se produce cuando dos clusters se combinan en uno único.

Esta medida puede emplearse con variables continuas y discretas, pero su utilización de esta distancia requiere el cumplimiento de dos principios básicos:

- las variables continuas siguen una distribución normal y las variables discretas una distribución multinomial;
- las variables y observaciones son independientes.

Logaritmo de la función de máxima verosimilitud

Consideremos un conjunto de k clusters, cada uno de ellos con N observaciones. Sea el vector $\{\bar{x}_i, i=1, \dots, N_k\}$, un vector p -dimensional de las N observaciones agrupadas en los k clusters. Representaremos por $C_k = 1, \dots, K$ a cada uno de los clusters.

Obtención de grupos homogéneos

Sea $p(x_i / \theta_k)$ la función de densidad de x en el clúster C_k , donde θ_k es el vector de los parámetros del modelo. Se define la función de clasificación de máxima verosimilitud como la unión de la función de densidad de todas de todas las observaciones agrupadas $\{\bar{x}_i = 1, \dots, N_k; k = 1, \dots, K\}$. El Logaritmo de clasificación de verosimilitud sería de la forma:

$$l = \sum_{k=1}^K \sum_{i \in I_k} \log p(x_i / \theta_k) = \sum_{k=1}^K l_{C_k}$$

donde $l_{C_k} = \sum_{i \in I_k} \log p(x_i / \theta_k)$, representa la contribución del cluster C_k al logaritmo de máxima verosimilitud e $I_K = \{i : x_i \in C_k\}$ es el índice fijado para C_k .

La función de máxima verosimilitud estima $\hat{\theta}_k$ de los θ_k máximos logaritmos de verosimilitud, l . Este valor nos proporciona la mayor probabilidad, en escala logarítmica, de que las agrupaciones de las x_i observaciones sucedan dentro de la función de densidad estimada para el modelo, $p(x / \theta_k)$.

En el caso de tener un conjunto de P^A variables continuas independientes que siguen una distribución normal y un conjunto de P^B variables categóricas independientes que siguen una distribución multinomial en el clúster C_k , podemos representar $\bar{x}_i = \{x_{Ai1}, \dots, x_{AiP^A}, x_{Bi1}, \dots, x_{BiP^B}\}$, donde \bar{x}_{Aip} son variables normales independientes con media μ y varianza σ^2 y \bar{x}_{BiP^B} tiene L_{P^B} categorías y el vector de categorías \bar{x}_{Bip} es independiente multinomial distribuido con el vector de probabilidad $(q_{p1}, \dots, q_{pL_{P^B}})$ donde q_{pl} es la probabilidad de la categoría l , siendo $0 \leq q_{pl} \leq 1$ y $\sum q_{pl} = 1$. La estimación de μ , σ^2 y q_{pl} que maximizan el logaritmo de la función de verosimilitud, constituyen la estimación de máxima verosimilitud.

Si N_k es el número de observaciones agrupadas en el clúster k , la estimación de máxima verosimilitud de la media, μ , para cada variable continua perteneciente a un clúster es la media simple $\hat{\mu} = \sum_{i \in I_p} xp / N_k$ y la varianza, σ^2 , es la varianza simple

$\hat{\sigma}^2 = \sum_{i \in p_j} (x_{Aip} - \hat{\mu}_p)^2 / N_k$ de las observaciones asignadas en el clúster C_k . Si llamamos N_{pl} al número de observaciones asignadas en el clúster C_k cuyas variables categóricas toman las $l = 1, \dots, L_{p^B}$, la estimación de máxima verosimilitud de q_{kpl} es la proporción de $\hat{q}_{pl} = \frac{N_{kpl}}{N_k}$, donde el valor máximo de la función del logaritmo de máxima verosimilitud quedaría:

$$\hat{l}_{C_k} = \sum_{k=1}^K (\hat{l}_{AC_k} + \hat{l}_{BC_k})$$

donde $\hat{l}_{AC_k}, \hat{l}_{BC_k}$ representan la contribución de las variables continuas y categóricas respectivamente, pudiendo expresarse como:

$$\hat{l}_{AC_k} = -\frac{1}{2} N_k \left[P_A \{ \log(2\pi) + 1 \} + \sum_{p=1}^{P_A} \log(\hat{\sigma}^2) \right]$$

$$\hat{l}_{BC_k} = -N \sum_{p=1}^{P_B} \hat{E}_{kp}$$

donde $\hat{E}_{kp} = -\sum_{l=1}^{L_{p^B}} \hat{q}_{kpl} \log \hat{q}_{kpl}$, es la entropía de las k categorías agrupadas en el clúster k .

Medida de Distancia

La medida de distancia entre dos clusters se basa en la disminución que se produce en el logaritmo de máxima verosimilitud al unir dos clusters C_j y C_s , en un nuevo

conglomerado $C_{(j,s)}$. Tras la formación de este nuevo conglomerado, el logaritmo de máxima verosimilitud estimado quedaría:

$$\hat{l}_{nuevo} = \sum_{k \neq j,s}^K \hat{l}_{C_k} + \hat{l}_{C_{(j,s)}}$$

siendo $\hat{l}_{C_{(j,s)}}$ la contribución del nuevo cluster formado.

La disminución que se produce en el logaritmo de máxima verosimilitud después de la formación del nuevo cluster sería $\hat{l} - \hat{l}_{nuevo} = \hat{l}_{C_k} + \hat{l}_{C_s} + \hat{l}_{C_{(j,s)}}$, que después de efectuar las simplificaciones necesarias puede escribirse como:

$$\hat{l} - \hat{l}_{nuevo} = \xi_k + \xi_s + \xi_{(j,s)}$$

donde,

$$\xi_v = -N_v \left\{ \frac{1}{2} \sum_{p=1}^{P_A} \log(\hat{\sigma}_{vp}^2) + \sum_{p=1}^{P_B} \hat{E}_{vp} \right\}$$

para $v = s, j$ y $\xi_{(j,s)}$ se define de manera similar.

En el caso específico de encontrarnos con clusters únicos o cuando los valores de algunos atributos son constantes, es necesario añadir la restricción $x \log(x) = 0$ si $x = 0$

y sustituir la expresión anterior por $\tilde{\xi}_v = -N_v \left\{ \frac{1}{2} \sum_{p=1}^{P_A} \log(\hat{\sigma}_{vp}^2 + \Delta_p) + \sum_{p=1}^{P_B} \hat{E}_{vp} \right\}$, siendo $\Delta_p > 0$, un

escalar positivo.

La distancia entre dos clusters C_j y C_s , queda definida como:

$$d_{(j,s)} = \tilde{\xi}_j + \tilde{\xi}_s - \tilde{\xi}_{(j,s)}$$

y es una medida de distancia asimétrica y no negativa.

Algoritmo

El algoritmo consiste en dos etapas. En la primera de ellas, denominada preclúster, los datos almacenados son escaneados, agrupados en numerosos subclusters y guardados como resúmenes estadísticos. En la segunda etapa, los subclusters obtenidos en la fase anterior son considerados como observaciones individuales y se agrupan en el número de clusters deseado mediante un procedimiento jerárquico.

Como el número de subclusters creados en la primera etapa es mucho menor que las observaciones originales, el algoritmo jerárquico de la segunda etapa puede desarrollarse de forma eficaz.

▪ **Etapas 1: PRE CLUSTER**

Siguiendo la nomenclatura utilizada por Zhang et al. (1996), denominaremos *cluster feature* (CF), a los parámetros estadísticos que resumen la información de los subclusters. En el caso del subclúster k , $CF_k = \{N_k, s_{Ak}, s_{Ak}^2, \vec{N}_{Bk}\}$ donde N_k es el número de observaciones que constituyen el conglomerado, s_{Ak} es la suma de las variables continuas de los N_k observaciones, s_{Ak}^2 es la suma de cuadrados de las variables continuas y $\vec{N}_{Bk} = N_{B1}, \dots, N_{BK_B}$ es un vector $\sum_{k=1}^{K_B} (L_k - 1)$ -dimensional donde los k subvectores son la dimensión $(L_k - 1)$, dada por $\vec{N}_{Bk} = (N_{k1}, \dots, N_{kL_k - 1})$, donde N_{kl} es el número de observaciones del conglomerado C_k , cuyas categorías toman el valor de la categoría l , siendo $l=1, \dots, L_k - 1$.

Cuando dos conglomerados, C_j y C_s , se unen creando uno nuevo, $C_{j,s}$, se estima un nuevo CF por adición de CF_j y CF_s ,

$$CF_{\langle J,S \rangle} = \{N_{j+s}, s_{Aj} + s_{As}, s_{Aj}^2 + s_{As}^2, \vec{N}_{Bj} + \vec{N}_{Bs}\}$$

Los CF constituyen una práctica herramienta para representar los datos originales, ya que se almacenan en bases mucho más pequeñas y contienen toda la información necesaria para efectuar posteriormente un análisis cluster jerárquico.

Obtención de grupos homogéneos

En esta primera etapa el algoritmo va construyendo un *CF-tree*, que se va modificando conforme se van agrupando las observaciones en los distintos subclusters. Los parámetros que son necesarios definir para construir el CF-tree son:

- B: Número de entradas en cada nodo interno, no terminal.
- L: Número máximo de entradas del nodo terminal.
- T: holgura máxima, que coincide con el diámetro máximo del nodo terminal.

Esquemáticamente, en la figura 6.1, se muestra un ejemplo del funcionamiento del CF-tree.

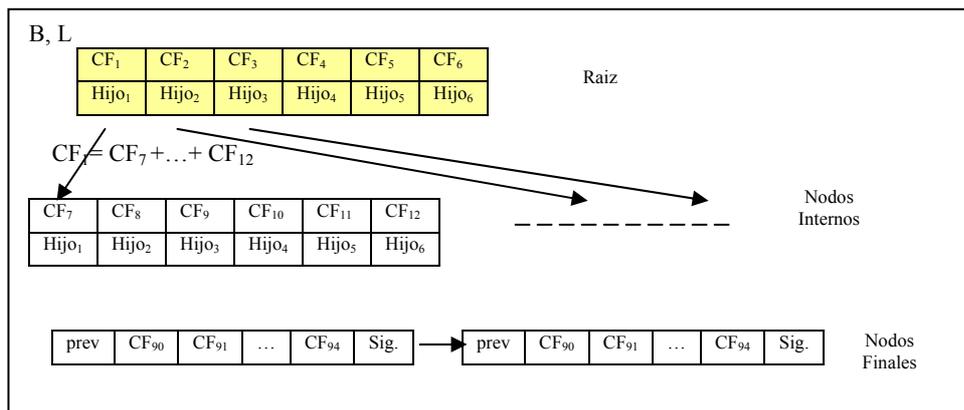


Figura 6.1. CF-tree

El proceso continúa recursivamente hasta que todas las observaciones descienden a través del CF-tree y llegan a un nodo terminal. Las observaciones se van agrupando en los nodos terminales en función del valor de holgura que se haya establecido.

En esta etapa es posible detectar datos anómalos evaluando aquellas observaciones que no se agrupan bien en ningún cluster. Si el número de observaciones que no se agrupan bien en un nodo terminal, es menor que una cierta fracción predefinida del tamaño del mayor nodo terminal, se consideran datos anómalos.

▪ **Etapas 2: CLUSTER**

En esta fase, los preclusters formados previamente son agrupados en el número deseado de clusters mediante el procedimiento jerárquico de aglomeración utilizando el logaritmo de máxima verosimilitud como medida de distancia. La precisión obtenida en los resultados finales depende del número inicial de preclusters formados, de manera que cuanto mayor sea éste, mejor precisión obtendremos en los resultados finales.

Para determinar el número de clusters más adecuado, se puede utilizar un procedimiento en dos fases que funciona bien en el caso de aplicar la aglomeración jerárquica. Este procedimiento consiste en estimar, en una primera fase, el índice BIC (Schwartz's Bayesian Criterion, también llamado Bayesian Information Criterion), o el AIC (Akaïke's Information Criterion), para cada número de clusters dentro de un rango especificado. Estos índices vienen definidos para k clusters como:

$$BIC(K) = -2 \sum_{k=1}^K \xi_k + m_k \log(N)$$

$$AIC(K) = -2 \sum_{k=1}^K \xi_k + 2m_k$$

donde,

$$m_k = K \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\}$$

Una vez estimados los índices, se escoge como número de clusters definitivo, aquel que produzca el mayor incremento en la distancia, o en el valor de los mismos, entre dos etapas seguidas del análisis cluster jerárquico.

6.2.5. Metodología del Análisis Discriminante

Con la finalidad de evaluar cuales de las variables o ítems incluidas en el cuestionario resultaban más relevantes para discernir entre los grupos obtenidos mediante el análisis clúster, se decidió efectuar un análisis discriminante. La utilización de esta técnica multivariante se manifiesta como adecuada en este caso, ya que permite:

- Analizar la pertenencia de observaciones a un determinado grupo, en función de un conjunto de variables independientes.
- Determinar si existen diferencias estadísticamente significativas entre los perfiles de puntuaciones medias sobre un conjunto de variables de unos grupos definidos a priori.
- Determinar cuales de las variables independientes cuantifica mejor las diferencias en los perfiles de las puntuaciones medias de los distintos grupos.

Se realizó un primer análisis discriminante utilizando como variable dependiente una nueva variable categórica, a la que denominamos QCL1, que nos indica el clúster de pertenencia de cada una de las observaciones incluyendo como variables independientes la totalidad de los ítems incluidos en el cuestionario, sin considerar los relativos a evaluación.

Asimismo, y debido a que la variable codificada como R19 puede considerarse un ítem criterio que está correlacionado con todos ítems del cuestionario, se decidió efectuar de nuevo el análisis discriminante eliminando esta variable del estudio para verificar si se modifica el poder discriminante del resto de variables. En este caso, se utilizó como variable categórica una nueva variable denominada QCL2 que nos indica el clúster de pertenencia de cada una de las observaciones tras haber repetido el análisis clúster sin incluir éste ítem global y como variables independientes los otros ítems incluidos en el cuestionario, sin considerar los referentes a evaluación.

6.2.5.1. Supuestos del Análisis Discriminante

Los supuestos clave para obtener una correcta aplicación del análisis discriminante son el de normalidad multivariante de las variables independientes y el de igualdad de matrices de covarianza y dispersión para los grupos.

El incumplimiento de la normalidad multivariante puede causar problemas en la estimación de la función de dispersión. Las matrices de covarianzas distintas pueden afectar desfavorablemente al proceso de clasificación. Otro supuesto implícito en algunas técnicas multivariantes es que todas las relaciones son lineales. Las relaciones no lineales no están reflejadas en la función discriminante, a menos que se realicen transformaciones específicas de las variables para representar los efectos no lineales.

Finalmente, la presencia de atípicos puede afectar considerablemente la precisión clasificatoria de cualquier resultado del análisis discriminante, siendo necesaria su detección y eliminación.

En nuestro caso, al trabajar con valores medios, como ya se comentó en el capítulo 5, por la aplicación del Teorema Central del Límite, la media de las observaciones tenderá a seguir una distribución normal y podremos evitar las consecuencias de posibles valores extremos.

6.2.5.2. Estimación del Modelo Discriminante

El análisis discriminante implica obtener una serie de funciones lineales de dos o más variables independientes que permitan interpretar las diferencias entre los grupos y clasificar a los individuos en algunas de las subpoblaciones definidas por la variable dependiente.

Una función discriminante es una combinación lineal de variables discriminantes que cumple ciertas condiciones. Así, si partimos de p variables (X_1, \dots, X_p) medidas para g grupos de individuos, la función discriminante quedaría de la forma:

$$y_{km} = u_0 + u_1 X_{1km} + u_2 X_{2km} + \dots + u_p X_{pkm}$$

donde,

y_{km} es el valor de la función discriminante para el caso m en el grupo k .

X_{ikm} es el valor de la variable discriminante X_i para el caso m en el grupo k .

u_i con $i = 1, 2, \dots, p$ es el conjunto de coeficientes que garantizan las características requeridas para la función.

Las condiciones que deben cumplir estas funciones responden fundamentalmente a que en el orden en el que van siendo obtenidas representen la máxima diferenciación posible entre grupos. Es decir, para la primera función discriminante que determinamos, los coeficientes u_i se obtienen de manera que los valores medios alcanzados en cada grupo sean entre sí lo más diferentes posible.

Para la segunda función, los coeficientes se determinan de manera que satisfagan el mismo criterio, pero además se añade la condición de que los valores obtenidos mediante la primera función no estén correlacionados con los obtenidos mediante la segunda. La tercera función tendrá en cuenta el criterio de maximizar las diferencias entre grupos y el no estar correlacionadas con ninguna de las funciones determinadas anteriormente. Las funciones discriminantes sucesivas se obtendrán de manera similar.

El número máximo de funciones discriminantes que podemos llegar a determinar será igual al número de variables discriminantes, o bien, al número de grupos menos uno, tomando el menor de ambos valores. Si denominamos q al número máximo de funciones posible,

$$q = \min(p, q - 1)$$

Si aplicamos las funciones discriminantes a los valores observados en las variables X_i para un individuo, obtenemos las denominadas puntuaciones discriminantes de ese individuo. Es decir, para un sujeto m perteneciente al grupo k , cuyos valores observados en las p variables originales son x_{ikm}, \dots, x_{pkm} , las funciones discriminantes permitirán obtener un nuevo conjunto de q puntuaciones y_{1km}, \dots, y_{qkm} .

Significado geométrico de las funciones discriminantes

Si n es el número de individuos y p el de variables independientes, los datos podrán ser dispuestos en una matriz de n filas por p columnas:

	X_1	X_2	...	X_p
Sujeto 1	x_{11}	x_{12}	...	x_{1p}
Sujeto 2	x_{21}	x_{22}	...	x_{2p}
⋮
Sujeto n	x_{n1}	x_{n2}	...	x_{np}

Cada individuo o fila de esta matriz, puede ser considerado un punto del espacio p -dimensional, definido al tomar a las variables discriminantes como ejes de dicho espacio. Los valores alcanzados por cada caso en las p variables constituirán sus coordenadas en los ejes del espacio de p dimensiones definido. Si los individuos de un grupo se comportan de manera similar respecto a las variables discriminantes, sus coordenadas espaciales serán similares y quedarán localizados en la misma región del espacio.

La posición de un grupo en el espacio puede ser caracterizada por su centroide y podemos evaluar las diferencias entre los grupos en función de las variables discriminantes, analizando la posición exacta de los centroides para determinar si los grupos quedan suficientemente diferenciados en el espacio multidimensional definido por las p variables. Los centroides de los g grupos definirán un espacio de dimensión $g-1$. Dentro del espacio definido por los centroides, el origen podría ser situado en el centroide general, correspondiente a todos los casos considerados globalmente. Localizado el origen del espacio formado por los centroides, podemos definir la orientación del primer eje en la dirección en la que los centroides aparezcan más separados, el segundo eje se define con el mismo criterio de conseguir la máxima diferenciación entre los centroides, pero añadiendo la condición de que sea perpendicular al primero y así sucesivamente.

En el caso de tres grupos, la representación gráfica se muestra en la figura 6.2.

Cada uno de esos ejes que maximizan la diferencia entre grupos y resultan ortogonales entre sí, corresponden a una función discriminante. La ecuación de la función discriminante facilitará la transformación de las coordenadas en el espacio p -dimensional

definido por las variables al nuevo sistema de ejes en el espacio determinado a partir de los centroides. Por tanto, cada ecuación discriminante permitirá obtener las puntuaciones discriminantes, que serán las coordenadas de los diferentes casos sobre un eje discriminante.

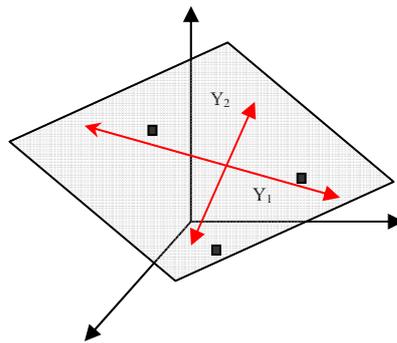


Figura 6.2. Ejes de máxima dispersión para tres centroides.

Obtención de los coeficientes de las funciones discriminantes

Las funciones lineales de discriminación deberán de ser determinadas como combinaciones lineales de las variables discriminantes que maximicen el cociente entre la variabilidad intergrupos y la variabilidad intragrupos. La primera función hará que ese cociente sea el máximo posible para las puntuaciones discriminantes. La segunda función deberá ser no correlacionada con la primera y hará máxima la razón entre la variabilidad residual intergrupos e intragrupos y así sucesivamente. De esta manera, las funciones discriminantes extraídas estarán ordenadas de forma decreciente en cuanto a la variación que explican en el sistema.

En consecuencia, los coeficientes estimados para las funciones discriminantes u_i , harán que las diferencias entre las medias de los grupos sobre estas funciones sean lo más grande posible. Para obtener dichos coeficientes son necesarias unas consideraciones previas.

Consideremos una combinación lineal de las p variables independientes que se desean utilizar para discriminar entre los g grupos en los que se clasifican n individuos.

Para determinar las diferencias entre medias de los g grupos, podemos utilizar la razón F, definida como:

$$F = \frac{SC_{entre} / (g - 1)}{SC_{intra} / (n - g)}$$

donde n y g son constantes para el número de individuos y grupos.

Para poder obtener la expresión de las sumas de cuadrados para la variable i , debemos definir previamente las matrices de sumas de cuadrados y productos cruzados intragrupos (W) e intergrupos (B).

Sea T la matriz total de sumas de cuadrados y productos cruzados, en la que cada elemento t_{ij} viene expresado por

$$t_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (X_{ikm} - \bar{X}_i)(X_{jkm} - \bar{X}_j)$$

donde g es el número de grupos, n_k es el número de casos en el grupo k , X_{ikm} y X_{jkm} son los valores que alcanzan las variables i y j para el caso m en el grupo k y \bar{X}_i y \bar{X}_j son las medias de los valores alcanzados por todos los individuos en las variables i y j .

Esta matriz de dimensiones $p \times p$ proporciona información sobre la variabilidad total que presenta el conjunto de las variables, ya que si dividimos el elemento t_{ij} por $(n-1)$, obtenemos la expresión de la covarianza entre dos variables. En el caso de los elementos de la diagonal principal, el cálculo se reduce a la suma de las desviaciones cuadradas respecto a la media global y si dividimos por $(n-1)$ obtendremos la expresión de la varianza para cada variable.

La matriz W, en la que cada elemento viene representado por w_{ij} , es similar a la anterior, excepto que las desviaciones en cada variable son calculadas respecto de la media del grupo al que pertenecen y recoge información sobre la variabilidad en el interior de los grupos.

La matriz B, en la que cada elemento viene representado por b_{ij} puede obtenerse como diferencia entre T y W y proporciona información sobre la variabilidad intergrupos.

A partir de estas matrices podemos definir las sumas de cuadrados para la variable i como,

$$\begin{aligned}SC_{inter} &= v'Bv \\ SC_{intra} &= v'Wv\end{aligned}$$

donde v es el vector columna obtenido por los pesos v_1, \dots, v_p y v' el vector fila obtenido por la transposición de éste. Es decir que el criterio de discriminación quedaría:

$$\frac{SC_{inter}}{SC_{intra}} = \frac{v'Bv}{v'Wv} = \lambda$$

donde λ es un escalar que nos proporciona un criterio para establecer la discriminación de grupos a lo largo de la dimensión especificada por el vector v . Tras decidir un criterio de discriminación, el objetivo es encontrar los coeficientes v_1, \dots, v_p que maximicen λ y para ello debemos estimar:

$$\frac{\partial \lambda}{\partial v} = 0$$

que con las simplificaciones necesarias equivale a estimar:

$$(B - \lambda W)v = 0$$

y asumiendo que el determinante de W es distinto de 0, podemos estimar su inversa, de manera que la expresión quedaría:

$$(W^{-1}B - \lambda I)v = 0$$

Para que las soluciones de λ y v correspondan respectivamente con los autovalores de la matriz $W^{-1}B$ y los autovectores asociados, llamaremos q al rango de dicha matriz y tendremos q autovalores, $\lambda_1, \dots, \lambda_q$, distintos de cero, cuya magnitud presenta un orden creciente y q autovectores asociados.

Si tenemos en cuenta, que el autovalor λ_m representa el valor asumido por el criterio de discriminación de grupos para la combinación lineal de las p variables construidas usando los elementos del autovector v_m como coeficientes y que el primer autovalor es el de mayor magnitud, la función definida como:

$$Y_1 = v_{11}X_1 + \dots + v_{1p}X_p$$

será la combinación lineal de las p variables que consiga una mayor discriminación lineal. La función Y_2 , será la de mayor capacidad de discriminación después de la primera y además no está correlacionada con ella y así sucesivamente.

Los coeficientes v_i obtenidos son útiles para la clasificación, pero no son interpretables como coeficientes y las puntuaciones a las que conducen no tienen un significado intrínseco ya que no se han calculado bajo ninguna restricción relativa al origen y la métrica del espacio discriminante. El ajuste de los valores v_i nos permitirá encontrar los coeficientes de las funciones discriminantes u_i :

$$u_i = v_i \sqrt{(n-g)}$$

$$u_0 = -\sum_{i=1}^p u_i \bar{X}_i$$

Esta transformación representa trasladar el origen de cada eje discriminante para hacerlo coincidir con el centroide global y varía la unidad de medida, haciendo que la puntuación discriminante asignada sobre un eje a cada caso represente el número de desviaciones típicas que se aleja del centroide. La función discriminante para todos los casos poseerá una media de 0 y una desviación típica intragrupos de 1. La puntuación obtenida por un individuo en una función discriminante representará su posición en el eje definido por esa función.

Existen dos posibles métodos de seleccionar las variables que deben incluirse en la función discriminante: el método simultáneo y el método por etapas. En el primer caso, la función discriminante se estima considerando todas las variables independientes de manera simultánea, sin considerar la capacidad discriminante de cada una de ellas.

En la estimación por etapas las variables son introducidas en la función discriminante de una en una, según su capacidad discriminatoria. Se comienza eligiendo la variable que mejor discrimina. Esta variable inicial se empareja en la siguiente etapa con cada una de las variables independientes restantes, escogiéndose la variable que más incrementa la capacidad discriminante de la función, en combinación con la primera variable. El resto de variables se van seleccionando de forma similar. A su vez, mientras se van incluyendo variables adicionales, pueden ir eliminándose algunas de las seleccionadas previamente, si la información que contienen sobre las diferencias del grupo está contenida en alguna combinación de otras variables incluidas en etapas posteriores. El conjunto reducido, es generalmente tan bueno y, algunas veces mejor que, el conjunto completo de variables (Hair et al., 1995).

Además, aplicando el procedimiento paso a paso y utilizando el criterio de la tolerancia para seleccionar las variables se soluciona el problema de trabajar con variables que presenten colinealidad. (Gil Flores et al., 2001).

Una vez estimada la función discriminante debe valorarse el nivel de significación estadística. Para ello se disponen de varios criterios estadísticos, siendo los más utilizados la medida del lambda de Wilks en el caso la estimación simultánea y la medida de la distancia de Mahalanobis en la estimación por etapas.

El criterio del lambda de Wilks promueve la selección de un modelo que maximiza las distancias entre los grupos y aumenta la cohesión de los mismos. La lambda de Wilks para un conjunto de p variables viene dada por la expresión:

$$\Lambda = \frac{|W|}{|W + B|}$$

El procedimiento de la distancia de Mahalanobis se basa en la distancia euclídea al cuadrado generalizada que se adecua a varianzas desiguales. Su expresión en el caso de dos grupos quedaría:

$$D^2_{i,j} = \sum_{k=1}^p (y_{ik} - y_{jk})^2$$

La principal ventaja de este método es que se calcula en el espacio original de las variables predictoras, pero presenta el inconveniente de poder llegar a ser crítico al incrementarse el número de variables ya que no se produce ninguna reducción de dimensionalidad. Por otro lado, la pérdida de dimensionalidad causa una pérdida de información al hacer disminuir la variabilidad entre las variables independientes, siendo en consecuencia, el procedimiento de la distancia Mahalanobis el más recomendado cuando se intenta aprovechar al máximo la información disponible.

6.2.5.3. Funciones de clasificación

Permiten asignar cada caso a uno de los grupos. Fisher fue el primero en proponer que los individuos podrían clasificarse mediante una combinación lineal de las variables discriminantes que maximicen las diferencias entre grupos y minimicen las diferencias intragrupos. La función de clasificación para un grupo, definida en el caso del grupo k , quedaría como:

$$h_k = b_{k0} + b_{k1}X_1 + \dots + b_{kp}X_p$$

donde h_k es la puntuación obtenida para ese grupo por un caso y los coeficientes b se obtienen como:

$$b_{ki} = (n - g) \sum_{j=1}^p a_{ij} \bar{X}_j \qquad b_{k0} = -0,5 \sum_{j=1}^p b_{kj} \bar{X}_j$$

siendo a_{ij} un elemento de la matriz inversa de W .

Estas puntuaciones tienen la propiedad de ser más elevadas cuanto mayor sea la proximidad del caso a grupo.

6.3. Resultados y discusión del Análisis Clúster

Con la finalidad de agrupar las valoraciones medias obtenidas a partir de las encuestas cumplimentadas por alumnos, en un número de grupos más manejable, se efectuó un análisis clúster. Con los conglomerados definidos y conocida la estructura subyacente,

se puede proceder a analizar las características más relevantes de cada uno de ellos e identificar si existen determinadas variables que hagan que los alumnos tengan una actitud u otra hacia los profesores.

A continuación se muestran los resultados obtenidos mediante la aplicación de los tres posibles algoritmos definidos para la obtención de los conglomerados, comparándose los resultados obtenidos mediante la utilización, por un lado, del algoritmo jerárquico y no jerárquico como técnicas complementarias y, por otro lado, del algoritmo two-step.

6.3.1. Análisis jerárquico y no jerárquico del periodo evaluativo 1999/2000

En primer lugar se utilizó el método jerárquico para identificar el número de conglomerados más apropiado, ya que este procedimiento presenta la ventaja frente al no jerárquico, de no necesitar ningún conocimiento a priori del número de clusters que se van a formar, siendo utilizado en muchas ocasiones como un estudio exploratorio para posteriormente aplicar el método no jerárquico y mejorar la solución clúster obtenida (Sharma, 1996).

Para la aplicación de este procedimiento fue necesario tomar una muestra, ya que aunque los procesos de formación de los conglomerados son, relativamente rápidos, los métodos jerárquicos no son susceptibles de analizar grandes bases de datos. Por ello se tomó una muestra aleatoria de unas 200 observaciones, mediante el procedimiento de muestreo aleatorio simple, debido a que este método resulta adecuado cuando se trabaja sobre bases de datos plenamente identificadas y es recomendable por su sencillez y facilidad para calcular errores en el muestreo. En este caso, aunque se sabe que el tamaño de muestra debe depender de la precisión de las estimaciones, de la varianza o dispersión de los datos, del tipo de muestreo considerado y del nivel de confianza deseado, entre otros factores, se consideró aproximadamente un 10% de los datos originales como adecuado, aunque en otros casos debe ser especificado en función de la homogeneidad y del tipo de investigación (Rodríguez Osuna, 1991).

El algoritmo utilizado es el método de Ward, para minimizar las diferencias dentro del conglomerado y evitar problemas con el encadenamiento de las observaciones (Hair et

al., 1995). Además, como nos muestran Pung y Steward (1983) en su revisión al respecto, existen numerosos estudios en los que se demuestra que esta técnica mejora los resultados obtenidos respecto de otros métodos.

La medida de distancia utilizada será la distancia euclídea al cuadrado, ya que como establecen Hair et al. (1995) y Sharman (1996) al trabajar con variables medidas en la misma escala tipo Likert y no ser necesaria la estandarización de las mismas, es la medida de distancia más recomendada. Además es la medida recomendada para los métodos de análisis clúster del centroide y Ward.

En la tabla 6.3 se muestran las etapas finales del historial de conglomeración, incluyéndose las observaciones que se han combinado en cada caso y el coeficiente de aglomeración, que en este método es la suma de los cuadrados dentro del conglomerado, mientras que en otros es la distancia euclídea entre los dos conglomerados que se están uniendo.

Tabla 6.3. Historial de conglomeración

Etapa	Conglomerado que se combina		Coeficientes	Etapa en la que el conglomerado aparece por primera vez		Próxima etapa
	1	2		1	2	
190	59	160	1713,182	178	180	194
191	44	287	1770,166	184	172	197
192	38	147	1833,695	181	164	196
193	113	445	1910,474	176	185	198
194	59	393	1990,603	190	0	197
195	29	47	2073,983	183	188	196
196	29	38	2259,170	195	192	199
197	44	59	2459,421	191	194	199
198	113	129	2701,260	193	189	200
199	29	44	3702,214	196	197	200
200	29	113	6838,044	199	198	0

Para determinar el número de conglomerados más adecuado, suele utilizarse el coeficiente de aglomeración como regla de parada, de manera que valores pequeños de este coeficiente indican que están surgiendo nuevos conglomerados homogéneos, mientras que

valores del coeficiente elevados o cambios muy elevados en el valor del mismo indican que se están uniendo conglomerados heterogéneos. En nuestra muestra se detecta un cambio elevado del coeficiente de aglomeración al pasar de la etapa 198 a la 199, lo que nos indicaría la existencia de 3 conglomerados.

El dendograma del proceso de aglomeración se muestra en la figura 6.3. En el eje de abscisas se representan las observaciones y en el eje de coordenadas la distancia euclídea entre los centroides de los clusters. El análisis de este dendograma también nos permite identificar posibles atípicos como ramas que se unieron muy tarde, cosa que no se detecta en nuestro caso.

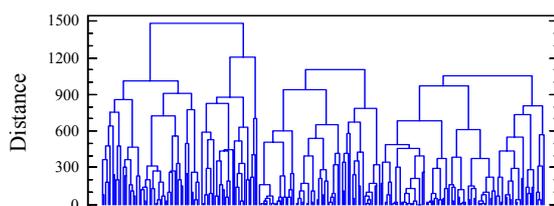


Figura 6.3. Dendograma

En él se puede detectar de nuevo, la existencia de tres conglomerados.

Tras la realización del análisis cluster jerárquico se utilizaron técnicas no jerárquicas para comparar y ajustar los resultados obtenidos. En las tablas 6.4 y 6.5 se muestran los valores obtenidos para los centros finales de los conglomerados, así como la distancia euclídea entre los centros de los conglomerados y el número de casos u observaciones en cada conglomerado. Las variables han sido codificadas como $MEDIA R_i$, siendo R_i cada uno de los ítems incluidos en el cuestionario.

El conglomerado 1 está constituido por observaciones que poseen todos los ítems con puntuaciones bajas, casi siempre inferiores a 5, por lo que podemos establecer que está integrado por las puntuaciones de aquellos alumnos que poseen menor nivel de satisfacción; el conglomerado 2 presenta puntuaciones en casi todos los ítems superiores a 7, es decir, está constituido por estudiantes con un elevado nivel de satisfacción, tanto en lo

relativo con el profesorado como con la asignatura; el conglomerado 3 está constituido por observaciones que poseen puntuaciones medias, entre 6 y 7.

Tabla 6.4. Centros de los conglomerados finales

	Conglomerado		
	1,00	2,00	3,00
MEDIA R1	4,40	8,29	6,64
MEDIA R2	5,36	8,64	7,26
MEDIA R3	4,49	7,99	6,37
MEDIA R4	4,96	8,20	6,75
MEDIA R5	5,05	8,37	6,94
MEDIA R6	4,63	7,65	6,28
MEDIA R7	5,13	8,28	6,91
MEDIA R8	4,11	6,84	5,48
MEDIA R9	4,50	7,86	6,31
MEDIA R10	3,90	7,58	5,75
MEDIA R11	4,91	8,43	6,86
MEDIA R12	3,25	7,39	5,37
MEDIA R16	4,78	7,51	6,13
MEDIA R17	5,24	7,97	6,70
MEDIA R18	4,84	7,73	6,40
MEDIA R19	4,56	8,58	6,85

Tabla 6.5. Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3
1		13,42	7,31
2	13,42		6,13
3	7,31	6,13	

Como puede observarse, la distancia euclídea entre los centros finales de los conglomerados es en todos los casos superior a 4, lo que nos indica que los clusters son diferentes entre si.

Obtención de grupos homogéneos

En la tabla 6.6 se muestra el número de casos en cada conglomerado, pudiendo apreciarse, que el conglomerado 1 presenta un tamaño relativamente inferior al resto, lo que nos indica que el número de alumnos que poseen valoraciones bajas es mucho menor que los que las poseen altas o medias.

Tabla 6.6. Número de casos en cada conglomerado

Conglomerado	1	385
	2	833
	3	917
Válidos		2135
Perdidos		6

Finalmente se evaluó la significación estadística de las 16 variables incluidas en el estudio. Los resultados obtenidos se muestran en la tabla 6.7.

Tabla 6.7. ANOVA

	Conglomerado		Error		F
	Media cuadrática	Gl	Media cuadrática	gl	
MEDIA R1	2039,11	2	,99	2132	2046,41
MEDIA R2	1450,97	2	,95	2132	1518,97
MEDIA R3	1688,49	2	1,14	2132	1485,87
MEDIA R4	1429,18	2	,62	2132	2328,99
MEDIA R5	1492,64	2	,68	2132	2188,89
MEDIA R6	1245,31	2	,84	2132	1485,41
MEDIA R7	1338,38	2	1,19	2132	1126,93
MEDIA R8	1044,34	2	,96	2132	1088,65
MEDIA R9	1552,21	2	,87	2132	1783,76
MEDIA R10	1897,63	2	1,25	2132	1515,52
MEDIA R11	1689,74	2	,89	2132	1892,73
MEDIA R12	2396,93	2	,86	2132	2789,08
MEDIA R16	1054,82	2	,92	2132	1145,62
MEDIA R17	1019,47	2	,88	2132	1157,02
MEDIA R18	1146,21	2	1,10	2132	1039,95
MEDIA R19	2182,21	2	,64	2132	3421,99

Para cada variable podemos evaluar la varianza atribuida a los conglomerados y la varianza de error. El estadístico F-ratio entre la varianza de los clusters y la de error, nos puede proporcionar una idea de las variables que mejor sirven para diferenciar entre los conglomerados. Cabe comentar aquí, que en este caso las pruebas F sólo deben considerarse con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados y minimizar la variabilidad dentro de cada conglomerado y los niveles críticos no son corregidos, por lo que no pueden considerarse como pruebas de hipótesis de que los centros de los conglomerados son iguales (Everitt, 1980). A la vista de los valores obtenidos del estadístico F, podemos considerar que todas las variables resultaban relevantes para la formación de los conglomerados.

6.3.2. Análisis clúster “two-step”

Finalmente, se aplicó el algoritmo “two-step”, recientemente propuesto por su adecuación cuando existen tanto variables continuas como categóricas. Este método presenta la ventaja de definir automáticamente el número de clusters que deben formarse y además efectúa un tratamiento previo de los atípicos excluyendo del estudio aquellas observaciones que no resulten adecuadas. La medida de distancia utilizada se basa en que la distancia entre dos clusters es equivalente a la disminución de la función del logaritmo de verosimilitud como resultado de los clusters que van surgiendo. Los cálculos de esta medida son bastante factibles, al depender únicamente de los dos clusters que se van uniendo en cada etapa, pudiéndose utilizar cuando se dispone de muestras de elevado tamaño, ventaja que no presentan los procedimientos jerárquicos. En este caso, se utilizó el Criterio Bayesiano de Schwartz’s para determinar el número de clusters, también denominado Criterio Bayesiano de Información, que en las salidas de los programas se suele denominar BIC (Bayesian Information Criterion). Aunque no vamos a profundizar en el razonamiento, se fundamenta en comparar la plausibilidad de dos modelos, frente a simplemente comparar las diferencias absolutas del ajuste. Cuando comparamos dos modelos un valor inferior del BIC indica un mejor modelo según ese criterio.

De nuevo mediante este procedimiento, se volvió a detectar que la estructura más simple que representaba soluciones homogéneas, correspondía a una solución de tres conglomerados, aunque los clusters obtenidos varían en tamaño y en los valores de los centroides de cada variable respecto a los obtenidos mediante los procedimientos tradicionales jerárquico y no jerárquico. En este caso se observa que los grupos obtenidos tienen un tamaño más homogéneo que los obtenidos mediante el procedimiento no jerárquico.

Tabla 6.8. Distribución de conglomerados

		N	% de combinados	% del total
Conglomerado	1	561	26,3%	26,2%
	2	964	45,2%	45,0%
	3	610	28,6%	28,5%
	Combinados	2135	100,0%	99,7%
Casos excluidos		7		,3%
Total		2142		100,0%

Los perfiles de cada conglomerado se muestran en la tabla 6.9 y en el gráfico 6.4. El conglomerado 3 está constituido por observaciones que poseen todos los ítems con puntuaciones bajas, casi siempre inferiores a 5; el conglomerado 2, está constituido por observaciones que poseen puntuaciones medias, entre 6 y 7; el conglomerado 1 estaría constituido por aquellos profesores con puntuaciones medias en casi todos los ítems superiores a 8. Asimismo se observa que las desviaciones típicas de cada uno de los ítems son mayores en el conglomerado 3, precisamente el que presenta puntuaciones medias inferiores a cinco, tomando en todas las variables valores superiores a la unidad, mientras que en los otros dos conglomerados se encuentran siempre por debajo de ella.

En este caso, se puede observar que los perfiles de los distintos grupos son prácticamente paralelos, lo que hace pensar que el estudiante gradúa su valoración a través de todo el cuestionario.

Tabla 6.9. Medias y desviaciones típicas de los centroides para los tres conglomerados

		Conglomerado			
		1,00	2,00	3,00	Combinados
MEDIA R1	Media	8,56	7,14	4,94	6,88
	Desv. típica	,75	,93	1,35	1,70
MEDIA R2	Media	8,88	7,67	5,81	7,46
	Desv. típica	,65	,84	1,40	1,52
MEDIA R3	Media	8,31	6,78	4,97	6,67
	Desv. típica	,89	1,01	1,34	1,65
MEDIA R4	Media	8,46	7,16	5,39	6,99
	Desv. típica	,66	,67	1,05	1,40
MEDIA R5	Media	8,59	7,37	5,50	7,16
	Desv. típica	,60	,69	1,21	1,44
MEDIA R6	Media	7,91	6,66	5,02	6,52
	Desv. típica	,82	,80	1,15	1,42
MEDIA R7	Media	8,51	7,29	5,59	7,12
	Desv. típica	,84	1,02	1,45	1,56
MEDIA R8	Media	7,11	5,83	4,42	5,76
	Desv. típica	,96	,88	1,11	1,39
MEDIA R9	Media	8,16	6,70	4,96	6,59
	Desv. típica	,79	,84	1,23	1,52
MEDIA R10	Media	7,95	6,21	4,33	6,13
	Desv. típica	,99	1,06	1,28	1,74
MEDIA R11	Media	8,70	7,29	5,40	7,12
	Desv. típica	,68	,86	1,33	1,57
MEDIA R12	Media	7,75	5,92	3,73	5,78
	Desv. típica	,88	,85	1,11	1,76
MEDIA R16	Media	7,80	6,50	5,05	6,43
	Desv. típica	,85	,84	1,13	1,38
MEDIA R17	Media	8,21	7,05	5,57	6,93
	Desv. típica	,80	,82	1,18	1,35
MEDIA R18	Media	7,98	6,78	5,18	6,64
	Desv. típica	,93	,97	1,23	1,47
MEDIA R19	Media	8,84	7,38	5,10	7,11
	Desv. típica	,57	,70	1,18	1,64

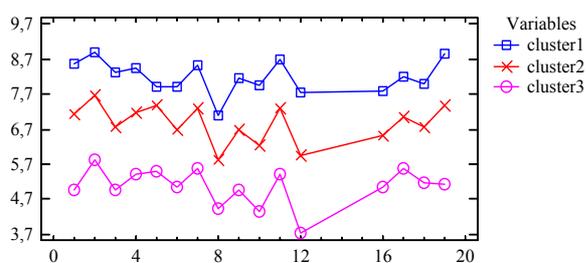


Figura 6.4. Valores de Centroides para los Conglomerados

En consecuencia, debido a la mayor homogeneidad en el tamaño de los grupos y las ventajas que presenta el algoritmo two-step comentadas anteriormente, utilizaremos la solución obtenida mediante este procedimiento para efectuar posteriormente el análisis discriminante.

6.4. Análisis de la estabilidad de los grupos homogéneos de calidad docente en distintos periodos evaluativos

Con la finalidad de realizar un estudio de carácter longitudinal, verificando la capacidad del cuestionario para establecer tipologías o perfiles de calidad docente a través de diversos periodos evaluativos, se repitieron los análisis anteriores con los datos procedentes de los cuestionarios de evaluación docente de la UPV durante el curso académico 00/01. No se efectuará en el resto de aplicaciones debido a que como ya se ha comentado anteriormente, es habitual en este tipo de estudios que la información presente diferentes estructuras entre aplicaciones, ya que en cada aplicación puede variar la información que se recoge de las distintas variables objeto de estudio, por el carácter dinámico de las mismas. La muestra para abordar este estudio estaba constituida por las puntuaciones medias obtenidas para cada uno de los ítems en las 541 posibles combinaciones profesor- asignatura.

En este caso, se efectuó un análisis clúster mediante procedimientos no jerárquicos y otro mediante el algoritmo two-step para determinar el número de conglomerados existentes. De nuevo mediante todas las técnicas se obtuvo que la mejor estructura de agrupación correspondía a la solución de tres conglomerados. A continuación se muestran las soluciones obtenidas en ambos casos.

6.4.1. Resultados del análisis clúster mediante el procedimiento no jerárquico

En la tabla 6.10 se muestran los valores obtenidos para los centros finales de los conglomerados. Al igual que sucedía en el curso académico 99/00, existe un conglomerado constituido por profesores que obtuvieron calificaciones relativamente bajas, entre 3 y 5,5, otro conglomerado constituido por profesores que obtuvieron calificaciones intermedias, entre 5,5 y 7,3 y un último conglomerado constituido por los profesores que obtuvieron calificaciones relativamente altas en todos los ítems.

Tabla 6.10. Centros de los conglomerados finales

	Conglomerado		
	1	2	3
MEDIA R1	6,70	4,40	8,40
MEDIA R2	7,27	5,39	8,73
MEDIA R3	6,47	4,46	7,96
MEDIA R4	6,75	5,02	8,13
MEDIA R5	7,00	5,00	8,00
MEDIA R6	6,40	4,80	7,70
MEDIA R7	7,14	5,53	8,35
MEDIA R8	5,68	4,23	6,95
MEDIA R9	6,48	4,71	7,86
MEDIA R10	6,07	4,24	7,68
MEDIA R11	7,00	5,20	8,40
MEDIA R12	5,56	3,35	7,54
MEDIA R16	6,36	5,02	7,62
MEDIA R17	6,98	5,37	8,07
MEDIA R18	6,51	4,97	7,81
MEDIA R19	6,89	4,59	8,61

A continuación se muestra la distancia euclídea entre ellos (tabla 6.11) y el número de casos u observaciones en cada conglomerado (tabla 6.12).

Tabla 6.11. Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3
1		7,301	5,800
2	7,301		13,089
3	5,800	13,089	

Las distancias euclídeas entre los centros de los conglomerados resultaron en todos los casos superiores a 4, lo que nos indica que son diferentes.

Tabla 6.12. Número de casos en cada conglomerado

Conglomerado	1	208,000
	2	91,000
	3	216,000
Válidos		515,000
Perdidos		2,000

Evaluando el número de casos de cada conglomerado, detectamos que igual que sucedía en el curso 99/00, el conglomerado con menor número corresponde al de los profesores que obtuvieron calificaciones relativamente bajas.

6.4.2 Resultados del análisis clúster mediante el procedimiento two-step

Procediendo como en los datos referentes al periodo 2000/2001, se utilizó el Criterio Bayesiano de Schwartz's para determinar el número de clusters y se utilizó el criterio del logaritmo de máxima verosimilitud para medir la distancia entre clusters. En las tablas 6.13 y 6.14, se muestran los perfiles y distribución de los tres conglomerados.

Tabla 6.13. Centroides de los conglomerados

		Conglomerado			
		1,00	2,00	3,00	Combinados
MEDIA R1	Media	4,42	6,84	8,53	7,02
	Desv. típica	1,20	,96	,70	1,70
MEDIA R2	Media	5,41	7,40	8,81	7,56
	Desv. típica	1,23	,94	,62	1,49
MEDIA R3	Media	4,50	6,59	8,05	6,74
	Desv. típica	1,27	,97	,84	1,58
MEDIA R4	Media	5,04	6,82	8,26	7,02
	Desv. típica	,91	,75	,63	1,35
MEDIA R5	Media	5,11	7,11	8,52	7,26
	Desv. típica	1,04	,75	,57	1,41
MEDIA R6	Media	4,81	6,51	7,77	6,66
	Desv. típica	,98	,79	,75	1,31
MEDIA R7	Media	5,56	7,24	8,42	7,37
	Desv. típica	1,11	,94	,81	1,36
MEDIA R8	Media	4,23	5,77	7,06	5,96
	Desv. típica	1,01	,83	,97	1,35
MEDIA R9	Media	4,74	6,57	7,97	6,75
	Desv. típica	1,21	,97	,77	1,48
MEDIA R10	Media	4,25	6,17	7,82	6,42
	Desv. típica	1,17	1,08	,92	1,63
MEDIA R11	Media	5,24	7,09	8,52	7,28
	Desv. típica	1,32	,95	,74	1,50
MEDIA R12	Media	3,38	5,73	7,64	6,00
	Desv. típica	,99	,89	,85	1,74
MEDIA R16	Media	5,00	6,45	7,73	6,65
	Desv. típica	,90	,87	,79	1,28
MEDIA R17	Media	5,36	7,04	8,18	7,15
	Desv. típica	1,17	,84	,73	1,31
MEDIA R18	Media	4,96	6,61	7,91	6,78
	Desv. típica	1,10	,98	,79	1,39
MEDIA R19	Media	4,60	7,05	8,71	7,21
	Desv. típica	1,01	,77	,52	1,61

En este caso, el tamaño de los conglomerados es similar al obtenido mediante los procedimientos no jerárquicos, siendo de nuevo el conglomerado de menor tamaño el correspondiente a los profesores que habían obtenido puntuaciones medias relativamente bajas.

Tabla 6.14. Distribución de conglomerados

	N	% de combinados	% del total
Conglomerado 1	92	17,9%	17,8%
Conglomerado 2	237	46,0%	45,8%
Conglomerado 3	186	36,1%	36,0%
Combinados	515	100,0%	99,6%
Casos excluidos	2		,4%
Total	517		100,0%

Evaluando el gráfico de los perfiles de los conglomerados, de nuevo detectamos que existe cierto paralelismo entre ellos.

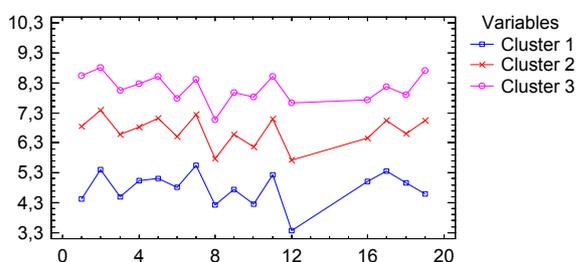


Figura 6.5. Valores de Centroides para los Conglomerados

Comentar finalmente, que a la vista de los resultados obtenidos, parece plausible proponer la existencia de tres tipologías docentes, en función de las puntuaciones medias del cuestionario de evaluación. Estas tipologías se mantienen en los distintos periodos evaluativos estudiados y parecen diferenciar entre profesores con puntuaciones medias relativamente bajas, intermedias o relativamente altas en casi todos los ítems incluidos en el cuestionario. El siguiente paso, será intentar caracterizar esas tipologías en función de

ciertos parámetros descriptivos adicionales, de manera que podamos conocer si existen ciertos rasgos que los alumnos tengan en cuenta a la hora de evaluar al personal docente.

De los distintos algoritmos utilizados en la obtención de los conglomerados, parece razonable proponer el procedimiento two-step, como el más apropiado, debido a las que permite determinar automáticamente el número de conglomerados más adecuado, lo que nos evita tener que realizar el doble análisis que suponía utilizar las técnicas tradicionales jerárquicas y no jerárquicas como complementarias y, por otro lado, elimina del estudio las observaciones atípicas, lo que podría ocasionar tener soluciones con un número de conglomerados inadecuado.

6.5. Caracterización de los clusters

Tras determinar el número de clusters que mejor representaba la estructura subyacente existente en los datos, se efectuó un nuevo estudio con la información adicional respecto a diferentes variables adicionales:

- Descriptivas del profesor: edad del profesor, categoría docente, tiempo de dedicación, si es doctor o no.
- Descriptivas de la asignatura: titulación, curso, edad del profesor, categoría docente, tiempo de dedicación, si es doctor o no, departamento al que pertenece y semestre en el que se imparte la asignatura.

El objetivo es determinar las características más relevantes de cada uno de los conglomerados y evaluar cuales de ellas permiten caracterizar los objetos a agrupar y, en consecuencia, están relacionadas con las valoraciones de los estudiantes a la hora de evaluar al profesorado.

Para evaluar la relación de dependencia entre el clúster de pertenencia y las distintas variables descriptivas se construyeron las tablas de frecuencia y se aplicó un contraste de hipótesis basado en el estadístico chi-cuadrado para evaluar la independencia de las variables en el caso de que ambas puedan considerarse categóricas. Es decir, dadas

Obtención de grupos homogéneos

dos variables X e Y, con p y q categorías respectivamente, sea $N=(n_{ij})$ la tabla de contingencia $p \times q$ que las resume, el estadístico chi-cuadrado quedaría definido por:

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n_{..}} \right)^2 \bigg/ \left(\frac{n_{i.} \cdot n_{.j}}{n_{..}} \right)$$

donde,

$$n_{i.} = \sum_{j=1}^q n_{ij}; \quad n_{.j} = \sum_{i=1}^p n_{ij}; \quad n_{..} = \sum_{i=1}^p \sum_{j=1}^q n_{ij}$$

Los resultados obtenidos, se compararán posteriormente, en el capítulo siete, con los resultantes de aplicar la técnica de segmentación basada en la detección automática de interacciones con Chi-cuadrado (CHAID). Esta técnica de segmentación proporciona los criterios y datos necesarios para establecer grupos homogéneos en función de determinadas características.

En los análisis efectuados para detectar si existe relación entre las tipologías identificadas y determinadas variables descriptivas, en general no se obtuvo ninguna posible relación con las variables descriptivas referentes al profesor, como sexo, edad, categoría docente, etc. Es decir, no se ha encontrado ninguna vinculación que permita señalar que existe una tendencia clara en los datos como para poder afirmar que los profesores de determinada edad, sexo, categoría docente o tiempo de dedicación se engloban más en determinados grupos puntuales.

Únicamente las variables titulación, departamento y semestre parecen tener cierta relación con los conglomerados obtenidos. Así, se detecta que algunas titulaciones presentan mayor porcentaje en el conglomerado de valoraciones medias más altas, entre las que cabe citar ETSIA, ETSIA TECNO-ALIM, o FI, mientras que otras presentan mayor porcentaje en el grupo de valoraciones medias más bajas como ETSII QUÍMICO o ETSIGCT-TOPOGRAFÍA.

Por departamentos, se detecta que los departamentos codificados como 9, 11, 16 y 31 presentan una proporción superior en el conglomerado de valoraciones medias altas,

mientras que los codificados como 15, 19, 39 y 42 lo hacen en el de puntuaciones medias más bajas.

Finalmente, las asignaturas del segundo semestre parecen tener una proporción ligeramente superior en el conglomerado de valoraciones medias altas.

Los resultados completos se muestran en el anexo 6 y serán comparados posteriormente con los obtenidos al aplicar la técnica de clasificación CHAID.

6.6. Análisis Discriminante

Tras la obtención de los conglomerados mediante la aplicación del análisis clúster, se utilizará el análisis discriminante considerando como variables independientes los ítems del cuestionario, con la finalidad evaluar más detalladamente las diferencias entre los conglomerados obtenidos y discernir cuáles son las variables que mejor contribuyen a explicar el perfil obtenido en cada grupo y, en función de las cuales pueden establecerse esas diferencias. Asimismo, se pretende identificar una función discriminante que permita clasificar futuras observaciones en base a características precisas y fáciles de determinar.

Para conseguir los objetivos propuestos se han construido dos modelos de análisis discriminante utilizando dos conjuntos de variables:

- Para analizar cuáles de las variables o ítems del cuestionario resultaban más relevantes para discriminar entre los conglomerados obtenidos, se decidió efectuar un análisis discriminante utilizando como variable dependiente la variable QCL1, que indica el clúster de pertenencia de cada observación y como variables independientes los 16 ítems del cuestionario, ya que no se incluían en el estudio los tres referentes a evaluación.
- Debido a que la variable codificada como R19 puede considerarse un ítem criterio, se decidió efectuar de nuevo el análisis discriminante eliminando esta variable del estudio, considerando como variable dependiente una nueva variable QCL2, que nos indica el clúster de pertenencia de cada observación sin

haber incluido el ítem R19 en el análisis clúster y como variables dependientes los otros 15 ítems del cuestionario.

Cabe comentar aquí, que no es adecuado emplear LDA cuando existe colinealidad entre las variables (Gil-Flores et al., 2001). La colinealidad se presenta cuando dos variables de la matriz de correlaciones presentan una correlación perfecta o casi perfecta y cuando muestran el mismo patrón de correlaciones con las restantes variables. Es decir, la colinealidad implicaría que dos variables se comportaran del mismo modo y aportan una información redundante. Entre las variables del cuestionario existe cierta colinealidad y como consecuencia de ello la aplicación de LDA provocaría que la matriz inversa presente valores muy inestables, sensibles al mínimo cambio en los valores de las correlaciones presentes en la matriz original. Sin embargo, este problema se puede solventar en parte trabajando con el procedimiento paso a paso y el criterio de tolerancia para seleccionar las variables (Gil-Flores et al., 2001).

El procedimiento de clasificación, en los LDA se ha basado en la utilización de las funciones de clasificación. Otro procedimiento se basaría en el cálculo de su distancia a los centroides de cada uno de los grupos. La observación sería clasificada en el grupo con cuyo centroide existe una menor distancia. La distancia de Mahalanobis es una distancia adecuada para valorar la proximidad entre centroides y casos.

6.6.1. Análisis discriminante incluyendo todos los ítems del cuestionario

En este caso se obtuvieron dos funciones discriminantes significativas con las que se explica más del 99% de la variabilidad. En las tablas adjuntas se muestran los valores de los autovalores (λ_i) y de la correlación canónica, estando este estadístico definido como:

$$CR_i = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}, \text{ para } i = 1, \dots, r$$

que mide, en términos relativos, el poder discriminante de la *i*-ésima función discriminante, ya que es el porcentaje de la variación total en dicha función el que es explicada por las diferencias entre los grupos.

Toma valores entre 0 y 1 de forma que, cuanto más cerca de 1 esté su valor, mayor es la potencia discriminante de la *i*-ésima función discriminante.

Tabla 6.13. Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica	Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	3,903	98,7	98,7	,892	1 a la 2	,194	3488,325	20	,000
2	,051	1,3	100,0	,220	2	,952	105,738	9	,000

Con la primera función obtenemos un porcentaje de varianza explicada del 99,0%. Aunque la segunda función no vaya a ser utilizada, consideramos conveniente mantenerla para posteriormente efectuar la representación gráfica.

En las tablas 6.14 y 6.15 se muestran la matriz de estructura y los coeficientes estandarizados de la función discriminante estimada. En ella, las variables no empleadas en el análisis vienen indicadas mediante una (a). Los coeficientes estandarizados de las funciones discriminantes constituyen la ponderación discriminante asignada a cada variable para calcular las funciones discriminantes y, representan la contribución relativa de la variable considerada asociada a esa función. La matriz de estructura nos muestra las cargas discriminantes o correlaciones de estructura que miden la correlación lineal simple entre cada variable independiente y la función discriminante. Estas cargas discriminantes reflejan la varianza que las variables independientes comparten con la función discriminante y pueden ser interpretadas como cargas de los factores para valorar la contribución relativa de cada variable a la función discriminante. A partir de ambos parámetros se puede deducir la expresión matemática de las funciones discriminantes en términos de las variables originales estandarizadas. Comentar aquí, que cuando existe cierto grado de colinealidad los coeficientes de la matriz de estructuras (Tabla 6.14.), al ser coeficientes de correlación bivariantes no están afectados por las relaciones con las restantes variables y son más

fiabiles que los estandarizados de la función discriminante (tabla 6.15), por lo que la interpretación de los resultados se hará a partir de ellos (Hair et al., 1995).

Tabla 6.14. Matriz de estructura

	Función	
	1	2
MEDIA R19	,848	-,460
MEDIA R12	,802	,158
MEDIA R4	,728	-,114
MEDIA R5(a)	,654	-,241
MEDIA R11(a)	,641	,060
MEDIA R1(a)	,639	-,277
MEDIA R9(a)	,619	,107
MEDIA R10	,609	,303
MEDIA R2	,586	-,307
MEDIA R6(a)	,580	-,056
MEDIA R3	,578	,124
MEDIA R16	,554	,219
MEDIA R8	,517	,236
MEDIA R18	,509	-,046
MEDIA R17(a)	,499	-,019
MEDIA R7	,492	-,115

Tabla 6.15. Coeficientes estandarizados de las funciones discriminantes canónicas

	Función	
	1	2
MEDIA R2	,156	,371
MEDIA R3	,132	,276
MEDIA R4	,096	,009
MEDIA R7	,094	-,425
MEDIA R8	,064	,360
MEDIA R10	,236	,300
MEDIA R12	,231	,677
MEDIA R16	,199	,304
MEDIA R18	,083	,010
MEDIA R19	,238	-1,486

En estas tablas podemos detectar que las variables con mayor poder discriminante son los ítems codificados como R19, R12, R10, R4, R3, R2 y R16, todos ellos correlacionados con la primera función discriminante y con puntuaciones positivas. Eso nos indicaría que los profesores con puntuaciones medias altas en la primera función son sujetos considerados, en global, como buenos profesores y que además presentan puntuaciones medias altas en motivar a los alumnos por la asignatura (R12), dialogar con los alumnos y considerar sus opiniones (R10), sintetizar los conceptos más importantes (R4), llevar un ritmo de la clase adecuado (R3), tener bien preparadas las clases (R2) y haber preparado y recomendado materiales útiles para la asignatura (R16). En este primer eje tendríamos representadas las 4 dimensiones extraídas mediante el Análisis Factorial efectuado en el capítulo cinco, para conocer la estructura subyacente del cuestionario, siendo algunas de ellas las variables indicadoras de cada dimensión.

La tabla 6.16 contiene las puntuaciones medias para cada grupo, conocidas como centroides de grupo. Fijándonos en los de la primera función, ya que la segunda cuenta con escaso poder de discriminación, podemos comprobar como los grupos 1 (profesores con puntuaciones relativamente bajas) y 3 (profesores con puntuaciones relativamente altas) prácticamente resultan opuestos mientras que el grupo 2 (profesores con puntuaciones intermedias) se sitúa en una posición intermedia cercana al origen. Respecto a la segunda función vemos que los grupos no se distancian substancialmente puesto que todos ellos poseen en esta función coordenadas próximas entre sí.

Tabla 6.16. Funciones en los centroides de los grupos

Número de conglomerados	Función	
	1	2
1	2,563	,239
2	,242	-,247
3	-2,740	,171

Finalmente se muestra en la gráfica 6.15 la discriminación entre grupos obtenida mediante las dos funciones estimadas.

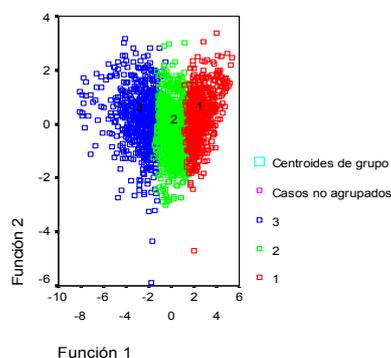


Figura 6.5. Funciones discriminantes

En este gráfico apreciamos de nuevo que la única función significativa es la primera de las dos obtenidas, en virtud de la cual los grupos aparecen separados.

El análisis discriminante se ha utilizado en esta aplicación con dos finalidades básicas: interpretar las diferencias existentes entre los grupos obtenidos mediante el análisis clúster y pronosticar la clasificación de nuevas observaciones.

En cuanto a la primera finalidad, con las funciones discriminantes obtenidas y mediante la representación en ellas de los datos correspondientes, se evidencia la diferencia entre grupos. Respecto a la segunda finalidad, si observamos la tabla de clasificación (Tabla 6. 17) vemos que los distintos clusters se distinguen casi completamente, estimándose que son clasificados correctamente el 94,0% de los casos agrupados originales.

Tabla 6.17. Resultados de la clasificación

		Número de conglomerados	Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	540	21	0	561
		2	50	903	11	964
		3	0	47	563	610
	%	1	96,3	3,7	,0	100,0
		2	5,2	93,7	1,1	100,0
		3	,0	7,7	92,3	100,0

6.6.2. Análisis discriminante sin incluir el ítem R19

Como se ha comentado anteriormente, debido a que la variable codificada como R19 puede considerarse un ítem criterio, se decidió efectuar de nuevo el análisis discriminante eliminando esta variable del estudio. De nuevo se obtuvo que con una sola función discriminante se explicaba el 99,1% de la varianza total.

Tabla 6.18. Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica	Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	3,895	98,9	98,9	,892	1 a la 2	,196	3463,503	24	,000
2	,041	1,1	100,0	,199	2	,960	86,126	11	,000

Los coeficientes estandarizados de la función discriminante y la matriz de estructuras se muestran a continuación.

Tabla 6.19. Coeficientes estandarizados de las funciones discriminantes canónicas

	Función	
	1	2
MEDIA R1	,135	,720
MEDIA R2	,159	-,185
MEDIA R3	,108	-,439
MEDIA R4	,098	,067
MEDIA R5	,085	,295
MEDIA R7	,102	,516
MEDIA R8	,066	-,354
MEDIA R10	,250	-,578
MEDIA R11	,019	,703
MEDIA R12	,270	-,483
MEDIA R16	,187	-,396
MEDIA R18	,106	,143

Tabla 6.20. Matriz de estructura

	Función	
	1	2
MEDIA R12	,803	-,145
MEDIA R4	,729	,153
MEDIA R5	,688	,417
MEDIA R1	,668	,410
MEDIA R11	,634	,108
MEDIA R9(a)	,618	-,070
MEDIA R10	,610	-,314
MEDIA R2	,586	,362
MEDIA R3	,579	-,116
MEDIA R6(a)	,578	,053
MEDIA R16	,555	-,223
MEDIA R8	,517	-,242
MEDIA R18	,509	,070
MEDIA R17(a)	,495	,021
MEDIA R7	,493	,146

En ellas se detecta que las variables con mayor poder discriminante son las mismas que en el caso anterior, habiéndose introduciendo en lugar del ítem criterio, los ítems R1 (el profesor explica con orden y claridad) y R5 (cuando responde a una pregunta lo hace realmente a lo preguntado con claridad y precisión), ambos muy correlaciones con el mismo y pertenecientes a la misma dimensión subyacente. Es decir, que al eliminar el ítem criterio del estudio, se modifican algunas de las variables de mayor poder discriminante entre los grupos obtenidos, siendo necesario introducir dos nuevas, ambas relacionadas con la primera dimensión subyacente obtenida mediante el análisis factorial, para obtener prácticamente el mismo poder discriminante entre los grupos. En este caso la variable con mayor poder discriminante pasa a ser el conseguir que los alumnos estén motivados por la asignatura (R12), seguido por sintetizar los conceptos más importantes (R4), por responder realmente a lo preguntado (R5) y por explicar con orden y claridad (R1).

La representación gráfica de las funciones discriminantes y la tabla con el porcentaje de observaciones correctamente clasificadas se muestra a continuación. De nuevo podemos observar que mediante la primera función discriminante se consigue separar bastante bien a los tres grupos.

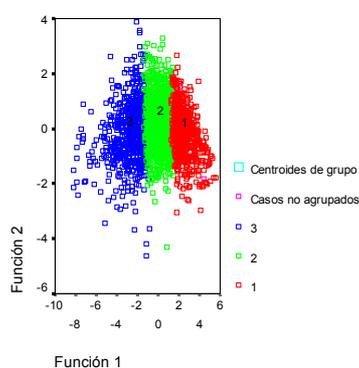


Figura 6.6. Funciones canónicas discriminantes

Finalmente en la tabla 6.21 se muestra el porcentaje de observaciones correctamente clasificadas.

Tabla 6.21. Resultados de la clasificación

		Número de conglomerados en dos fases			Grupo de pertenencia pronosticado			Total
					1	2	3	
Original	Recuento	1			539	22	0	561
		2			51	905	8	964
		3			0	43	567	610
	%	1			96,1	3,9	,0	100,0
		2			5,3	93,9	,8	100,0
		3			,0	7,0	93,0	100,0

El porcentaje de observaciones clasificadas correctamente es del 94,2% de los casos originales agrupados, ligeramente inferior al obtenido en el caso anterior.

6.7. Obtención de grupos homogéneos tras la aplicación del análisis discriminante

Con la finalidad de no utilizar variables irrelevantes en el análisis clúster y tras conocer las variables incluidas en el análisis discriminante, se volvió a efectuar un análisis clúster, incluyendo únicamente estos ítems y sin incluir el ítem global R19, para comprobar si únicamente incluyendo las variables con mayor poder discriminante, obteníamos el mismo número de tipologías o perfiles docentes.

De nuevo obtuvimos tres clusters. El número de observaciones en cada uno de ellos y el valor de los centros finales de cada conglomerado se muestra en las tablas 6.22 y 6.23.

Tabla 6.22. Número de casos en cada conglomerado

		N	% de combinados	% del total
Conglomerado	1	760	35,5%	35,5%
	2	512	23,9%	23,9%
	3	866	40,5%	40,4%
	Combinados	2138	100,0%	99,8%
Casos excluidos		4		,2%
Total		2142		100,0%

En este caso, se volvió a detectar que el cluster 3 corresponde a profesores que han obtenido puntuaciones medias bajas en casi todos los ítems considerados; el conglomerado 1 corresponde a profesores con puntuaciones medias altas, superiores a 7, en todos los ítems considerados y el conglomerado 2 corresponde a profesores con puntuaciones medias aceptables.

Es decir, que únicamente considerando los ítems de mayor poder discriminante, se consigue identificar las mismas tipologías o perfiles docentes que si se hubiesen incluido todas.

Tabla 6.23. Centros de los conglomerados finales

		Conglomerado			
		1,00	2,00	3,00	Combinados
MEDIA R1	Media	9,12	6,99	3,52	6,89
	Desv. Típica	,63	1,34	1,14	1,71
MEDIA R2	Media	9,35	7,55	4,63	7,46
	Desv. Típica	,54	1,21	1,46	1,52
MEDIA R3	Media	8,97	6,73	3,70	6,67
	Desv. Típica	,66	1,33	1,18	1,65
MEDIA R4	Media	9,07	7,06	4,25	7,00
	Desv. Típica	,54	1,06	1,01	1,40
MEDIA R5	Media	9,12	7,25	4,26	7,16
	Desv. Típica	,53	1,09	1,20	1,44
MEDIA R7	Media	9,04	7,20	4,44	7,13
	Desv. Típica	,68	1,29	1,43	1,56
MEDIA R8	Media	7,94	5,78	3,47	5,76
	Desv. Típica	,85	1,11	1,03	1,39
MEDIA R10	Media	8,81	6,17	3,14	6,13
	Desv. Típica	,80	1,40	1,16	1,74
MEDIA R11	Media	9,32	7,20	4,08	7,12
	Desv. Típica	,52	1,22	1,27	1,57
MEDIA R12	Media	8,55	5,83	2,48	5,78
	Desv. Típica	,80	1,37	,92	1,76
MEDIA R16	Media	8,49	6,45	4,20	6,43
	Desv. Típica	,72	1,12	1,19	1,38
MEDIA R18	Media	8,70	6,67	4,26	6,64
	Desv. Típica	,76	1,23	1,19	1,48

6.8. Conclusiones

La realización del análisis clúster pone de manifiesto que el cuestionario permite identificar adecuadamente, tanto por ítems como por dimensiones, tipologías de calidad docente percibida. Estas tipologías son paralelas, lo cual es concurrente con la capacidad de discriminación analizada en los ítems. Este hecho avala:

- Los perfiles identificados se producen por nivel de satisfacción con la docencia recibida, es decir, reflejan una graduación de calidad percibida.
- La identificación de tipos se mantiene a través de distintos periodos evaluativos.
- Esta diferenciación se produce en todos los ítems o dimensiones por lo que es posible explorar la capacidad de diferenciación de una puntuación de síntesis (como la media global) para puntuar al profesor. Este hecho concuerda con lo obtenido en otros estudios (González Such, 1997; González Such et al., 1990; Gonzalez Such et al., 1993), en los que se indica que el estudiante tiende a globalizar la valoración del profesor en este tipo de instrumentos.

Por otra parte, los grupos identificados no pueden ser caracterizados por otras variables ajenas al cuestionario, como sexo, categoría docente, tiempo de dedicación, si es doctor o no, etc. Únicamente parecen verse vinculados el departamento y la titulación.

Finalmente, mediante el análisis discriminante de los ítems del cuestionario se identificaron las principales variables que permiten discernir entre los grupos obtenidos. Mediante los ítems de mayor poder discriminante tenemos representadas todas las dimensiones subyacentes del cuestionario, lo que nos indicaría que los alumnos consideran todas ellas relevantes a la hora de evaluar a los profesores.

Por otro lado, obtuvimos una única función discriminante que nos permitiera clasificar a nuevos sujetos, es decir, que conociendo las puntuaciones medias obtenidas en los ítems de mayor capacidad discriminante, podemos conocer el perfil dentro del cual quedaría agrupado.

7. Identificación y caracterización de tipologías docentes

7.1. Introducción

En el presente capítulo, se aplica la técnica de árboles de clasificación CHAID (*Chi-Squared Automatic Interaction Detector*), como herramienta complementaria al análisis clúster realizado anteriormente y a la posterior caracterización de los grupos homogéneos obtenidos mediante dicha técnica en el capítulo 6. La finalidad de aplicar este método es segmentar los datos agrupados previamente en los distintos clusters y establecer grupos homogéneos en función de determinadas variables predictoras referentes a ciertas características de las asignaturas y/o de los profesores de las que disponíamos información (curso, titulación, departamento, semestre en el que se imparte, edad, categoría docente, si es doctor o no y tiempo de dedicación).

Esta técnica, como se verá posteriormente, presenta la ventaja de trabajar con cortes univariantes, examinando las variables predictoras una cada vez, en lugar de en conjunto, lo que nos permitirá detectar en cada caso la variable predictora de mayor nivel significativo para realizar la ramificación en cada caso concreto y, encontrar la división óptima de cada una de ellas.

La ventaja de utilizar este método frente a otras técnicas clásicas como el LDA, es que CHAID maneja con facilidad las interacciones entre variables que tantas dificultades plantean a otras técnicas de modelización y que pueden afectar sensiblemente a los resultados. Además los modelos CHAID pueden manejar simultáneamente variables independientes tanto categóricas como continuas que estén relacionadas. Concretamente, si las variables independientes son categóricas e implícitamente contienen relaciones, CHAID dará aún mejor resultado. Este modelo previene automáticamente el sobreparametrizaje y se ocupa de los datos perdidos.

Además, frente a otras técnicas multivariantes como el análisis de correspondencias múltiples, combinado con el análisis de conglomerados, la principal diferencia responde a que mientras en este último caso los grupos obtenidos finalmente no están definidos a priori, en el método CHAID sí lo están y responden a unos criterios lógicos determinados.

Tras la obtención de la segmentación y clasificación óptima de la variable dependiente se efectuará un LDA dentro de cada uno de los grupos homogéneos obtenidos, para determinar, entre los ítems que resultaron ser los de mayor poder discriminante en la diferenciación de los perfiles realizada en el capítulo 6, aquellos que tienen mayor relevancia en cada caso, lo que podría permitir en investigaciones posteriores, determinar los estándares y puntos de corte. Comentar brevemente aquí, que según Jornet y Suarez (1989) es conveniente diferenciar entre si un procedimiento se encamina a determinar un punto de corte en la escala de puntuaciones verdaderas, es decir, libres de error (a lo cual denominaremos procedimiento para determinar un estándar), o bien, si está destinado a determinar un punto de corte en la escala de puntuaciones observadas (a lo que denominaremos procedimiento para determinar un punto de corte propiamente dicho).

Finalmente, se aplicará el método SIMCA (Soft Independent Modelling of Class Analogy), que es un método de modelización y clasificación que se suele aplicar cuando las muestras que pertenecen a las clases son demasiado diferentes entre sí como para formar un solo modelo bilineal.

En nuestro caso, la aplicación de esta técnica, no se realiza con la finalidad de obtener una modelización y clasificación de las observaciones, que ya se realizó en el capítulo sexto mediante las técnicas rígidas, sino para obtener las variables de mayor poder discriminante entre los grupos formados, utilizando los modelos construidos con los departamentos, titulaciones y cursos, agrupados en los nodos de primera ramificación obtenidos con cada una de las variables predictoras utilizadas.

7.2. Método CHAID Exhaustivo

Las técnicas de árboles de clasificación, son particularmente útiles para analizar y extraer la estructura subyacente en grandes bases de datos. Se utilizan para predecir la clasificación de objetos en distintas clases de una variable dependiente categórica, en función de varias variables predictoras. Básicamente la técnica consiste en realizar particiones recursivas de la base de datos que sean mutuamente excluyentes y que mejor describan a la variable dependiente o variable respuesta. Se basan en segmentar la muestra inicial de datos, construyendo “árboles” donde cada nodo (no terminal), identifica una condición de partida para lograr obtener una segmentación y clasificación óptima de la variable dependiente.

Este método constituye un sistema de aprendizaje basado en un conjunto de reglas de decisión que predicen o clasifican las observaciones futuras. Los árboles de decisión son gráficos que ilustran unas reglas de decisión. Parten de un nodo raíz que contiene todas las observaciones de la muestra. A medida que se desplaza por el árbol, los datos se ramifican en subconjuntos de datos que se excluyen mutuamente. Se examinan todas las variables observadas en cada uno de los objetos para encontrar aquellas que proporcionen la mejor clasificación al dividir los datos en subgrupos. Este proceso se aplica de forma recursiva a los diferentes subgrupos obtenidos para definir otros grupos dentro de ellos, y así sucesivamente, hasta que se haya finalizado el árbol.

Finalmente, como se muestra en la figura 7.1, en el árbol se obtienen los nodos terminales que constituyen grupos homogéneos de observaciones, bien diferenciados del resto, en función de los valores que hayan tomado las variables predictoras.

Puede considerarse una técnica complementaria o alternativa a los métodos tradicionales de clasificación, como por ejemplo, el análisis discriminante o el análisis clúster, presentando algunas ventajas frente a ellos. En primer lugar, es una técnica flexible, lo que evita el cumplimiento de algunas hipótesis en las que se sustentan las técnicas multivariantes clásicas. En segundo lugar, trabaja con cortes univariantes, examinando las variables predictoras una cada vez, en lugar de en conjunto, lo que le permite ser menos

exigente y restrictivo con las medidas de las variables predictoras, pudiendo trabajar con variables ordinales, nominales y continuas.

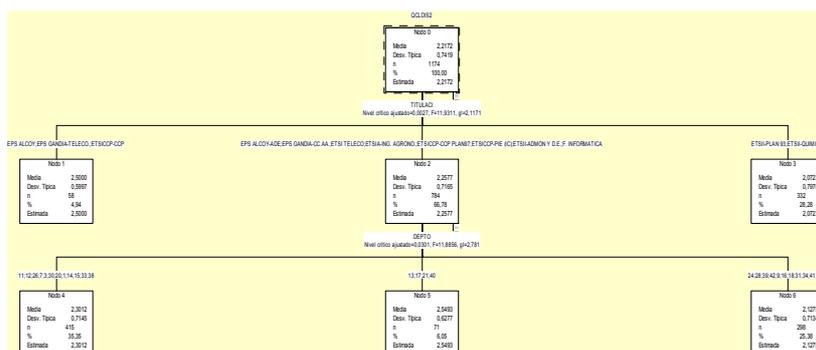


Figura 7.1. Árbol de clasificación

La técnica de los árboles de clasificación fue originada con el desarrollo del Detector Automático de Interacciones (AID) por Morgan y Sonquist (1963a y 1963b); Sonquist y Morgan (1964) y Sonquist et al. (1971). Sin embargo, esta técnica fue criticada por no utilizar test de significación estadística (Kass, 1975a y 1975b), identificar relaciones espúreas (Einhorn, 1972), tener tendencia a trabajar mal con pequeñas bases de datos (Doyle, 1975), tener unas reglas de parada inadecuadas (Doyle y Fenwick, 1975) y trabajar únicamente realizando particiones binarias (Kass, 1980).

Posteriormente, Kass (1980), propuso otro método, conocido como CHAID), que se trata de una técnica estadística muy eficiente para segmentar o generar árboles, determinando la mejor segmentación posible de los datos. Utilizando como criterio la significación de un contraste estadístico, evalúa todos los valores de una variable predictor potencial, funde los valores evaluados como estadísticamente homogéneos respecto a la variable criterio y conserva inalterados todos los valores que resulten heterogéneos. A continuación, selecciona la mejor variable predictor para formar la primera rama del árbol, de forma que cada nodo del mismo esté compuesto por un grupo de valores homogéneos de la variable seleccionada. La prueba estadística utilizada depende del nivel de medida de la

variable criterio, utilizándose la prueba del estadístico F si la variable es continua o el test del estadístico χ^2 si es categórica.

No es un procedimiento binario, pudiendo generar más de dos categorías en cualquier nivel del árbol, funciona con todo tipo de variables y aprovecha los valores perdidos tratándolos como una categoría individual.

Esta técnica, sin embargo, también presentaba algunos inconvenientes. En concreto, no puede trabajar con variables predictoras con un elevado número de categorías nominales y, en ocasiones, el método CHAID puede no encontrar la división óptima de una variable, ya que detiene la fusión de categorías, en cuanto constata que todas las categorías restantes son estadísticamente distintas.

Quinlan (1986), propuso una nueva técnica conocida como AI (Artificial Intelligence), que se basaba en técnicas inductivas para desarrollar el árbol. Sin embargo, estas técnicas inductivas presentaban el inconveniente de no poder segmentar la base de datos en un número de particiones superior al número de categorías de la variable predictora. Como consecuencia de esto, los árboles de clasificación obtenidos tenían muchas reglas redundantes que daban lugar a soluciones poco comprensibles en muchos casos.

Finalmente, Biggs propuso la técnica de segmentación basada en la detección automática de interacciones con Chi-cuadrado, conocida como CHAID Exhaustivo (Biggs et al., 1991), que es una extensión de la técnica CHAID y que supera los inconvenientes anteriormente citados de otras técnicas. Por ejemplo, resuelve el problema que presentaba el método CHAID de no poder encontrar la división óptima de una variable, al continuar fundiendo las categorías de la variable predictora hasta que sólo quedan dos supercategorías. A continuación, examina la serie de fusiones del predictor, busca el conjunto de categorías que proporciona la mayor asociación con la variable criterio y calcula un valor p (*nivel crítico*) corregido para esa asociación. De esta manera, el método CHAID exhaustivo puede encontrar la mejor división para cada predictor y, a continuación, elegir el predictor que se debe dividir comparando los niveles críticos corregidos.

El algoritmo utilizado en el CHAID Exhaustivo es muy similar al propuesto por Kass (1980). Funciona segmentando recursivamente cada agrupación o nodo no terminal en un conjunto de k nuevos nodos ($2 \leq k \leq$ número de categorías de la variable predictora (c)). El algoritmo comienza con un único nodo inicial que contiene todas las observaciones y continúa recursivamente hasta que no se pueden encontrar más segmentaciones significativas. En cada nodo, se examinan todas las variables predictoras como candidatas para realizar la segmentación del mismo, evaluándose la mejor partición que se podría obtener con cada una de ellas y analizando su significación estadística. Se utilizará el estadístico χ^2 para conocer el nivel de significación estadística de cada variable predictora, en el caso de que las variables sean categóricas y el estadístico F en el caso de variables continuas, seleccionándose como segmentación óptima aquella de mayor nivel de significación. A continuación se selecciona la variable predictora con mayor nivel de significación para efectuar la segmentación del nodo.

Kass utiliza la desigualdad de Bonferroni para estimar la significación estadística de cada variable. Sin embargo, su método de estimar los ajustes de Bonferroni tiende a favorecer particiones simples y a discriminar contra las variables con escala nominal y elevado número de categorías sobre otro tipo de variables. En el método CHAID Exhaustivo se propone un procedimiento alternativo, como se detalla a continuación.

a) Algoritmo CHAID

El algoritmo del método CHAID Exhaustivo puede resumirse en los siguientes pasos:

1. Para cada variable predictora, se busca el par de categorías cuya diferencia presente la menor significación estadística (es decir, aquella que presenta el menor nivel crítico) respecto a la variable criterio. El método utilizado para calcular el nivel crítico depende del tipo de variable que sea la variable criterio:
 - Si es una variable continua, utilizar la prueba F .

- Si es una variable nominal, se forma una tabla de contingencia con las categorías de la variable predictora como filas y las categorías de la variable criterio como columnas. Se utiliza la prueba chi-cuadrado de Pearson o la prueba de razón de verosimilitud.
 - Si es una variable ordinal, se ajusta un modelo de asociación de la variable criterio (Clogg y Eliasin, 1987; Goodman, 1979 y Magdison, 1992). Se utiliza la prueba de razón de verosimilitud.
2. Fundir en una categoría compuesta la pareja que proporcione el mayor nivel crítico.
 3. Calcular el nivel crítico basado en el nuevo conjunto de categorías de la variable predictora.
 4. Repetir los pasos anteriores hasta que sólo queden dos categorías. Después, de entre todos los conjuntos de categorías de la variable predictora, buscar aquel cuyo nivel crítico sea menor en el tercer paso.
 5. Calcular el nivel crítico corregido mediante Bonferroni para el conjunto de categorías de la variable predictora y de la variable criterio.
 6. Seleccionar la variable predictora cuyo nivel crítico corregido sea el menor (es decir, aquella que sea la más significativa). Comparar su nivel crítico con el nivel α preestablecido (α_{corte}):
 - Si el nivel crítico es menor o igual que el α_{corte} , dividir el nodo conforme al conjunto de categorías de la variable predictora.
 - Si el nivel crítico es mayor que el α_{corte} , no dividir el nodo. Este nodo constituye un nodo terminal.
 7. Continuar con el proceso de desarrollo del árbol hasta que se cumpla una de las reglas de parada que se definen posteriormente.

El procedimiento descrito debe repetirse para cada variable predictora.

Un paso adicional nos permitiría obtener el número óptimo de categorías en las que segmentar cada nodo en función de una determinada variable predictora. Después de realizar el paso (1), en caso de haber obtenido un número de categorías mayor o igual que tres en un determinado nodo, se intentará buscar una segmentación binaria que proporcione un valor del estadístico χ^2 o F superior al obtenido con la segmentación inicial realizada en el paso 1. En caso de que exista, se debe segmentar el grupo y continuar con el resto de pasos.

b) Factor de Ajuste de Bonferroni

La última fase del algoritmo propuesto requiere la utilización de un test para evaluar la significación de las variables predictoras. Como se ha dicho, para ello puede utilizarse el estadístico χ^2 , test que depende del número de categorías de la variable predictora. Sin embargo, Kass comprobó que los valores críticos de Bonferroni eran similares a los obtenidos mediante un procedimiento exacto, proponiendo su utilización en aquellos casos en los que se desconocen los resultados exactos.

El factor de ajuste de Bonferroni toma en consideración el hecho de que el test de significación se basa en la agrupación de categorías con el mayor nivel de significación, lo cual implica ser bastante estricto en las agrupaciones iniciales si se desea mantener el valor del error de primera especie (probabilidad de encontrar una segmentación significativa cuando realmente no existe relación entre la variable respuesta y la variable predictora) por debajo de un cierto nivel crítico, α . La desigualdad de Bonferroni, proporciona un método de fijar el nivel de significación para la mejor segmentación encontrada de una determinada variable predictora, conservando el valor del error de primera especie, por debajo de un cierto nivel crítico α (Miller, 1981).

Teóricamente, una vez determinada la mejor forma de agrupar las categorías de la variable predictora, el factor de ajuste de Bonferroni debería proporcionarnos el número total de posibles formas de distribuir las c categorías de la variable predictora en k grupos

($2 \leq k \leq c$). El nivel de significación, α_0 , de un test simple al cual debe evaluarse la mejor agrupación hallada, será según la desigualdad de Bonferroni:

$$\alpha_0 = \frac{\alpha}{N_B(c)}$$

donde,

- α es el error de tipo de primera especie establecido.
- $N_B(c) = \sum_{k=2}^c N(c, k)$ es el factor de ajuste de Bonferroni para todas las posibles formas de agrupar las categorías de la variable predictora.
- $N(c, k)$, es el número de posibles maneras de agrupar las c categorías en k grupos.

El valor de $N(c, k)$ puede ser muy elevado (superior a 120000), en el caso de variables predictoras con un número de categorías superior a 10 y además, en ese caso, para asegurarse de que existe una relación real, la mejor forma de agruparlas debe ser significativa a un nivel de significación del 0,0000004 (Biggs et al., 1991). La aplicación del factor de ajuste de Bonferroni para todas las posibles formas de agrupar las categorías de la variable, nos conducirá a obtener un test de significación demasiado conservador, que puede proporcionar elevados valores del error de segunda especie (probabilidad de no detectar una posible segmentación cuando sí que existe relación entre la variable predictora y la variable respuesta).

Kass calculó el ajuste de Bonferroni, asumiendo que el número de grupos k es fijo e igual al número final de grupos obtenidos mediante otros test previos. Sin embargo, la posibilidad de fijar el valor de k es cuestionable, ya que disminuye el factor de ajuste y, en consecuencia, aumenta la significación de las variables predictoras. Esto último puede ser contraproducente, ya que como las variables predictoras son ordenadas en función de su nivel de significación, es importante que los niveles de significación estimados no favorezcan a unas variables u otras.

Partiendo de la propiedad de que el ajuste de Bonferroni, no descarta una posible agrupación de categorías porque otra similar haya resultado no significativa y de que el algoritmo de búsqueda propuesto trabaja basándose en la estimación por etapas, es posible obtener otro factor de ajuste menos conservador.

Supongamos que en cada etapa, a la que llamaremos r , hay $(c-r+1)$ grupos y se combinan los dos grupos de características más similares. De esta manera, la primera etapa reduciría las c categorías iniciales a $(c-1)$ y la etapa $(c-2)$ reduciría el número de agrupaciones a 2. En este sentido el número de agrupaciones que obtendremos en la etapa r , puede estimarse mediante (Kass, 1980):

- $N(c-r+1, c-r) = c-r$ en el caso de variables con escala ordinal. (1a)

- $N(c-r+1, c-r) = \sum_{v=0}^{c-r-1} (-1)^v \frac{(c-r-v)^{c-r+1}}{v!(c-r-v)!}$ en el caso de variables nominales. (1b)

Asumiendo una variable respuesta de tipo categórico y denominando S_{rj} a las j posibles agrupaciones consideradas en la etapa r , donde $j = 1, \dots, N(c-r+1, c-r)$, podemos definir

$$I(S_{rj}) = \begin{cases} 0 & \text{si } X^2(S_{rj}) < x_r \\ 1 & \text{si } X^2(S_{rj}) \geq x_r \end{cases}$$

donde x_r satisface,

$$P(X_r^2 \geq x_r / H_0) = \alpha_0$$

y X_r^2 tiene una distribución χ^2 con los grados de libertad correspondientes a la etapa r . En el caso de variables continuas el estadístico χ^2 puede ser sustituido por el estadístico F.

Si denominamos S_r^* , a la agrupación encontrada con mayor valor de χ^2 (o F), podemos establecer que:

$$I(S_r^*) = 0 \text{ equivale a } \sum_j I(S_r) = 0 \quad (2)$$

La hipótesis nula, H_0 , que se desea contrastar es: “no existe relación entre la variable respuesta y las variables predictoras”. El test de H_0 para el conjunto completo de las c categorías podemos definirlo como aceptar H_0 si $I(S_{01}) = 0$ y el error de tipo de primera especie en este test quedaría como:

$$P(I(S_{01}) = 1 / H_0) = \alpha_0$$

Consideremos en cada etapa, que la probabilidad de cualquier posible agrupación tiene un valor significativo del estadístico χ^2 , dado que no se encontraron agrupaciones significativas en etapas previas y, H_0 es α_r :

$$P(I(S_{rj}) = 1 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0) = \alpha_r \quad (3)$$

El conjunto de agrupaciones $\{S_{rj}\}$ viene restringido por el número de agrupaciones encontradas en las etapas previas, al crearse mediante fusión de pares de grupos, la mejor agrupación posible hallada en dicha etapa. Si la mejor agrupación posible en la etapa previa ha sido detectada aleatoriamente, $\{S_{(r-1)j}\}$, se disminuye la probabilidad de encontrar agrupaciones significativas en la etapa actual, lo que conllevaría a que α_r debería ser menor que $\alpha_{(r-1)}$. Sin embargo, si la mejor agrupación ha sido hallada mediante fusión de los grupos más diferentes en la etapa previa, se incrementaría la posibilidad de que la agrupación fuese significativa en la etapa actual. Estos dos efectos, no pueden ser fácilmente demostrados, ya que no es posible que sucedan simultáneamente. Así, si se cumple:

$$P(I(S_{rj}) = 1 / \sum_j I(S_{(r-1)j}) = 0 \text{ y } H_0) \approx \alpha_0$$

en cada etapa la probabilidad de que la mejor agrupación posible, S_r^* , no sea significativa, conocido que en la etapa anterior no era significativa y conocido H_0 , puede hallarse utilizando la inecuación de Bonferroni:

$$P\left(\sum_j I(S_{rj}) = 0 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0\right) \geq 1 - \sum_j P(I(S_{rj}) = 1 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0) \geq 1 - Nc - r + 1, c - r \alpha_0$$

Utilizando un estimador del error de primera especie igual al límite superior especificado y empleando la equivalente relación en la ecuación (2), quedaría:

$$P(I(S_r^*) = 1 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0) = \alpha_0 N(c - r + 1, c - r)$$

De manera similar la inecuación de Bonferroni puede ser utilizada para determinar la probabilidad de no encontrar la mejor agrupación, S_r^* , significativa en ninguna etapa ($r = 1, \dots, (c-2)$):

$$\begin{aligned} P(I(S_r^*) = 0 / H_0) &= P\left(\sum_r \sum_j I(S_{rj}) = 0 / H_0\right) = \\ &P\left(\sum_r \left[\sum_j I(S_{rj}) = 0 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0 \right]\right) = \\ &P\left(\sum_r \left[I(S_r^*) = 0 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0 \right]\right) \geq \\ &1 - \sum_{r=1}^{c-2} P(I(S_r^*) = 1 / \sum_i I(S_{(r-1)i}) = 0 \text{ y } H_0) \geq \\ &1 - \sum_{r=1}^{c-2} \alpha_0 N(c - r + 1, c - r) \geq 1 - \alpha_0 N_{BC}(c) \end{aligned} \quad (4)$$

donde $N_{BC}(c)$ es el factor de ajuste de Bonferroni para una variable categórica de c categorías y se calcula como:

$$N_{BC}(c) = 1 + \sum_{r=1}^{c-2} N(c-r+1, c-r) \quad (5)$$

El test de hipótesis nula, H_0 , utilizando todas las agrupaciones y considerando una variable respuesta categórica, quedaría como:

$$\text{Aceptar } H_0 \text{ si } X^2(S_{rj}) < x_r, \forall r \text{ y } j \quad (6)$$

donde $X^2(S_{rj})$ es el valor de χ^2 para la agrupación S_{rj} , con los grados de libertad correspondientes y x_r se elige de manera que el ratio del error de primera especie sea α^* . De la ecuación (4), se obtiene que α^* cumple:

$$\alpha^* = P(X^2(S_{rj}) \geq x_r), \forall r \text{ y } j / H_0 \leq N_{BC}(c)\alpha_0$$

Es decir, que el error de primera especie del test global debe ser menor o igual que α^* si x_r es elegido de manera que se cumpla:

$$\alpha_0 = \frac{\alpha^*}{N_{BC}(c)}$$

siendo x_r el valor crítico al nivel α_0 de una χ^2 con los grados de libertad correspondientes a la etapa

El desarrollo mostrado anteriormente, es aplicable cuando trabajamos con una sola variable. Sin embargo, en la mayoría de los casos trabajamos con varias variables predictoras simultáneamente, seleccionando la que resulta más significativa para realizar la primera segmentación del nodo. En este caso, es también posible utilizar el factor de ajuste de Bonferroni, para estimar la significación estadística de cada variable predictora y seleccionar la que resulte más significativa. En este caso, el factor de ajuste en el nodo t vendría definido por:

$$N_{BV}(t) = N_V - N_n(t)$$

donde N_V es el número de variables predictoras y $N_n(t)$ el número de niveles o segmentaciones realizados hasta alcanzar el nodo t .

El error de primera especie asociado al nivel de significación, vendría definido por:

$$\alpha_0 = \frac{\alpha}{N_{BV}(t)N_{BC}(c)}$$

Este factor de corrección de Bonferroni no permite, sin embargo, evaluar la correlación entre las variables predictoras, siendo recomendable aplicar previamente otras técnicas para determinar las variables independientes que expliquen un mayor porcentaje de varianza de las variables predictoras.

c) Tamaño del árbol

La correcta aplicación del método de clasificación mediante el algoritmo CHAID exhaustivo, requiere especificar cuando debemos dejar de segmentar, es decir, cual es el tamaño de árbol más adecuado. Para ello, deben definirse los siguientes parámetros:

- **Profundidad máxima del árbol.** Permite controlar la profundidad (número de niveles bajo el nodo raíz) del árbol generado.
- **Número mínimo de casos.** Permite especificar el número mínimo de casos para los nodos. Los nodos que no cumplen estos criterios no se dividen.
- **Nodo parental.** Permite especificar el número mínimo de casos en un nodo parental. Los nodos con menos casos no se dividen.
- **Nodo filial.** Permite especificar el número mínimo de casos en los nodos filiales. El nodo no se divide si, al efectuar la división del mismo, el número de casos en los nodos filiales resultantes es menor que este valor.

- **Información de la partición.** Muestra el número de semilla aleatoria utilizado, el número total de casos y el porcentaje de casos en los conjuntos de datos de entrenamiento y de comprobación.

d) *Precisión del árbol*

Después de generar un árbol, siempre es importante considerar su *precisión*. En el caso de que la variable criterio sea de tipo nominal, la precisión indica el porcentaje de clasificación correcta. En ocasiones, la precisión se denomina validez predictiva. Por otra parte, la imprecisión se llama *riesgo*. La estimación del riesgo y el error típico de la estimación del riesgo (ET de la estimación del riesgo), son valores que indican la capacidad del clasificador.

CHAID Exhaustivo puede estimar el riesgo del árbol mediante tres métodos distintos:

- La reestimación de toda la muestra.
- La creación de un conjunto de entrenamiento y otro de comprobación.
- La validación cruzada

El método más sencillo para calcular el riesgo es la reestimación, pero normalmente subestima el riesgo verdadero. La partición de los datos en dos subconjuntos, uno para el entrenamiento y otro para la comprobación, es un buen método cuando el conjunto de datos es suficientemente grande. El riesgo se calcula a partir de la muestra de comprobación. La validación cruzada implica dividir la muestra en una serie de muestras más pequeñas. A continuación, se generan los árboles, que no incluyen los datos de cada submuestra. Para cada árbol se calcula el riesgo de clasificación errónea aplicando el árbol a la submuestra que se excluyó al generarlo. La estimación de riesgo mediante validación cruzada para todo el árbol se calcula como el promedio de los riesgos de todos los árboles.

Cuando se utiliza algún procedimiento de validación los casos se asignan de manera aleatoria a las particiones o submuestras. En estos casos, CHAID Exhaustivo

dispone de un parámetro denominado semilla aleatoria que es útil si se desea poder duplicar en forma exacta la partición en cada sesión ya que los conjuntos definidos con los mismos números de semilla aleatoria siempre asignaran los mismos casos a las mismas particiones.

La configuración de la semilla permite, en consecuencia, especificar el valor inicial que utiliza el generador de número aleatorio para asignar casos. El valor utilizado en nuestro caso es de 2.000.000. por ser el más recomendado (Biggs et al., 1991).

e) Análisis del Árbol

En el caso de que la variable criterio sea categórica, la denominada tabla de ganancias proporciona los estadísticos de los nodos que describen el árbol de clasificación respecto a la categoría considerada de la variable criterio. Si la variable criterio es continua, la tabla de ganancias proporciona estadísticos de nodos respecto a la media de la variable criterio.

Los estadísticos proporcionados son:

- *Nodo*, que nos indica el nodo que se está analizando.
- *Nodo: n*, indica el número de casos de la variable criterio y la categoría criterio capturados por ese nodo.
- *Nodo: %*, indica el porcentaje de casos de la variable criterio y la categoría criterio capturados por ese nodo.
- *Resp: n*, indica el número de casos que han sido identificados como la correspondiente categoría criterio.
- *Resp: %*, indica el porcentaje de casos que han sido identificados como la correspondiente categoría criterio respecto del total.
- *Ganancia (%)*, muestra el porcentaje de casos del nodo que representan el valor criterio en la variable criterio.

- *Índice (%)*, nos indica la composición del nodo comparándola con la de toda la muestra. Se obtiene dividiendo la ganancia (%) por el porcentaje de la categoría criterio del nodo raíz y multiplicando por cien.

7.3. Análisis Discriminante

Una vez obtenida la segmentación de los datos en función de las distintas variables predictoras, se efectuó un LDA para evaluar cuales de los ítems del cuestionario caracterizan mejor cada uno los grupos obtenidos mediante el método de clasificación CHAID Exhaustivo e identificar los que mejor sirven para discernir los perfiles de tipologías docentes existentes dentro de cada uno de los grupos obtenidos.

Para ello, partimos de los resultados obtenidos en el LDA efectuado en el capítulo 6, en el que obtuvimos los ítems de mayor poder discriminante entre los tres perfiles obtenidos mediante el análisis clúster. Utilizamos estos ítems como variables independientes y como variable dependiente, una nueva escala que se definirá posteriormente y que es una categorización de las puntuaciones medias del ítem criterio.

El procedimiento de análisis discriminante utilizado ha sido el método por etapas empleando como criterio para la inclusión o eliminación de variables en cada etapa la Distancia de Mahalanobis, por las ventajas que presentaba frente a la estimación en un solo paso, ya comentadas en el capítulo anterior.

7.4. Método SIMCA

El método SIMCA (Soft Independent Modelling of Class Analogy) es un método bilineal (BLM: *bi-linear modelling*) multibloque de clasificación que, como se ha comentado, se suele aplicar cuando las muestras que pertenecen a las clases son demasiado diferentes entre sí como para formar un solo modelo bilineal. Es una técnica de PR (Pattern

Recognition) supervisado, que fue formulada por primera vez por Wold en 1976 (Wold, 1976). Es un método flexible (soft) ya que no asume ninguna hipótesis sobre la distribución de las variables.

Con este método se realizan modelizaciones independientes mediante ACP para cada una de las clases. Después, cada muestra es ajustada en un modelo ACP para chequear la separación entre clases. Con ello se puede posteriormente clasificar nuevas muestras cuya clase es desconocida.

Más concretamente, el método SIMCA utiliza como técnica básica el análisis de componentes principales, ACP, para extraer variables latentes y proporcionar una visualización más sencilla y resumida de la información de un conjunto de datos. Éste es precisamente uno de los principales objetivos de la modelización multivariante. Una importante propiedad del método BLM es su capacidad de mitigar los efectos negativos causados por el problema de colinealidad. Al tratar con datos como los utilizados en este trabajo, este hecho se convierte en algo muy positivo, debido precisamente a que tenemos bastantes variables correlacionadas entre sí.

En los capítulos anteriores, las caracterizaciones y clasificaciones multivariantes se han realizado mediante LDA (Linear Discriminant Analysis), HCA (Hierarchical Cluster Analysis), etc., que pueden considerarse como técnicas de modelización “hard”, que se caracterizan por que cada observación se clasifica en uno y sólo uno de los grupos establecidos para la clasificación. Además los métodos hard requieren el cumplimiento razonable de una serie de supuestos teóricos. El primero de ellos es la normalidad multivariante. Los datos que no la cumplen pueden causar problemas en la estimación de la función discriminante. El segundo supuesto es que las matrices de covarianzas, aunque desconocidas, deben ser iguales para los grupos. El incumplimiento de este supuesto puede afectar desfavorablemente al proceso de clasificación. Si los tamaños muestrales son pequeños y las matrices de covarianzas son distintas, la significación estadística del proceso de estimación se ve afectada negativamente.

Frente a estas técnicas rígidas, actualmente se ha comenzado a utilizar técnicas flexibles (soft modelling), que se basan en la variación y correlación de los datos, recogidos

en la matriz de covarianza. Son métodos soft debido a que requieren muy pocas o ninguna hipótesis previas para construir el modelo. Estas técnicas de modelización soft constituyen una forma de trabajar más pragmática que permiten extraer la principal información sobre las relaciones contenidas en una tabla de datos (Martens y Martens, 2001).

En la figura 7.2 se representa el procedimiento SIMCA. La modelización de las clases o categorías se realiza independientemente para cada una de ellas mediante la construcción de un modelo ACP. Este modelo utiliza el número óptimo de componentes principales que mejor describe y agrupa una clase individual. Generalmente se obtienen después de autoescalar separadamente los datos. Las componentes principales se obtienen normalmente mediante el algoritmo NIPALS (*non iterative partial least squares*), que se describe a continuación, con los datos autoescalados con la media y la desviación típica de su categoría. Los modelos construidos para las distintas clases se comparan estudiando sus diferencias y analogías.

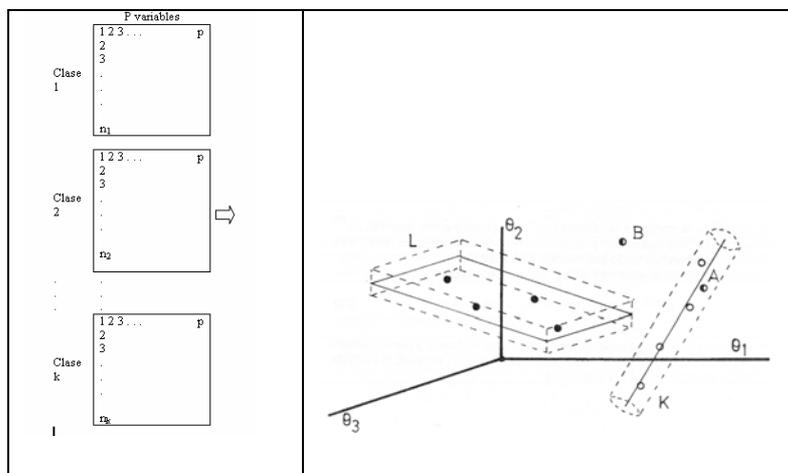


Figura 7.2. Soft Independent Modelling of Class Analogy

Para cada clase, c , el número de componentes principales significativos, A_c , se obtienen mediante alguno de los algoritmos de validación cruzada (normalmente NIPALS) y puede ser diferente en cada categoría. Este número de componentes principales determina la dimensionalidad del “espacio interno”, es decir, el espacio de la información

estructurada; las otras (p-A_c) componentes son el “espacio externo” o ruido, siendo p el número de variables que se han observado.

Debido a sus características, algunos autores como Frank y Friedman, (1989) y McLachlan, (1992) entre otros, lo señalan como uno de los mejores métodos de clasificación.

a) Algoritmo NIPALS

El algoritmo NIPALS (*nonlinear iterative partial least squares*) es un proceso iterativo para obtener las A componentes principales de un ACP. Fue desarrollado en 1976 por Herman Wold y ha sido desde los años setenta un método estándar de cálculo para los métodos BLM (ACP y PLSR sobre todo). La historia del algoritmo NIPALS puede encontrarse en Esbensen y Geladi (1990), Geladi y Esbensen (1990), Geladi y Kowalski (1986) y Wold et al. (1987).

Las etapas del algoritmo son las siguientes:

0. Centrar y escalar apropiadamente, si es necesario, la matriz X.
1. Se inicia el algoritmo tomando el vector \vec{t} como cualquiera de las p variables de la matriz X, aunque es conveniente tomar la columna de mayor amplitud, es decir, la que cumpla $\max|\vec{X}_i|$.
2. Se utiliza t para predecir el vector de loadings \vec{p} de acuerdo con la expresión $\vec{p} = X'\vec{t} / |X'\vec{t}|$. Geométricamente, este paso corresponde a la proyección de los vectores columna de la matriz X sobre la dirección de \vec{t} , en el espacio de las observaciones.
3. Se utiliza el vector \vec{p} para predecir el vector de scores $\vec{t} : \vec{t} = X\vec{p}$. El paso 3 corresponde a la proyección de los vectores fila de X sobre la dirección de \vec{p} en el espacio de las variables.

4. Se comprueba la convergencia: si $|\vec{t}_{nuevo} - \vec{t}_{anterior}| < criterio$ se para; en otro caso volver al paso 2.
5. Se actualiza $X = X - \vec{t}\vec{p}'$ y se vuelve al paso 1.

Este proceso se repite A veces (número de componentes principales a extraer), ya que cada vez que se completa una iteración, por haberse logrado la convergencia se extrae una componente principal.

Las matrices de puntuaciones T y cargas P están formadas por los vectores columna \vec{t} y \vec{p} obtenidos en cada iteración. La matriz de residuos E es el resultado de restar a X la parte explicada por las componentes:

$$E = X - \vec{t}_1\vec{p}'_1 - \vec{t}_2\vec{p}'_2 - \dots - \vec{t}_A\vec{p}'_A$$

En cuanto al paso 4, decir que el algoritmo NIPALS con frecuencia converge hacia un vector \vec{t} estable en menos de 20 a 40 iteraciones, con un criterio por ejemplo de 10^{-6} o menos.

Una de las características del algoritmo NIPALS es que los componentes principales se calculan uno cada vez. NIPALS utiliza este proceso iterativo de cálculo de las componentes principales para trabajar directamente con la matriz X. Esto es una aproximación numérica al análisis bilineal que se diferencia de otros métodos como los llamados de diagonalización directa, cuya descripción puede encontrarse en (Martens y Martens, 2001).

b) Validación

Para cada una de las observaciones i se define el estadístico (con frecuencia llamado simplemente SPE_i):

$$Q_i = \sum_{k=1}^p e_{ik}^2$$

que mide la parte de la variabilidad total de la observación no recogida por las componentes. Este estadístico mide el ajuste del modelo a la observación i , utilizando la distancia euclídea de esa observación a su proyección en el hiperplano formado por las nuevas componentes del modelo.

La parte de la variabilidad de cada observación explicada por los factores viene dada por los scores. Se define el estadístico T_i^2 como la suma de cuadrados de los scores tipificados que proporciona una estimación de la distancia de Mahalanobis de la muestra:

$$T_i^2 = \sum_{a=1}^A t_{ia}^2 .$$

Del mismo modo se puede calcular la varianza residual de una variable k como:

$$e_k^2 = \sum_{i=1}^n e_{ik}^2$$

y la varianza total para las variables como:

$$e_{\text{tot},v}^2 = \sum_{k=1}^p e_k^2 .$$

La varianza residual total se calcula a partir de la matriz E , que es una medida del error del modelo. Esta varianza se suele llamar varianza del error de modelización o error de modelización (*modeling error*). En ACP el error de modelización significa la desviación entre el modelo y los datos reales X .

7.4.1 Procedimiento de aplicación del método SIMCA

Los dos pasos en la aplicación del método SIMCA son:

1. **Modelización:** Construcción de modelos separados para cada una de las clases.
2. **Clasificación de muestras:** asignar cada muestra a las clases y decidir si la muestra pertenece o no a la clase correspondiente.

▪ **Paso 1: Modelización**

En primer lugar, se realiza un autoescalado separado para cada una de las clases. El valor autoescalado de un dato x_{ijc} (i indica el objeto, j la variable y c la clase) es:

$$x_{ijc}^* = \frac{x_{ijc} - \bar{x}_{jc}}{s_{jc}}$$

donde \bar{x}_{jc} y s_{jc} son respectivamente, la media y desviación típica de la variable j para la clase c.

Algunas veces, después de autoescalar las clases, las variables son ponderadas. Los pesos pueden ser el resultado de experiencias previas, o de una primera aplicación de algún procedimiento del SIMCA, como el análisis del poder discriminante o del poder de modelización. En ocasiones, en lugar de autoescalar, se puede solamente centrar los datos o incluso trabajar con los valores absolutos (esto último mucho menos frecuente), dependiendo de las características de estos datos. En adelante se hablará de la matriz X a la que se le habrá realizado el pretratamiento más adecuado.

Como hemos dicho, para cada clase c se calculan sus componentes principales, A_c . A veces, después de este paso, los objetos anómalos, outliers, pueden ser detectados y eliminados de la matriz de datos. En este caso, el algoritmo vuelve a comenzar en el paso 1.

Cuando en el gráfico de componentes principales se observan desviaciones en la linealidad, pueden utilizarse transformaciones o productos cruzados de variables para mejorar el modelo.

Se calcula el rango de las puntuaciones (scores) en cada componente a ($a=1, \dots, A_c$). A este valor se le denomina Rango Normal.

La letra griega θ se usa con frecuencia en la literatura del método SIMCA para denotar las puntuaciones, a pesar de que esta letra se use fundamentalmente para parámetros poblacionales. Aquí, θ , la usaremos sólo para los límites de los rangos de las puntuaciones.

El rango normal se denota por NR y se define como:

$$NR \equiv [\theta_{ac}^{\min}, \theta_{ac}^{\max}]$$

siendo θ_{ac}^{\min} y θ_{ac}^{\max} los valores mínimo y máximo respectivamente de las puntuaciones de la componente principal a de la clase c.

Cuando SIMCA se emplea con conjuntos de datos con pocos objetos, el rango normal es una subestimación del verdadero rango. Por esta razón, se calcula el rango expandido, ER:

$$ER \equiv [\theta_{ac}^{\min} - s_{ac}, \theta_{ac}^{\max} + s_{ac}]$$

donde s_{ac} es la desviación típica de las puntuaciones de los objetos de la categoría c sobre la componente principal a. De esta forma el modelo de la clase c está definido por A_c componentes principales y por un rango para cada uno de ellos.

De acuerdo con la descomposición para cada clase:

$$X_c = T_c P'_c + E_c$$

donde X es la matriz de datos, T es la matriz de puntuaciones, P es la matriz de cargas y E representa la matriz de residuos.

Si el número de variables es p, el número de componentes principales de la clase c es A_c y el número de objetos es n_c , entonces la expresión algebraica equivalente a la anterior es:

$$x_{ijc} = \sum_{a=1}^{A_c} t_{iac} \cdot p_{ajc} + e_{ijc} \quad \forall i = 1, \dots, n_c \quad \forall j = 1, \dots, p$$

La varianza residual general de la clase c es una medida de la dispersión de los objetos en el conjunto de entrenamiento (training set) de la clase alrededor del modelo matemático. Se define como:

$$s_0^{2(c)} = \sum_{j=1}^p \frac{e_{ijc}^2}{(p - A_c) \cdot (n_c - A_c - 1)}$$

La varianza residual de un objeto individual i del conjunto de entrenamiento de la clase c , es la distancia al cuadrado de ese objeto al modelo SIMCA y se define como:

$$s_i^2 = \sum_{j=1}^p \frac{e_{ijc}^2}{(p - A_c)} \cdot \frac{n_c}{(n_c - A_c - 1)}$$

La distancia para otros objetos (conjunto de validación y objetos de otras categorías) quedaría:

$$s_i^{2*} = \sum_{j=1}^p \frac{e_{ij}^2}{(p - A_c)}$$

Notar que no se pierden grados de libertad cuando el objeto no se utiliza para construir el modelo.

a) Poder de modelización

El poder de modelización (*Modeling Power*) de una variable es una medida de la contribución de esa variable al modelo matemático de su clase.

La varianza inicial de una variable j en la clase c es:

$$s_j^2(x) = \sum_{i=1}^{n_c} \frac{(x_{ijc} - \bar{x}_{jc})^2}{n_c - 1}$$

La varianza residual de la variable j , respecto al modelo de la clase c es:

$$s_{jc}^2 = \sum_{i=1}^{n_c} \frac{e_{ijc}}{n_c - A_c - 1}$$

El poder de modelización, MP, de una variable j en la clase c se define como:

$$MP_j = 1 - \frac{s_{jc}^2}{s_j^2(x)}$$

Este valor está comprendido en el intervalo [0,1]. Cuanto más se acerca a 1 el valor de MP para una variable, en mayor grado es tenida en cuenta esa variable en el modelo de la clase y mayor es su influencia.

El poder de modelización puede utilizarse en ocasiones como un peso o ponderación en una construcción más detallada del modelo SIMCA. En problemas con una sola categoría, las variables con bajo poder de modelización pueden ser eliminadas

b) Poder discriminante

Algunos parámetros estadísticos como el poder discriminante, o la distancia entre clases, pueden calcularse sólo en el caso de que el problema tenga más de una categoría.

Sean c1 y c2 dos categorías. El cuadrado del poder discriminante, DP (llamado *discriminant power* o *classification power*), de la variable j entre las categorías c1 y c2 es:

$$DP_{j,c1,c2}^2 = \frac{s_{jc1c2}^2 + s_{jc2c1}^2}{s_{jc1}^2 + s_{jc2}^2}$$

donde la desviación típica de los objetos de la clase c2 respecto al modelo de la clase c1 es:

$$s_{jc1c2}^2 = \sum_{i=1}^{n_{c1}} \frac{e_{ij}^2(c1,c2)}{n_{c2}}$$

siendo $e_{ij}^2(c1,c2)$ el residuo para la variable j, del objeto i de la clase c2 respecto a la clase c1, como se muestra en la figura 7.3..

El poder discriminante de una variable indica la capacidad que tiene esa variable para distinguir entre los dos modelos. Así, una variable con un alto poder discriminante (con respecto a dos modelos) es muy importante para la diferenciación entre las dos clases correspondientes. El poder discriminante es una medida que debe ser comparada con 1. DP

toma el valor 1 cuando no existe en absoluto capacidad de discriminación en esa variable.

Suele aceptarse que a partir de un valor de DP superior a 3, la variable debe ser considerada como importante en la discriminación de clases.

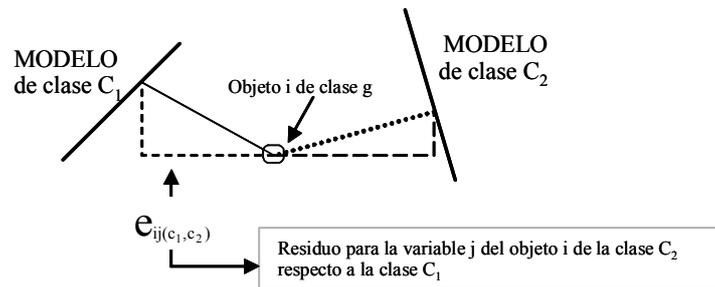


Figura 7.3. Clasificación en SIMCA

Al igual que el poder de modelización, el poder discriminante puede ser usado como una ponderación en una mejora de la construcción del modelo SIMCA, cuando el objetivo es mejorar la clasificación. En problemas con más de una clase, las variables que tienen un bajo poder de modelización y clasificación a la vez, pueden ser eliminadas.

c) Distancia entre clases

La distancia SIMCA al cuadrado entre dos categorías c1 y c2 puede calcularse como:

$$D_{c1c2}^2 = \left[\frac{\sum_{j=1}^p (s_{jc1c2}^2 + s_{jc2c1}^2)}{\sum_{j=1}^p (s_{jc1}^2 + s_{jc2}^2)} \right] - 1$$

Esta medida muestra cuán diferentes son dos modelos uno del otro y se calcula a partir de los resultados del ajuste de todas las muestras a cada clase (a su propio modelo y

al otro). Un valor grande de esta distancia implica que probablemente las dos clases serán bien distinguidas una de la otra.

▪ **Paso 2: Clasificación de objetos**

El espacio de clase SIMCA (llamado *SIMCA Box* en la literatura del método) es la región determinada por el modelo matemático de componentes principales. El rango expandido, ER, de cada componente proporciona los límites del modelo de la clase.

Las puntuaciones de los objetos en el conjunto de entrenamiento están dentro del rango de la componente. Para otros objetos (conjunto de validación o de otras clases), sus puntuaciones, en una o más componentes, pueden caer fuera del modelo de la clase. En este caso, se utiliza una distancia SIMCA aumentada \tilde{s}_i^2 del objeto i a la clase c , que se calcula, como:

$$\tilde{s}_i^2 = s_i^{2*} + \sum_a (t_{ia} - \theta_{ac}^{LIM})^2 \frac{s_0^{2(c)}}{s_{ac}^2}$$

donde el sumatorio se calcula sólo para las componentes en los que la puntuación correspondiente al objeto queda fuera del rango expandido y θ_{ac}^{LIM} puede representar el valor máximo o mínimo de las puntuaciones. Notar que la ponderación mediante el cociente s_0^2 / s_{ac}^2 evita que \tilde{s}_i^2 tenga el significado de la hipotenusa (Figura 7.4).

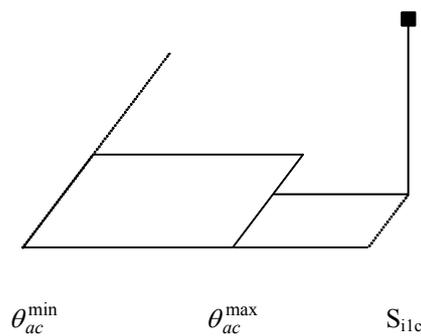


Figura 7.4. Clasificación en SIMCA

SIMCA aplica un test F para evaluar la probabilidad de que un objeto pertenezca a una clase, que quedaría definido como:

$$F_{(p-A_c), (p-A_c)(n_c-A_c-1)} = \frac{s_i^2}{s_0^2}$$

y utilizando la distancia aumentada \tilde{s}_i^2 , en lugar de s_i^2 , cuando sea necesaria.

Si F_α es el valor crítico de la función F para un nivel de significación α entonces:

$$s_{CRIT} = s_0 \sqrt{F_\alpha}$$

Este valor determina los límites del espacio de clase SIMCA, es decir, la región de confianza alrededor del modelo de componentes principales.

Teniendo en cuenta la distancia aumentada de los objetos en el conjunto de evaluación y validación y el factor $n_c / (n_c - A_c - 1)$ y los objetos en el conjunto de calibrado (cuyos scores no exceden el rango normal) el SIMCA box tiene la forma que se muestra en la figura 7.5.

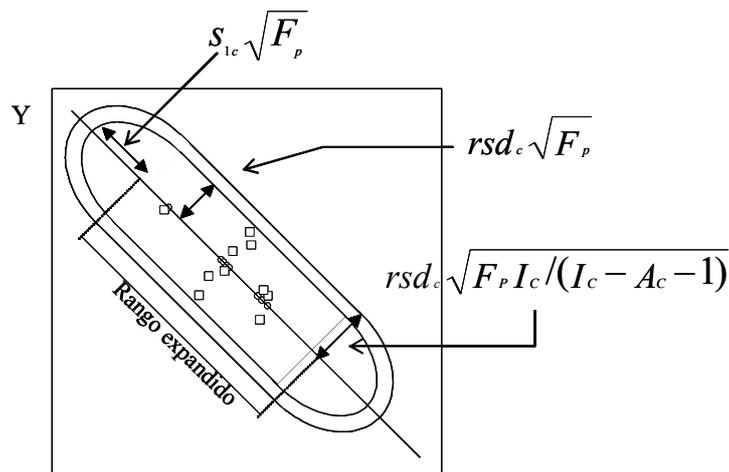


Figura 7.5. Método SIMCA

7.4.2 Clasificación con SIMCA

La clasificación se obtiene en base a la distancia del objeto que se desea clasificar respecto al modelo de la clase: el objeto se clasifica en la clase para la cual la distancia SIMCA (s_i o \tilde{s}_i) es mínima.

Los nuevos datos que se desean clasificar deben lógicamente estar descritos por las mismas variables que el conjunto de calibración (conjunto de muestras con el que han sido calculados los modelos para cada una de las clases). Si los datos del conjunto de calibración han sufrido algún tipo de transformación o pre-procesamiento, los nuevos datos deben ser transformados de la misma manera.

Un objeto cuya clase de pertenencia se desconoce, con un vector de datos \bar{x}_i se comprueba si pertenece a una clase determinada mediante regresión del vector \bar{x}_i en el modelo de clase c . Multiplicando el vector de datos por la matriz de cargas obtenemos una estimación de un vector de puntuaciones nuevo \bar{t}_i . Con el vector de puntuaciones se calculan los residuos y se decide a qué clase se adjudica el objeto:

$$\bar{t}_i = \bar{x}_i P$$

$$\bar{e}_i = \bar{x}_i - \bar{t}_i P'$$

Se calcula la varianza residual del objeto i como:

$$s_i^2 = \sum_{j=1}^p \frac{e_{ij}^2}{p - A_c}$$

El objeto es asignado a la clase c si las varianzas s_i^2 y s_0^2 son similares en magnitud. En caso de que s_i^2 sea mayor que s_0^2 , el objeto no es de la clase c . Esta comparación de varianzas se realiza con la razón F anterior. De forma equivalente, un objeto se clasifica en la clase c si está en el espacio de la clase c (para un determinado nivel de significación α).

Al clasificar nuevos objetos puede encontrarse alguna de estas situaciones:

- El objeto está clasificado únicamente en una clase, es decir, sólo hay un modelo donde pueda colocarse dentro de los límites críticos proporcionados. Además la distancia respecto a la siguiente clase más cercana es generalmente mucho mayor que la distancia al modelo en el que es aceptado el objeto.
- El objeto puede clasificarse en varias clases, es decir, tiene una distancia que está dentro de los valores críticos de varias clases simultáneamente. Esta situación puede deberse a dos razones: los datos son insuficientes para distinguir entre diferentes clases, o bien, el objeto pertenece a varias clases. Éste puede ser un elemento situado en la frontera o tener propiedades de varias clases. Si el objeto es clasificado en varias clases, pero quiere asignarse a una solamente, puede realizarse un estudio conjunto de la distancia (s_i) y el Leverage (H_i) para determinar en qué clase encaja mejor. (Este estudio se expone más adelante). Con distancias similares a las distintas clases, el objeto está más cercano al modelo para el que tiene un Leverage más bajo.
- El objeto no se clasifica dentro de ninguno de los límites críticos de las clases. Debe prestarse especial atención a este caso, ya que puede significar que el objeto procede de una clase distinta a las utilizadas en la clasificación o que simplemente es un dato anómalo

La posibilidad de clasificar un objeto en una, varias o ninguna clase, constituye uno de los potenciales más importantes del método SIMCA que le proporciona una poderosa capacidad de detectar clasificaciones erróneas.

Algunas herramientas para clasificación y detección de outliers son:

- El gráfico Coomans
- El gráfico distancia frente a Leverage
- El gráfico de la distancia normalizada (s_i / s_0) frente al Leverage.

a) *Gráfico de Coomans*

Es un gráfico donde se muestran las distancias SIMCA de un conjunto de muestras a dos modelos (Coomans et al., 1984). En este gráfico, como se muestra en la figura 7.6, se incluyen también los valores críticos s_{CRIT} para cada una de las clases, lo que permite visualizar qué muestras están dentro de una clase, de la otra, de ambas o de ninguna.

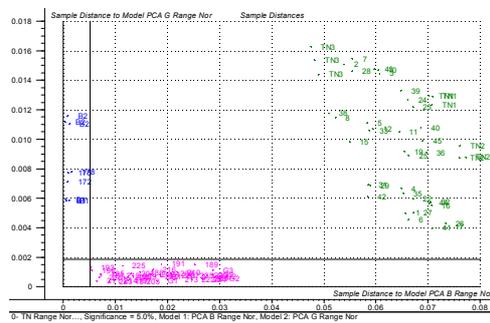


Figura 7.6. Gráfico distancia frente a Leverage (Si vs. Hi)

b) *Gráfico distancia frente a Leverage (s_i vs. H_i)*

Este gráfico representa cada uno de los objetos de una clase mediante su distancia al modelo (s_i) y su distancia al centro del modelo, llamada leverage (H_i). El leverage se calcula, para un objeto i , como:

$$H_i = \frac{1}{n_c} + \sum_{a=1}^{A_c} \frac{(s_{ia} - \bar{s}_a)^2}{\sum_{i=1}^{n_c} (s_{ia} - \bar{s}_a)^2}$$

donde \bar{s}_a es la media de las puntuaciones en la componente a .

H_i puede interpretarse como una medida de lo extremo que es un punto en comparación con el resto.

En este gráfico se muestran también los valores críticos para la distancia al modelo y para el leverage. El valor crítico para la distancia al modelo depende del nivel de

significación escogido y el valor crítico del leverage depende del número de componentes principales usados y es fijo para una determinada clasificación.

c) *Gráfico de la distancia normalizada frente al leverage (s_i/s_0 vs H_i)*

El cociente s_i/s_0 es la distancia relativa de una muestra al modelo, y se denomina distancia normalizada. Este gráfico es similar al anterior (s_i frente a H_i) y se utiliza con el mismo objetivo. Sin embargo, los valores del eje de ordenadas en el gráfico anterior están en valores absolutos mientras que en éste se representa el valor de s_i/s_0 , que indica la distancia relativa respecto a la distancia media representativa del modelo (s_0). Este gráfico se interpreta de la misma forma que el anterior.

7.4.3 Sensibilidad y Especificidad

Las técnicas de modelización asignan una región acotada en el espacio de las variables a cada una de las categorías. Esta región es el modelo de la clase. La calidad de cada una de ellas puede evaluarse por su sensibilidad, especificidad y capacidad de predicción. La sensibilidad de clase (*sensitivity*) es el porcentaje de observaciones de una clase que el modelo acepta correctamente. La especificidad de clase (*specificity*) se define como el porcentaje de observaciones de otras clases que son rechazadas correctamente por el modelo. De acuerdo con estas definiciones, la sensibilidad es una estimación del nivel de confianza del test y la especificidad es una estimación de la potencia.

Las técnicas de modelización deben obtener el máximo valor de especificidad y sensibilidad, a la vez que tener una capacidad de predicción para datos desconocidos. Es decir, se deben evitar sobreajustes causados por modelos con una tendencia a describir los datos actuales pero que no tienen una capacidad suficiente o no son estables en la predicción.

7.4.4 Ventajas del SIMCA sobre otros métodos estadísticos clásicos

El método SIMCA presenta muchas ventajas importantes frente a la utilización de otros métodos clásicos como por ejemplo el LDA o el análisis cluster.

En primer lugar, esta técnica no está limitada a las situaciones en las que el número de objetos es significativamente mayor que el número de variables, como ocurre en las técnicas estadísticas clásicas. Los modelos bilineales basados en variables latentes son estables con respecto a una relación desequilibrada entre objetos y variables, tanto si hay muchas más variables como al contrario. Debido al producto externo entre puntuaciones y cargas de los modelos bilineales, toda la estructura de una matriz de datos puede ser razonablemente bien modelada incluso en los casos en los que una dimensión de la matriz de datos es mucho más pequeña que la otra.

El método SIMCA también se utiliza con frecuencia como herramienta de clasificación rígida (hard modeling), es decir, donde los objetos sólo pueden pertenecer a una clase. Sin embargo, el método tiene, como hemos visto, una capacidad mucho mayor y es aplicable a los casos donde los objetos pertenecen a más de una clase. Además, las colinealidades son manejadas más fácilmente.

Otra ventaja es que todos los resultados pueden representarse gráficamente, lo que favorece una comprensión intuitiva de la estructura subyacente a los datos.

7.5. Resultados y Discusión

7.5.1. Definición de tipologías docentes

La técnica de árboles de decisión, ha sido aplicada como técnica complementaria al análisis clúster realizado en el capítulo anterior, con la finalidad de segmentar los datos, determinar las variables predictoras más significativas para determinar el clúster de

pertenencia de las distintas observaciones y poder clasificar observaciones futuras en función de dichas variables descriptivas.

Para ello, partimos de los resultados obtenidos en el capítulo 6 mediante la aplicación del análisis clúster two-step, considerando únicamente como variables independientes los ítems que habían sido incluidos en el análisis discriminante efectuado sin considerar la variable criterio (R12, R4, R5, R1, R11, R10, R2, R3, R16, R8, R18 y R7), ya que posteriormente este ítem global es utilizado para crear una nueva variable categórica y efectuar un análisis discriminante en cada uno de los nodos obtenidos.

Se utilizó como variable criterio o variable respuesta, una nueva variable categórica denominada “qcldis2”, que nos indica el cluster de pertenencia de cada observación tras efectuar el análisis clúster. Como variables predictoras se utilizaron las características descriptivas de la asignatura y del profesor que la impartía: titulación, departamento, curso, semestre en el que se imparte, categoría del profesor, si es doctor o no, edad y dedicación. De esta manera, conseguiremos segmentar la muestra en función de las variables predictoras más significativas (aquellas con un valor de *p ajustado* más pequeño), estableciendo los grupos de individuos que pertenecen a cada categoría, diferenciados con respecto a esas variables predictoras.

Previo a la realización del análisis, nos planteamos la necesidad de dividir la base de datos en tres grupos en función de que la titulación sea técnica, superior o de segundo ciclo, ya que dependiendo del tipo de titulación que estemos considerando, puede variar el enfoque o la situación de los alumnos respecto a algunas de las características descriptivas de las asignaturas. Así, por ejemplo, el tercer curso en una titulación técnica sería el último mientras que en una titulación superior sería intermedio y, en consecuencia, el grado de valoración que los alumnos presentan en cada caso, puede verse afectado por esta circunstancia.

En todos las situaciones, para determinar el tamaño más adecuado del árbol definimos como reglas de parada a la hora de segmentar, una profundidad máxima del árbol de 5 niveles bajo el nodo raíz. El estudio de divisiones sucesivas ayudará poco a comprender mejor el problema, ya que las divisiones sucesivas incluyen en muchas

ocasiones, un número de casos muy reducido. Para verificar la precisión predictiva del mismo efectuamos una validación cruzada con 10 submuestras y definiendo un valor de semilla aleatoria de 2000000, por lo comentado anteriormente.

7.5.1.1. Titulaciones superiores

Como resultado de la aplicación del método CHAID Exhaustivo con los criterios descritos anteriormente, se ha obtenido un árbol con un único nivel de ramificación en función de la variable departamento y 6 nodos terminales. En la figura 7.7 se muestra el árbol obtenido.

Los nodos terminales permiten clasificar observaciones futuras. En las imágenes de los nodos se muestra el número del nodo, el número de observaciones y el porcentaje de cada categoría que han sido clasificadas en ese nodo y la categoría a la que se asignarían las nuevas observaciones (en sombreado gris).

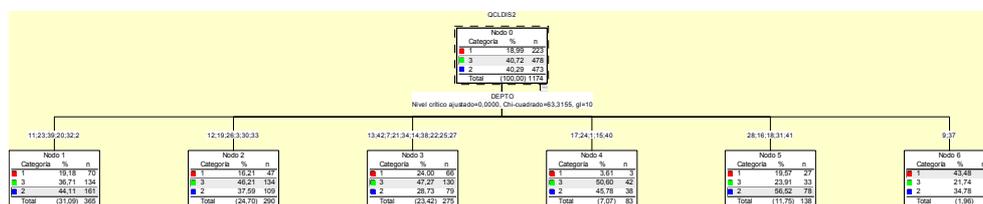


Figura 7.7. Árbol de clasificación

En el primer nodo, que se muestra en la figura 7.8, se obtuvieron un total de 365 observaciones correspondientes a los departamentos 11, 23, 39, 20, 32 y 2, perteneciendo el 19,18% de las mismas al cluster 1, en el que casi todos los ítems presentaban puntuaciones medias altas, el 36,71% al cluster 3, en el que casi todos los ítems presentaban puntuaciones medias bajas y el 44,11% al cluster 2, en el que predominaban puntuaciones intermedias en casi todos los ítems. Los departamentos agrupados en cada nodo son básicamente equivalentes desde el punto de vista estadístico, no existiendo diferencia estadística

significativa entre ellos. Las características de los nodos restantes se muestran en la tabla 7.1.

Nodo 1			
Categoría	%	n	
1	19,18	70	
3	36,71	134	
2	44,11	161	
Total	(31,09)	365	

Figura 7.8. Nodo 1 del árbol de clasificación

Tabla 7.1. Características de los nodos terminales del árbol de clasificación

	Departamentos	Clúster	Nº de observaciones	% de Observaciones
Nodo 1	11, 23, 39, 20, 32, 2	1	70	19,16
		2	161	44,11
		3	134	36,71
Nodo 2	12, 19, 26, 3, 30, , 33	1	47	16,21
		2	109	37,59
		3	134	46,21
Nodo 3	13, 42, 7, 21, 34, 14, 38, 22, 25, 27	1	66	24
		2	79	28,73
		3	130	47,27
Nodo 4	17, 24, 1, 15, 40	1	3	3,61
		2	38	45,68
		3	42	50,60
Nodo 5	28, 16, 18, 31, 41	1	27	19,57
		2	78	56,52
		3	33	23,91
Nodo 6	9, 37	1	10	43,48
		2	8	34,78
		3	5	21,74

En ella se observa que los nodos 2, 3 y 4 presentan una mayor proporción de profesores que pertenecen al cluster 3, es decir, que poseen puntuaciones relativamente

Identificación y caracterización de categorías docentes

bajas en todos los ítems. Por otro lado, en el nodo 6, el 43,48% de las observaciones, que corresponden fundamentalmente a los departamentos 9 y 37, pertenecen al cluster 1, es decir, a profesores con puntuaciones relativamente altas en todos los ítems. Finalmente, el nodo 1 y 5, presentan una mayor proporción de observaciones en el clúster 2, es decir, en el que los profesores tenían puntuaciones intermedias en casi todos los ítems.

Destacar aquí, que prácticamente ninguno de los nodos obtenidos se asocia mayoritariamente con el clúster 1, correspondiente a profesores con puntuaciones medias altas, pero sí con el clúster 2 y especialmente con el 3, correspondiente a profesores con puntuaciones medias bajas. Esto pondría de manifiesto, en este caso, que sí existe una cierta asociación entre determinadas variables descriptivas y las valoraciones bajas, mientras que esta relación es prácticamente nula con las valoraciones altas. Debemos verificar si este hecho se mantiene en el resto de segmentaciones obtenidas en otras titulaciones y al efectuar la segmentación con otras variables predictoras.

Tras efectuar este primer análisis, se estudió lo que sucedía al desarrollar la segmentación de los datos seleccionando como variable predictora la variable curso y/o titulación, ya que por los resultados derivados de otros trabajos (Jornet y Suarez, 1997), ambas variables parecen tener relación con las puntuaciones medias que se obtienen en los cuestionarios de evaluación docente mediante opinión de los estudiantes. Los resultados obtenidos se muestran en las tablas y gráficos siguientes.

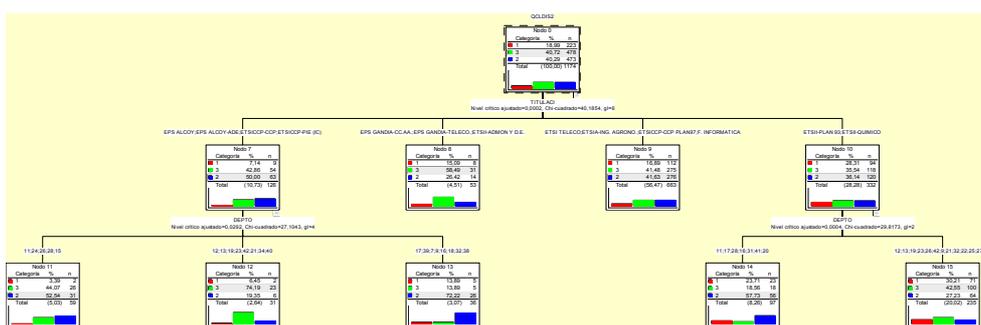


Figura 7.9. Árbol obtenido con variable predictora titulación

En el caso de utilizar como variable predictora la titulación, se obtuvo un árbol con dos niveles de ramificación. En el primer nivel de clasificación se obtuvieron cuatro nodos, con un nivel crítico ajustado de 0,0002 y un estadístico χ^2 igual a 40,1854 con 6 grados de libertad. El nodo 7 y el nodo 10, sufrían una segunda ramificación en función de la variable departamento.

En la tabla 7.2 se muestra el número de observaciones de cada uno de los nodos obtenidos mediante la primera ramificación y las titulaciones a las que pertenecían a cada uno de ellos. Comentar de nuevo, que las categorías agrupadas en un mismo nodo se pueden considerar básicamente equivalentes desde un punto de vista estadístico.

Tabla 7.2. Titulaciones agrupadas en cada nodo

	Titulaciones	Clúster	% de observaciones	Número de Observaciones
Nodo 7	EPS ALCOY	1	7,14	9
	EPS ALCOY ADE	2	50	63
	ETSII CCP-CCP	3	42,66	54
	ETSII CCP-CCP PIE (IC)			
Nodo 8	EPS GANDIA CCAA	1	15,09	8
	EPS GANDÍA TELECO	2	26,42	14
	ETSII ADMON Y D.E.	3	58,49	31
Nodo 9	ETSI TELECO	1	16,89	112
	ETSIA ING. AGRONOM.	2	41,63	276
	ETSII CCP-CCP PLAN 97	3	41,48	275
	F. INFORMÁTICA			
Nodo 10	ETSII PLAN 93	1	28,31	94
	ETSII-QUÍMICO	2	36,14	120
		3	35,54	118

En ella podemos detectar que el nodo 10 que clasifica a las observaciones pertenecientes a las titulaciones ETSII PLAN 93 Y ESTSII QUÍMICO, es el que mayor porcentaje de observaciones tiene en el clúster 1, correspondiente a profesores con puntuaciones medias altas, mientras que el nodo 8, correspondiente a las titulaciones de EPS GANDÍA CCAA, EPS GANDÍA TELECO Y ETSII ADMON y DE presenta un

58,49% de observaciones pertenecientes al cluster 3, correspondiente a profesores con puntuaciones medias bajas.

De nuevo se detecta que en todos los nodos existe cierta asociación entre determinadas características y el clúster 2 y 3 pero que ésta no se produce con el clúster 1.

Como se ha comentado, los nodos 7 y 10 se desglosan en función de la variable predictora departamento. En la tabla 7.3 se muestra la composición de los nodos terminales obtenidos mediante esta segunda ramificación. Sin embargo, los nodos 8 y 9 se definen como nodos terminales, no siendo posible dividirlos y mejorar el rendimiento del árbol.

Tabla 7.3. Titulaciones y departamentos agrupadas en los nodos tras la segunda ramificación

Nodo	Titulaciones	Departamento	Clúster	% de observaciones	Número de Observaciones
11	EPS ALCOY	11, 24, 26, 28, 15	1	3,39	2
	EPS ALCOY ADE		2	52,54	31
	ETSIICCP-CCP		3	44,01	26
	ETSICCP-CCP PIE (IC)				
12	EPS ALCOY	12,13,19,23,42,21,34,40	1	6,45	2
	EPS ALCOY ADE		2	19,35	6
	ETSIICCP-CCP		3	74,19	23
	ETSICCP-CCP PIE (IC)				
13	EPS ALCOY	17,39,7,9,16,18,32,38	1	13,89	5
	EPS ALCOY ADE		2	72,22	26
	ETSIICCP-CCP		3	13,89	5
	ETSICCP-CCP PIE (IC)				
14	ETSII PLAN 93	11,17,28,16,31,41,20	1	23,71	23
	ETSII-QUÍMICO		2	57,73	56
			3	18,56	18
15	ETSII PLAN 93	12,13,19,23,26,42,9,21,32,22,25,27,37	1	30,21	71
	ETSII-QUÍMICO		2	27,23	64
			3	42,55	100

En este caso, podemos observar que en el nodo 12, correspondiente a las titulaciones EPS ALCOY, EPS ALCOY ADE, ETSIICCP-CCP y ETSICCP-CCP PIE y a

los departamentos 12, 13, 19, 23, 42, 21, 34 y 40 el porcentaje de observaciones clasificadas en el clúster 3 se incrementa considerablemente, alcanzando un valor del 74,19%. En el nodo 13, correspondiente a las mismas titulaciones, se agrupan los departamentos 17, 39, 7, 9, 16, 18, 32 y 38, con una mayor proporción de observaciones pertenecientes al clúster 2 (72,22%).

En los nodos filiales del nodo 10, correspondiente a las titulaciones ETSII PLAN 93 y ETSII-QUÍMICO, podemos observar que mientras en el nodo 14 se incrementa el porcentaje de observaciones pertenecientes al clúster 2 y disminuyen las pertenecientes al clúster 3, en el nodo 15, se incrementan las observaciones en el clúster 1 y disminuyen las del 3.

En el caso de utilizar la variable curso como variable predictora, se obtuvo árbol con dos niveles de ramificación. En el primer nivel obtuvimos dos nodos, uno en el que se agrupaban las observaciones correspondientes a los cursos 1, 2, 3, 6 y asignaturas optativas (codificadas como 0) y otro en el que se clasificaban las observaciones correspondientes a los cursos 4 y 5. En el primer nodo, el porcentaje de observaciones pertenecientes al cluster 1 fue de 17,13% mientras que en el nodo 2, fue de un 22,9%, porcentaje ligeramente superior. Este resultado pone de manifiesto que no todos los profesores de asignaturas optativas reciben puntuaciones medias elevadas, siendo similares a las que se obtienen en los cursos de primero, segundo y tercero, en lugar de a los de cuarto y quinto, como se hubiese esperado en un primer momento. En ambos nodos, existía una segunda ramificación en función de la variable departamento, como se muestra en la figura 7.10.

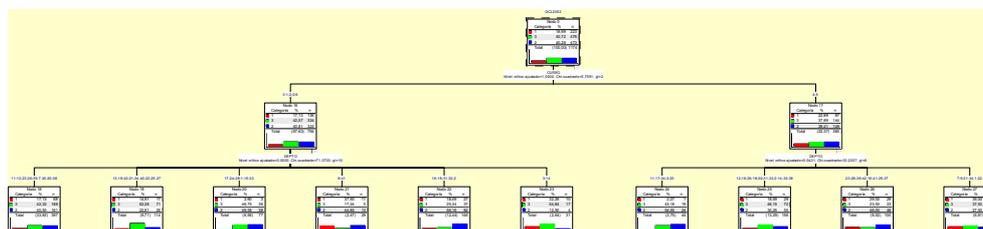


Figura 7.10. Árbol obtenido utilizando como predictor la variable curso

En la tabla 7.4 se muestran los resultados de todos los nodos finales.

Tabla 7.4. Cursos y departamentos agrupados en los nodos tras la segunda ramificación

Nodo	Cursos	Departamento	Clúster	% de observaciones	Número de Observaciones
18	0,1,2,3,6	11,12,23,26,39,7,30,20,38	1	17,13	68
			2	40,55	161
			3	42,32	168
19	0,1,2,3,6	13,19,42,21,34,40,22,25,27	1	14,91	17
			2	22,81	26
			3	62,28	71
20	0,1,2,3,6	17,24,28,1,15,33	1	3,90	3
			2	49,35	38
			3	46,75	36
21	0,1,2,3,6	9,41	1	37,93	11
			2	44,83	13
			3	17,24	5
22	0,1,2,3,6	16,18,31,32,2	1	18,49	27
			2	56,16	82
			3	25,34	37
23	0,1,2,3,6	3,14	1	32,26	10
			2	12,90	4
			3	54,84	17
24	4,5	11,17,24,33,20	1	2,27	1
			2	54,55	24
			3	43,18	19
25	4,5	12,19,26,18,30,31,32,2,14,33,38	1	18,59	29
			2	35,26	55
			3	46,15	72
26	4,5	23,28,39,42,16,41,25,37	1	29	29
			2	48	48
			3	23	23
27	4,5	7,9,21,34,1,22	1	35	28
			2	27,5	22
			3	37,5	30

Tabla 7.5. Titulaciones pertenecientes a cada nodo

	Titulaciones	Clúster	% de observaciones	Número de Observaciones
Nodo 1	EPS ALCOY ITI DISEÑO	1	8,60	8
	ETSICCP-OOPP	2	61,29	57
	ETSIICCP-OOPP PLAN 97	3	30,11	28
	EU ARQUITECTURA TÉC.			
Nodo 2	EPS ALCOY -ITI TELECO	1	18,34	53
	EU INFORMÁTICA	2	36,66	106
	EUIT INDUSTRIAL	3	44,98	130
Nodo 3	EPS ALCOY ITI E. IN.	1	37,06	63
	EPS ALCOY ITI ELEC.	2	17,06	29
	EPS ALCOY ITI QUIM.	3	45,88	78
	EPS ALCOY ITI TEXTIL			
	EPS GANDÍA FORESTAL			
	ETSIGCT TOPOGRAFÍA			
Nodo 4	EPS ALCOY ITI MEC.	1	22,63	43
	EPS GANDÍA ELEC.	2	24,74	47
	EPS GANDÍA SONIDO	3	52,63	100
	EUITA EXPL. AGR.			
	EUITA HORTO Y JARDIN.			
	EUITA IND .AGR.			
Nodo 5	EPS GANDÍA TURISMO	1	3,57	1
	ETSICCP-CCP PLAN 97	2	21,43	6
	EUITA MECAN y C.R.	3	75	21

En este caso, las titulaciones pertenecientes al nodo 5, presentan proporciones sensiblemente superiores de observaciones clasificadas en el cluster 3 que en el resto de nodos.

De nuevo todos los nodos presentan una mayor proporción de observaciones clasificadas en el cluster 2 y 3, lo que indicaría la existencia de cierta asociación entre puntuaciones medias y bajas y ciertas características descriptivas, no sucediendo así con las puntuaciones altas.

En la tabla 7.6 se muestra la composición de los nodos filiales del nodo 2, al dividirse en función de la variable departamento. En este caso observamos que los departamentos 27, 39, 19, 25, 16, 35 y 10 son los que presentan mayor proporción de observaciones en el clúster 3, con un 44,9% mientras que en los departamentos 28, 42, 20, 14, 15 y 6, la proporción de observaciones en el clúster 1 aumenta hasta un 65,22%. Es decir, en este caso puntual si que se detecta la existencia de cierta asociación entre las puntuaciones medias altas y las variables descriptivas consideradas.

Tabla 7.6. Titulaciones y departamentos agrupadas en los nodos tras la segunda ramificación

Nodo	Titulaciones	Departamento	Clúster	% de observaciones	Número de Observaciones
6	EPS ALCOY IT TELEC	11,22,26,13,21,23,31,9,41	2	53,19	50
	EU INFORMÁTICA		3	7,45	7
	EUIT INDUSTRIAL		1	39,36	37
7	EPS ALCOY IT TELEC	12,17,32	2	38	38
	EU INFORMÁTICA		3	15	15
	EUIT INDUSTRIAL		1	47	47
8	EPS ALCOY IT TELEC	27,39,19,25,16,35,10	2	22,45	11
	EU INFORMÁTICA		3	44,90	22
	EUIT INDUSTRIAL		1	32,65	16
9	EPS ALCOY IT TELEC	28,42,20,14,15,6	2	15,22	7
	EU INFORMÁTICA		3	19,57	9
	EUIT INDUSTRIAL		1	65,22	30

Esto pondría de manifiesto que un mismo departamento puede presentar mayor proporción de observaciones en un clúster u otro, en función de la titulación que estemos considerando, y de si ésta es técnica, superior o de segundo ciclo.

Seleccionando como predictor la variable departamento, se obtuvo de nuevo un árbol con dos niveles de ramificación. El primero de ellos, en función de la variable departamento, daba lugar a cuatro nodos. En un segundo nivel de ramificación, el nodo 10

y 11 se ramificaban en función de la variable titulación, dando lugar a un total de 8 nodos terminales.

En la tabla 7.7 se muestran los departamentos agrupados en cada nodo de la primera ramificación y el porcentaje de observaciones clasificadas en cada cluster.

En ella se puede apreciar que los departamentos 21, 35 y 6, clasificados en el nodo 13, presentan un elevado porcentaje de observaciones clasificadas en el clúster 3, mientras que los departamentos 24, 39, 19, 26, 3, 15, 38 y 5, pertenecientes al nodo 12, presentan una proporción relativamente elevada de observaciones clasificadas en el clúster 2.

Tabla 7.7. Características de los nodos de la primera ramificación del árbol de clasificación

	Departamentos	Clúster	% de observaciones	Número de Observaciones
Nodo 10	11, 22, 21, 34, 40, 9, 41	1	10,48	13
		2	63,23	66
		3	36,29	45
Nodo 11	12, 26, 27, 28, 42, 13, 17, 20, 32, 23, 14, 16, 30, 31, 18, 10, 33	1	18,40	85
		2	33,33	154
		3	46,27	223
Nodo 12	24, 39, 19, 25, 3, 15, 38, 5	1	41,14	65
		2	44,30	23
		3	14,66	70
Nodo 13	21, 35, 6	1	19,23	5
		2	7,69	2
		3	73,08	19

En la tabla 7.8 se muestran las características de los nodos filiales obtenidos de la segunda ramificación en los nodos 10 y 11. En este caso se detecta que en el nodo 15 y 18 se incrementa el porcentaje de observaciones agrupadas en el clúster 1, mientras que en el nodo 16 y 17 disminuye este porcentaje. También detectamos que en el nodo 15 y 19, con más de un 50% de observaciones en el clúster 3, muchas de las titulaciones agrupadas en ellos coinciden, a pesar de que los departamentos del nodo parental correspondiente son distintos.

Tabla 7.8. Titulaciones y departamentos agrupadas en los nodos tras la segunda ramificación

Nodo	Departamento	Titulación	Clúster	% de observaciones	Número de Observaciones
14	11, 22, 21, 34, 40, 9, 41	EPS ALCOY DISEÑO	2	81,08	30
		EPS GANDÍA SONIDO	3	10,81	4
		ETSIIICCP OOPP PLAN 97	1	8,11	3
15	11, 22, 21, 34, 40, 9, 41	EPS ALCOY IT IN. EC	2	17,24	5
		EPS GANDÍA TURISMO	3	59,62	17
		ETSIGCT TOPOGRAFÍA	1	24,14	7
		EUITA EXPL AGR			
		EUTA IND AGR			
EUITA MECAN Y C.R					
16	11, 22, 21, 34, 40, 9, 41	EPS ALCOY IT ELECT	2	63,45	31
		EPS GANDÍA FOREST	3	41,38	24
		EU INFORMÁTICA	1	5,17	3
		EUIT INDUSTRIA			
		EUITA HORTO Y JARD.			
17	12, 26, 27, 28, 42, 13, 17, 20, 32, 23, 14, 16, 30, 31, 18, 10, 33	EPS ALCOY IT DISEÑ	2	60	27
		EPS ALCOY IT TELEC	3	37,78	17
		EU ARQUITEC. TÉCNICA	1	2,22	1
18	12, 26, 27, 28, 42, 13, 17, 20, 32, 23, 14, 16, 30, 31, 18, 10, 33	EPS ALCOY IT E. IN.	2	18,42	7
		EPS ALCOY IT ELECT	3	36,84	14
		EPS ALCOY IT MEC.	1	44,74	17
		EPS GACNDÍA FOREST			
19	12, 26, 27, 28, 42, 13, 17, 20, 32, 23, 14, 16, 30, 31, 18, 10, 33	EPS ALCOY IT. QUIM.	2	31,66	120
		EPS ALCOY IT TEXT	3	50,66	182
		EPS GANDÍA ING ELECT.	1	17,68	67
		EPS GANDÍA SONIDO			
		EPS GANDÍA TURISMO			
		ETSI CCP OOPP			
		ETSCCP-OOPP PLAN 97			
		EYSGCT TOPOGRAFÍA			
		EU INFORMÁTICA			
		EUIT INDUSTRIALES			
		EUITA EXPL AGR			
		EUITA HORTO Y JARD.			
		EUITA IND. AGR.			

Finalmente, seleccionando como variable predictora la variable curso obtuvimos dos nodos, uno para los cursos primero, segundo y asignaturas optativas (codificadas como 0) y otro para tercero. Cada uno de estos nodos sufría un segundo nivel de ramificación, pero en el nodo correspondiente al tercer curso la nueva variable predictora era departamento, mientras que en el otro nodo era titulación. Ésto indicaría que en los cursos de primero, segundo y asignaturas optativas se detectan diferencias respecto a la variable titulación, mientras que en tercero las diferencias encontradas son respecto a la variable departamento.

En la tabla siguiente se muestra la composición de los nodos tras la primera ramificación, ya que al tener distintas variables predictoras en la segunda ramificación, en función del nodo considerado, no podremos extraer ninguna posible relación del análisis de ésta, excepto que en los cursos de primero, segundo y asignaturas optativas, si existen diferencias estadísticamente significativas en función del departamento que se considere e independientemente de la titulación, mientras que en tercero ocurre lo contrario, existiendo diferencias significativas en función de la titulación pero no del departamento.

Tabla 7.9. Características de los nodos tras la primera ramificación

Nodo	Cursos	Clúster	% de observaciones	Número de Observaciones
19	0,1,2	2	32,97	91
		3	42,75	118
		1	24,28	67
20	3	2	31,17	154
		3	48,38	239
		1	20,45	101

En este caso, el porcentaje de observaciones clasificadas en el clúster 3 fue de 42,75% en el nodo del tercer curso y de 48,38% en el nodo de los cursos primero, segundo y asignaturas optativas, porcentaje ligeramente superior, mientras que el porcentaje de observaciones clasificadas en el clúster 1 fue de un 24,28% en el nodo de primero, segundo y asignaturas optativas y de un 20,45% en el nodo de tercero.

Recordar que en las titulaciones superiores obteníamos que el último curso se agrupaba junto con las asignaturas optativas, primero, segundo y tercero, y presentaba un porcentaje de observaciones clasificadas en el clúster 1 de 17,13% y de un 42,07% en el clúster 3, mientras que en el nodo de los cursos cuarto y quinto se obtuvo un porcentaje de observaciones en el clúster 1 de 22,89% y de 37,89% en el clúster 3.

7.5.1.3. Titulaciones de Segundo Ciclo

Finalmente, se efectuó la segmentación de los datos de las titulaciones de segundo ciclo. Se obtuvo un árbol con dos niveles de ramificación. El primero de ellos en función de la variable departamento daba lugar a cuatro nodos y, posteriormente el tercer nodo se ramificaba en función de la variable predictora denominada doctor y que hace referencia a si los profesores considerados eran doctores o no. Este hecho pondría de manifiesto que mientras los alumnos de las titulaciones técnicas y superiores, no puntúan de manera diferente si el profesor es doctor o no, los alumnos de segundo ciclo si opinan de manera diferente en función de esta característica.

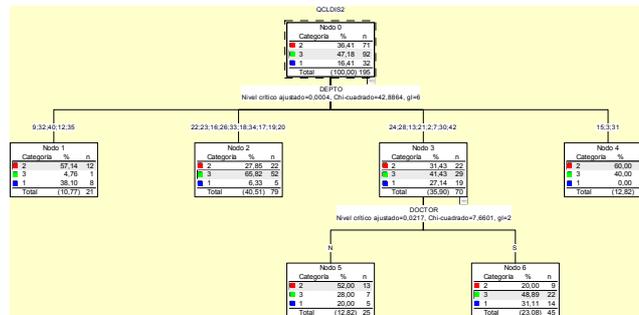


Figura 7.12. Árbol de clasificación

En la tabla 7.10 se muestran los departamentos clasificados en cada nodo de la primera ramificación y el número de observaciones que pertenecen a cada clúster en cada uno de ellos.

Tabla 7.10. Características de los nodos de la primera ramificación del árbol de clasificación

	Departamentos	Clúster	% de observaciones	Número de Observaciones
Nodo 1	9, 32, 40, 12, 35	1	38,10	8
		2	57,14	12
		3	4,76	1
Nodo 2	22, 23, 16, 26, 33, 18, 34, 17, 19, 20	1	6,33	5
		2	27,85	22
		3	65,82	52
Nodo 3	24, 28, 13, 21, 2, 7, 30, 42	1	27,14	19
		2	31,43	22
		3	41,43	29
Nodo 4	15, 3, 31	1	0	0
		2	60	15
		3	40	10

En ella se puede observar que mientras los departamentos 9, 32, 40, 12 y 35, pertenecientes al nodo 1, sólo presentan un 4,76% de observaciones clasificadas en el clúster 3, los departamentos 22, 23, 16, 26, 33, 18, 34, 17, 19 y 20, clasificados en el nodo 2, presentan un 65,82%. Destacar en este caso, que en el nodo 4 no hay ninguna observación correspondiente al clúster 1. Es decir, en este caso se podría considerar que el nodo 1 esta constituido por los profesores mejor valorados, mientras que el nodo 2 y, especialmente, el 4 lo estaría por los peor valorados.

La diferencia que se obtiene entre los departamentos parece ser más clara que en los casos de las titulaciones superiores y técnicas.

Dentro del nodo 3, correspondiente a los departamentos 24, 28, 13, 21, 2, 7, 30 y 42, se produce una segunda ramificación en función de si los profesores son doctores o no, detectándose que el porcentaje de observaciones clasificadas en el clúster 3 y en el clúster 1, es ligeramente superior, en caso de que lo sean.

Desarrollando el árbol mediante la variable predictora titulación, se obtuvo un único nivel de ramificación con tres nodos terminales, como se muestra en la figura 7.13.

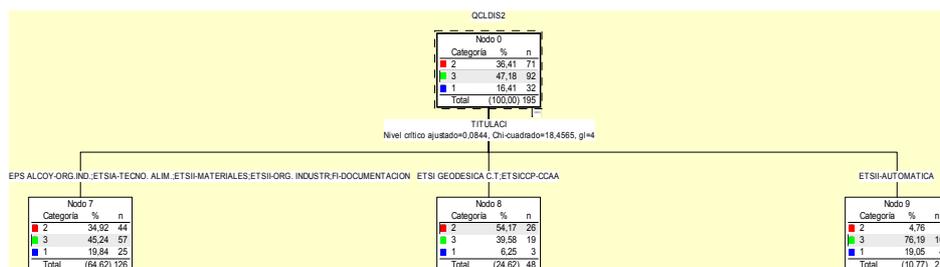


Figura 7.13. Árbol desarrollado con la variable predictora titulación

En la tabla siguiente se muestran las titulaciones clasificadas en cada nodo y la proporción de observaciones en cada uno de los clusters dentro de cada uno de ellos.

Tabla 7.11. Titulaciones pertenecientes a cada nodo

	Titulaciones	Clúster	% de observaciones	Número de Observaciones
Nodo 7	EPS ALCOY ORG. IND	1	19,84	25
	ETSIA TEC. ALIM	2	34,92	44
	ETSII MATERIALES	3	45,24	57
	ETSII ORG. IND.			
	FI DOCUMENTACIÓN			
Nodo 8	ETSI GEODÉSICA C. T.	1	6,25	3
	ETSICCP-CCAA	2	54,17	26
		3	39,58	19
Nodo 9	ETSII-AUTOMÁTICA	1	19,05	4
		2	4,76	1
		3	76,19	16

Se detecta que la titulación de ETSII automática presenta una proporción de observaciones clasificadas en el clúster 3, bastante superior al resto de nodos.

Identificación y caracterización de categorías docentes

Finalmente, seleccionando como variable predictora la variable curso, obtuvimos dos niveles de ramificación, como se muestra en la figura 7.14. El primero de ellos proporcionaba dos clusters, uno para los cursos primero, segundo, quinto y asignaturas optativas y el otro para el cuarto curso. Remarcar aquí, que al trabajar con bases de datos incompletas, en este caso no existía ninguna observación de tercer curso, lo que hace que los resultados deban interpretarse con cuidado. El primer nodo de ellos presentaba una segunda ramificación en función de la variable departamento.

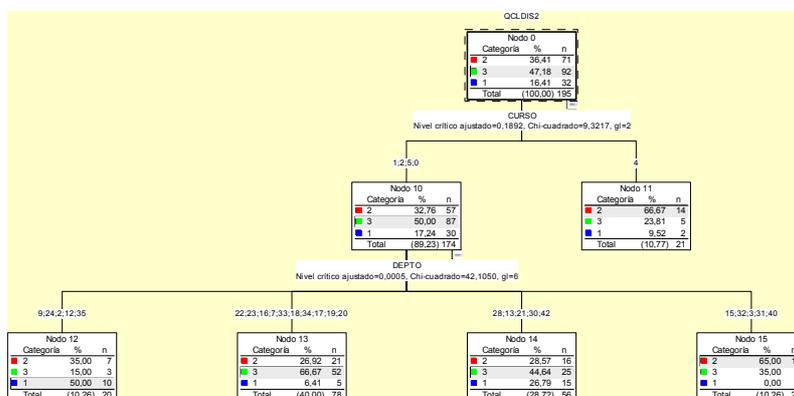


Figura 7.14. Árbol desarrollado con la variable predictora curso

De nuevo en este caso, las asignaturas optativas no presentan diferencias estadísticamente significativas con los primeros cursos y el último, como sucedía en las titulaciones superiores, pero si lo hacen con los cursos intermedios. Se obtuvo que el nodo en el que se clasificaban los cursos de 1º, 2º, 5º y optativas presentaban un porcentaje de observaciones clasificadas en el clúster 3 del 50%, mientras que en el otro nodo era sólo del 23,81%.

La composición de los nodos finales se muestra en la tabla 7.12. Puede observarse que la ramificación del nodo 10 por departamento, da lugar a nodos con muy diferente proporción de observaciones clasificadas como clúster 1, pasando de un 50 % en el nodo 9 a un 0 % en el nodo 12.

Tabla 7.12. Características de los nodos finales

Nodo	Curso	Departamento	Clúster	% de observaciones	Número de Observaciones
11	4		2	66,67	14
			3	23,81	5
			1	9,52	2
12	1,2,5,0	9,24,2,12,35	2	35	7
			3	15	3
			1	50	10
13	1,2,5,0	22,23,16,7,33,18,34,17,19,20	2	26,92	21
			3	6,67	52
			1	6,41	5
14	1,2,5,0	28,13,21,30,42	2	28,57	16
			3	44,65	25
			1	26,79	15
15	1,2,5,0	15,32,3,31,40	2	65	13
			3	35	7
			1	0	0

A la vista de todos los resultados obtenidos, podemos concluir que el método CHAID Exhaustivo representa una poderosa herramienta para segmentar los datos procedentes de las encuestas de opinión de los alumnos y relacionar esos resultados con características descriptivas relativas al profesor o asignatura.

En nuestro caso, independientemente de que la titulación sea superior, técnica o de segundo ciclo, se han detectado diferencias estadísticamente significativas en función de las variables predictoras curso, titulación, departamento y si el profesor es doctor o no.

En consecuencia, para el siguiente apartado se utilizarán como nodos terminales los obtenidos en la primera ramificación de cada árbol, aunque a nivel metodológico, para futuros análisis con bases de datos completas, lo conveniente sería trabajar con los nodos terminales obtenidos tras el completo desarrollo del árbol.

7.5.2. Identificación de los ítems de mayor poder discriminante en las tipologías docentes obtenidas

En este apartado se añade un paso más a la metodología desarrollada en este capítulo que consistirá en utilizar el análisis discriminante con la finalidad básica de interpretar las diferencias existentes entre los grupos obtenidos dentro de cada una de las posibles titulaciones (técnicas, superiores o de segundo ciclo) mediante el método de clasificación CHAID Exhaustivo. De esta manera se podrán identificar las variables o ítems incluidos en el cuestionario que mejor discriminan entre esas categorías dentro de cada uno de los nodos obtenidos en el primer nivel de ramificación de cada uno de los casos estudiados anteriormente y determinar las que mejor los caracterizan. Es decir, para cada nodo en particular nos permitirá detectar los ítems que más influyen en la valoración global de buen profesor obtenida mediante el ítem criterio.

Como variable dependiente se ha creado una nueva variable denominada “Escala R19”, en la cual se categorizan los valores medios obtenidos en el ítem criterio R19, como se muestra en la tabla siguiente. Esta escala ha sido elegida de manera que el número de observaciones en cada intervalo fuese lo más homogéneo posible, ya que los estudios previos mostraban que las puntuaciones intermedias son mucho más frecuentes que las extremas. En concreto, en ninguna ocasión se obtienen puntuaciones inferiores a 1 y, en muy pocos casos, se han obtenido puntuaciones del extremo inferior (por debajo de cinco).

Tabla 7.13. Valores de la variable Escala R19

Valor Escala R19	Valor Media R19
1	[0-5[
2	[5-6[
3	[6-7[
4	[7-8[
5	≥8

Como variables independientes se utilizaron las puntuaciones medias obtenidas de los ítems del cuestionario que resultaron ser los que mejor discriminaban entre los perfiles

de los grupos obtenidos mediante la aplicación del análisis clúster en el capítulo 6, tras efectuar un LDA sin incluir el ítem criterio, R19, ya que con ellos quedaban representadas todas las dimensiones subyacentes del cuestionario y se conseguía una adecuada diferenciación entre los tres grupos con una única función discriminante.

Los resultados completos de todos los análisis discriminantes pueden consultarse en el anexo 6, comentándose únicamente los más relevantes en este apartado.

7.5.2.1. Titulaciones Superiores

Se efectuó un análisis discriminante en cada uno de los grupos obtenidos mediante el método de clasificación CHAID Exhaustivo, tanto al utilizar como variable predictora el departamento, como al utilizar la titulación o el curso.

7.5.2.1.1 Variable predictora el departamento

En este caso se obtuvieron 6 nodos en el primer nivel de ramificación. El primer grupo estaba constituido por los departamentos codificados como 11, 23, 39, 20, 32 y 2, con un total de 365 observaciones. En este caso se obtuvieron tres funciones discriminantes, aunque sólo con la primera se lograba explicar el 99,5 % de la varianza.

Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R12, R1 y R5, todos ellos correlacionados positivamente con la primera función discriminante y con un valor muy similar de correlación.

La matriz de estructura se muestra en la tabla 7.14. En ella podemos detectar que los ítems R5, R12 y R1, de mayor poder discriminante, están más correlacionados con la primera función discriminante.

El porcentaje de observaciones correctamente clasificadas es de un 79,7 %.

Tabla 7.14. Matriz de estructura

	Función		
	1	2	3
MEDIA R1	,768(*)	,084	-,635
MEDIA R12	,730(*)	-,507	,459
MEDIA R5	,714(*)	,609	,346
MEDIA R4(a)	,610(*)	-,080	,049
MEDIA R2(a)	,516(*)	,093	-,410
MEDIA R3(a)	,495(*)	-,129	,011
MEDIA R16(a)	,495(*)	-,005	,008
MEDIA R8(a)	,413(*)	-,024	,303
MEDIA R18(a)	,359(*)	-,055	,054
MEDIA R7(a)	,337(*)	-,032	,308
MEDIA R11(a)	,427	-,070	,520(*)
MEDIA R10(a)	,363	-,172	,504(*)

(a) Variable no está incluida en el análisis

(*) Variable mayormente correlacionada con esa función discriminante

El segundo grupo estaba constituido por los departamentos codificados como 12, 19, 26, 3, 30 y 33, con un total de 290 observaciones.

En este caso obtuvimos cuatro funciones discriminantes, aunque sólo con la primera se lograba explicar el 98,3 % de la varianza.

Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R12, R1, R5 y R2, todas ellas correlacionadas con la primera función discriminante, aunque el ítem R2, presenta un mayor grado de correlación con la función 3. La matriz de estructura se muestra en la tabla 7.15.

En este caso se obtuvo un porcentaje de observaciones correctamente clasificadas del 79,3%, valor muy similar al del caso anterior.

Tabla 7.15. Matriz de estructura

	Función			
	1	2	3	4
MEDIA R1	,766(*)	-,576	-,023	,286
MEDIA R5	,727(*)	-,104	,219	-,642
MEDIA R12	,696(*)	,418	-,576	-,096
MEDIA R4	,572(*)	-,014	,050	-,090
MEDIA R16	,464(*)	,047	-,003	,055
MEDIA R3	,446(*)	-,135	-,430	,063
MEDIA R11	,434(*)	,269	-,181	-,369
MEDIA R18	,415(*)	,032	,018	,065
MEDIA R10	,368(*)	,342	-,282	-,242
MEDIA R8	,354(*)	,231	-,110	-,128
MEDIA R7	,313(*)	,301	-,048	-,200
MEDIA R2	,643	,000	,655(*)	,396

El tercer grupo estaba constituido por los departamentos codificados como 13, 42, 7, 21, 34, 14, 38, 22, 25 y 27, con un total de 275 observaciones. En este caso obtuvimos dos funciones discriminantes, aunque sólo con la primera se lograba explicar el 95,8% de la varianza. Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R1 y R11. La matriz de estructura en este caso quedaría como se muestra en la tabla 7.16. El porcentaje de observaciones correctamente clasificadas fue de un 66,7%, valor ligeramente inferior al obtenido en los casos anteriores.

El cuarto grupo estaba constituido por los departamentos codificados como 17, 24, 1, 15 y 40, con un total de 83 observaciones.

Se obtuvieron dos funciones discriminantes, aunque sólo con la primera se lograba explicar el 99,8% de la varianza.

Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R12 y R1, siendo este último el que presentaba mayor grado de correlación con la primera función discriminante, como se muestra en la tabla 7.17.

El porcentaje de observaciones correctamente clasificadas es de un 78,3%.

Tabla 7.16. Matriz de estructura

	Función	
	1	2
MEDIA R5	,558(*)	-,072
MEDIA R12	,554(*)	,090
MEDIA R4	,514(*)	-,023
MEDIA R3	,389(*)	-,063
MEDIA R8	,009(*)	,000
MEDIA R11	,640	,769(*)
MEDIA R1	,680	-,734(*)
MEDIA R10	,417	,573(*)
MEDIA R2	,271	-,509(*)
MEDIA R7	,144	,300(*)
MEDIA R16	,157	,183(*)
MEDIA R18	,019	-,043(*)

Tabla 7.17. Matriz de estructura

	Función	
	1	2
MEDIA R1	,825(*)	-,565
MEDIA R4	,588(*)	,245
MEDIA R3	,479(*)	,175
MEDIA R5	,447(*)	,111
MEDIA R18	,340(*)	-,073
MEDIA R16	,326(*)	,129
MEDIA R7	,278(*)	,252
MEDIA R12	,636	,772(*)
MEDIA R10	,394	,610(*)
MEDIA R2	,473	-,495(*)
MEDIA R11	,307	,490(*)
MEDIA R8	,361	,376(*)

En el quinto grupo, constituido por los departamentos codificados como 28, 16, 18, 31 y 41, se obtuvieron tres funciones canónicas discriminantes, explicando la primera de ellas el 96,5%. En este caso los ítems incluidos fueron el R12, R1, R5, R10 y R2, todos ellos correlacionados más fuertemente con la primera función excepto el ítem R10 que presentaba mayor correlación con la cuarta. En la tabla 7.18 se muestra la correlación de cada una de ellas con las funciones discriminantes. El porcentaje de observaciones correctamente clasificadas es de un 80,4%.

Finalmente, en el sexto grupo constituido por los departamentos 9 y 37, se obtuvieron 2 funciones canónicas discriminantes, explicándose con la primera el 98,2% de la varianza. Las variables incluidas en el análisis fueron R12 y R2.

En la tabla 7.19 se muestra su correlación con las funciones discriminantes.

Tabla 7.18. Matriz de estructura del quinto grupo

	Función			
	1	2	3	4
MEDIA R1	,683(*)	-,133	,448	-,561
MEDIA R5	,657(*)	-,482	,204	,485
MEDIA R2	,629(*)	-,105	-,363	-,452
MEDIA R12	,605(*)	,286	-,179	,376
MEDIA R4	,493(*)	,115	-,011	,179
MEDIA R3	,478(*)	,176	,400	,149
MEDIA R8	,417(*)	,109	,073	,331
MEDIA R18	,313(*)	,171	,077	,155
MEDIA R16	,284(*)	-,017	,135	,163
MEDIA R10	,424	,608	,211	,636(*)
MEDIA R11	,361	,449	,122	,564(*)
MEDIA R7	,272	-,053	-,071	,353(*)

Tabla 7.19. Matriz de estructura del sexto grupo

	Funciones	
	1	2
MEDIA R12	,761(*)	-,649
MEDIA R3	,547(*)	-,080
MEDIA R1	,520(*)	,507
MEDIA R4	,412(*)	,079
MEDIA R8	,322(*)	-,275
MEDIA R7	,226(*)	-,058
MEDIA R16	,173(*)	,070
MEDIA R2	,434	,901(*)
MEDIA R10	,233	-,785(*)
MEDIA R11	,066	-,734(*)
MEDIA R5	,013	-,330(*)
MEDIA R18	,157	-,313(*)

El porcentaje de observaciones correctamente clasificadas es de un 78,3%, valor similar al obtenido en los otros casos.

Finalmente, en la tabla 7.20 se muestra una relación de los ítems que más han pesado en la valoración global para cada grupo, es decir, aquellas características que más han influido a la hora de dar un juicio global.

Tabla 7.20. Ítems discriminantes

GRUPO	ITEMS
1	R12, R1, R5
2	R12, R1, R2 y R5
3	R1 y R11
4	R12 y R1
5	R12, R1, R2, R10 y R5
6	R12, R2

En ella, podemos detectar los ítems que mejor sirven para comprender a cada uno de los grupos, por ejemplo, mientras que en el grupo 1 los ítems que más han influido en la segmentación obtenida han sido los codificados como R1, R12 y R5, en el grupo 3 lo han hecho los codificados como R1 y R11. Asimismo, puede detectarse que los ítems R1 (El profesor explica en clase con orden y claridad) y R12 (Consigue que los alumnos estén motivados por la asignatura) aparecen en casi todos los grupos.

7.5.2.1.2 Variable predictora la titulación

De igual forma que en el apartado anterior, se efectuó un análisis discriminante en cada uno de los nodos obtenidos mediante el árbol de clasificación utilizando como variable predictora la titulación.

En este caso se habían obtenido cuatro grupos, el primero de ellos constituido por las titulaciones EPS ALCOY, EPS ALCOY-ADE; ETSICCP-CCP y ETSICCP-PIE (IC), con un total de 126 observaciones. Por no extendernos demasiado únicamente se comentan los principales resultados obtenidos, pudiendo consultar la totalidad de los mismos en el anexo 6.

De nuevo obtuvimos tres funciones discriminantes, explicando con la primera de ellas el 96,9% de la varianza. Las variables incluidas en el análisis fueron las medias de los ítems codificados como R1, R12 y R11. El porcentaje de observaciones clasificadas correctamente fue de 73 %.

El segundo grupo estaba constituido por las titulaciones de EPS GANDÍA CCAA, EPS GANDÍA-TELECOMUNICACIONES, ETSII ADMON y DE, con un total de 53 observaciones. Las variables incluidas en el análisis fueron las medias de los ítems codificados como R11 y R1, siendo esta última la que presenta mayor correlación. Se obtuvieron dos funciones discriminantes, pero sólo con la primera se explicaba un 95,9% de la varianza. El porcentaje de observaciones correctamente clasificadas fue de un 83%.

En el tercer grupo, constituido por las titulaciones de ETSII TELECOMUNICACIONES, ETSIA, ETSICCP-CCP PLAN 97 y F. INFORMÁTICA, con un total de 663 observaciones, se obtuvieron 4 funciones discriminantes, explicándose con la primera de ellas el 99,4% de la varianza. Las variables incluidas en el análisis fueron las medias de los ítems R1, R5, R12, R4, R2 y R16. El porcentaje de observaciones correctamente clasificadas fue del 80,7%.

Finalmente, en el cuarto grupo constituido por las titulaciones de ETSII PLAN 93 y ETSII QUÍMICO, con un total de 332 observaciones, se obtuvieron cuatro funciones discriminantes y las variables incluidas en el análisis fueron R1, R5, R12 y R2. El porcentaje de varianza explicada con la primera función fue de un 98,5%. El porcentaje de observaciones correctamente clasificadas fue de un 78,9%.

Es decir, en el caso de los grupos obtenidos mediante la variable predictora titulación, los ítems considerados en cada uno de ellos fueron los que se relacionan en la tabla 7.21, observándose que en todos los casos vuelve a estar presente el ítem R12 y, en casi todos, el ítem R1.

Tabla 7.21. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R12, R1 y R11
2	R1 y R11
3	R1,R12,R2,R5,R6 y R4
4	R5,R1,R12 y R2

7.5.2.1.3 Variable predictora el curso

En este caso se efectuó un análisis discriminante en cada uno de los grupos obtenidos mediante el árbol de clasificación utilizando la variable predictora el curso. Se obtuvieron dos grupos, el primero de ellos constituido por primero, segundo, tercero, sexto y las asignaturas optativas, con un total de 794 observaciones.

Las variables incluidas en el análisis fueron las medias de los ítems codificados como R1, R5, R12, R2 y R11. Se obtuvieron cuatro funciones discriminantes, explicando con la primera el 99,5% de la varianza. El porcentaje de observaciones clasificadas correctamente fue de 76,8%.

El segundo grupo estaba constituido por los cursos cuarto y quinto, con un total de 360 observaciones. Las variables incluidas en el análisis fueron las medias de los ítems codificados como R1, R12, R2, R7 y R11. Se obtuvieron 4 funciones discriminantes, pero sólo con la primera se explicaba un 99,0% de la varianza. El porcentaje de observaciones correctamente clasificadas fue de un 82,9%. En este caso, de nuevo los ítems R12 y R1 estaban presentes en ambos grupos, aunque el resto de ítems incluidos es ligeramente diferente a los que se obtenían mediante la variable predictora departamento o titulación. El ítem R1 es el de máxima correlación con la función discriminante en los dos casos.

Tabla 7.22. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R1, R5, R12, R2 y R11
2	R1, R12, R2, R7 y R11

7.5.2.2. Titulaciones Técnicas

Se efectuó un análisis discriminante en cada uno de los grupos obtenidos mediante el método de clasificación CHAID Exhaustivo, tanto al utilizar como variable predictora la titulación, como los obtenidos al utilizar como variable predictora el departamento o el curso.

7.5.2.2.1 Variable predictora titulación

Como en el apartado anterior se efectuó un análisis discriminante en cada uno de los grupos obtenidos mediante el árbol de clasificación utilizando como variable predictora la titulación. En el primero de ellos, se agruparon 93 observaciones, correspondientes a las titulaciones de EPS ALCOY-IT DISEÑO, ETSICCP OOPP PLAN 97 y EU ARQUITECTURA TÉCNICA. El segundo grupo estuvo constituido por 289 observaciones que pertenecían a las titulaciones de EPS ALCOY- IT TELECO, EU INFORMÁTICA y EUIT INDUSTRIAL. El tercer grupo estuvo constituido por 170 observaciones pertenecientes a EPS ALCOY-ITI E. IN, EPS ALCOY-IT QUÍMICOS, EPS ALCOY-ITI TEXTIL, EPS GANDÍA FOREST y ETSIGCT-TOPOGRAFÍA. El cuarto grupo, estuvo formado por 190 observaciones correspondientes a las titulaciones EPS GANDÍA SONIDO, EPS GANDÍA ELECTRICOS, EUITA EXPLOTACIONES AGROPECUARIAS, EUITA HORTO Y JARDINERÍA y EUITA INDUSTRIAS AGROALIMENTARIAS. Finalmente, el quinto grupo estuvo formado por 28 observaciones correspondientes a las titulaciones EPS GANDÍA TURISMO, ETSICCP-CCP PLAN 97 y EUITA MECANIZACIÓN y CAMINOS RURALES.

En el primer grupo, al efectuar el análisis discriminante se obtuvieron tres funciones canónicas discriminantes, explicándose con la primera de ellas el 97,4% de la varianza. Las variables incluidas en el análisis fueron las medias de los ítems codificados como R5, R2 y R12, siendo en este caso R5 la de mayor poder discriminante. El porcentaje de observaciones clasificadas correctamente fue de 79,6%.

En el segundo grupo, se obtuvieron también tres funciones discriminantes, explicando la primera de ellas el 98,3% de la varianza, siendo las variables incluidas en el análisis R12, R1 y R11. Se conseguía un porcentaje de observaciones correctamente clasificadas fue de un 74 %.

En el tercer grupo se obtuvieron dos funciones discriminantes, explicando la primera de ellas el 99,1% de la varianza. Las variables incluidas en el análisis fueron las medias de los ítems R1 y R12. El porcentaje de observaciones correctamente clasificadas fue del 70,6%.

En el cuarto grupo obtuvimos cuatro funciones discriminantes, explicando la primera de ellas el 99,0% de la varianza. Las variables incluidas en el análisis fueron R1, R2, R18, R11 y R12, clasificándose correctamente el 76,3% de las observaciones.

Finalmente, en el quinto grupo se obtuvieron dos funciones discriminantes, siendo las variables incluidas en el análisis R1 y R12. Con la primera función discriminante se explicaba un 99,4% de la varianza. El porcentaje de observaciones correctamente clasificadas fue de un 82,1%.

La relación de los ítems de mayor poder discriminante se muestra en la tabla 7.23. De nuevo podemos observar que los ítems R1 y R12 aparecen en todos los grupos.

Tabla 7.23. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R1, R12 y R11
2	R12, R1 y R11
3	R1 y R12
4	R1, R2, R18, R11 y R12
5	R12 y R1

7.5.2.2.2 Variable predictora departamento

Se efectuó un análisis discriminante en cada uno de los grupos obtenidos tras desarrollar el árbol de clasificación utilizando como variable predictora el departamento.

El primer grupo estaba constituido por los departamentos codificados como 11, 22, 21, 34, 40, 9 y 41, con un total de 124 observaciones. En este caso obtuvimos cuatro funciones discriminantes, aunque sólo con la primera se lograba explicar el 90,8% de la varianza. Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R12, R1, R11, R8 y R3. El porcentaje de observaciones correctamente clasificadas es de un 82,3%.

El segundo grupo estaba constituido por los departamentos codificados como 12, 26, 27, 28, 42, 13, 17, 20, 32, 23, 7, 14, 16, 30, 31, 18, 10 y 33, con un total de 462 observaciones. En este caso obtuvimos cuatro funciones discriminantes, explicándose con la primera el 99,2 % de la varianza. Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R1, R12, R2 y R11. Se obtuvo un porcentaje de observaciones correctamente clasificadas del 73,4%.

En el tercer grupo, constituido por los departamentos codificados como 24, 39, 19, 25, 3, 15, 38 y 5, con un total de 158 observaciones, obtuvimos dos funciones discriminantes, explicándose con la primera el 99,6% de la varianza. Las variables incluidas en el estudio fueron los valores medios de los ítems codificados como R12 y R2. El porcentaje de observaciones correctamente clasificadas es de un 67,7 %.

Finalmente, en el cuarto grupo, constituido por los departamentos 2, 1, 35 y 6, con 26 observaciones, las variables incluidas en el análisis fueron de nuevo R1 y R12, obteniéndose dos funciones discriminantes. Con la primera función se explicaba un 86,7% de la varianza. El porcentaje de observaciones correctamente clasificado fue de un 87%.

La relación de ítems de mayor poder discriminante en cada grupo se muestra a continuación. En este caso, en el grupo 3 no aparece el ítem R1.

Tabla 7.24. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R12, R1, R11, R8 y R3
2	R1, R12, R2 y R11
3	R12 y R2
4	R1 y R12

7.5.2.2.3 Variable predictora curso

Se efectuó un análisis discriminante en los dos grupos obtenidos mediante el árbol de clasificación utilizando como variable predictora el curso.

En el primer grupo, constituido por las observaciones correspondientes al curso de tercero, se obtuvieron cuatro funciones discriminantes, explicándose con la primera el 99,1% de la varianza. Las variables incluidas en el análisis fueron R1, R12, R5 y R18, clasificándose correctamente el 71,7% de los casos originales.

En el segundo grupo, constituido por las observaciones correspondientes a los cursos de primero, segundo y asignaturas optativas, se obtuvieron 4 funciones discriminantes, explicándose con la primera el 99,3% de la varianza. Las variables incluidas en el análisis fueron R1, R12, R2 y R11, clasificándose correctamente el 77,7 % de los casos originales.

La relación de ítems de mayor poder discriminante en cada caso se muestra a continuación.

Tabla 7.25. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R1, R12, R5 y R18
2	R1, R12, R2 y R11

7.5.2.3. Titulaciones de Segundo Ciclo

Finalmente se efectuó un análisis discriminante en cada uno de los grupos obtenidos en las observaciones pertenecientes a las carreras de segundo ciclo.

7.5.2.3.1 Variable predictora el departamento

En este caso obtuvimos cuatro posibles grupos. El primero de ellos estaba constituido por 21 observaciones pertenecientes a los departamentos 9, 32, 40, 12 y 35, el segundo estaba formado por 79 observaciones pertenecientes a los departamentos

codificados como 22, 23, 16, 26, 33, 18, 34, 17, 19 y 20, el tercero estaba formado por 70 datos que pertenecían a los departamentos 24, 28, 13, 21, 2, 7, 30 y 42 y, el último grupo estaba constituido por 25 observaciones de los departamentos 15, 3 y 31.

En el primer grupo, obtuvimos dos funciones canónicas discriminantes, correlacionadas con las variables R1 y R4. Sólo con la primera función discriminante se explica el 93,6% de la varianza. El porcentaje de casos originales clasificados correctamente fue del 81%.

En el segundo grupo se obtuvieron tres funciones discriminantes, pero en este caso las variables incluidas en el análisis fueron R2, R5 y R8, destacando en este caso, que no se incluyen en el análisis ni el ítem R1 ni el R12. La primera función canónica explicaba el 95,0 % de la varianza y se clasificaban correctamente el 63,3% de los casos originales.

En el tercer grupo obtuvimos dos funciones discriminantes, incluyéndose en el análisis las variables R1 y R10. El porcentaje de varianza explicada por la primera función discriminante fue de un 99,8%. El porcentaje de observaciones originales clasificadas correctamente fue de un 75,7%.

Finalmente, en el grupo cuarto obtuvimos una única función discriminante, incluyéndose sólo la variable R1 en el análisis. El porcentaje de casos originales correctamente agrupados fue del 80,0%.

La relación de ítems de mayor poder discriminante en cada caso se muestra en la tabla 7.26. Como puede observarse el ítem R12 no aparece como relevante para caracterizar ninguno de los grupos.

Tabla 7.26. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R1 y R4
2	R2, R5 y R8
3	R1 y R10
4	R1

7.5.2.3.2 Variable predictora la titulación

En este caso se efectuó el análisis discriminante en los tres grupos obtenidos al seleccionar como variable predictora la titulación.

El primer grupo estaba constituido por 126 observaciones correspondientes a las titulaciones de EPS ALCOY ORGANIZACIÓN INDUSTRIAL, ETSIA TECNOLOGÍA DE ALIMENTOS, ETSII MATERIALES, FI DOCUMENTACIÓN y ETSII ORGANIZACIÓN INDUSTRIAL. Se obtuvieron tres funciones discriminantes, la primera de las cuales explicaba un 98,1% de la varianza. Las variables incluidas en el análisis fueron R1, R5 y R10, clasificándose correctamente el 76,2% de los casos originales.

El segundo grupo estaba constituido por 48 observaciones, de las titulaciones ETSI GEODÉSICA y ETSICCP-CCAA. Se obtuvo una única función discriminante, correlacionada con la variable R1. El porcentaje de datos originales clasificados correctamente fue del 68,8%.

El tercer grupo estaba constituido por 21 observaciones, de las titulaciones ETSII AUTOMÁTICA. Se obtuvieron dos funciones discriminantes, correlacionada con las variables R10 y R5. La primera función discriminante explica el 98,3% de las observaciones. El porcentaje de datos originales clasificados correctamente fue del 76,2%.

La relación de ítems de mayor poder discriminante en cada caso se muestra en la tabla 7.27. De nuevo aparece un grupo en el que, ni el ítem R1 ni el ítem R12, resultan incluidos en el análisis discriminante y además, este último no aparece en ninguno de los grupos.

Tabla 7.27. Relación de los ítems discriminantes de cada grupo

GRUPO	ITEMS
1	R1, R5 y R10
2	R1
3	R5 y R10

7.5.2.3.3 Variable predictora el curso

Se aplicó el análisis discriminante a los tres grupos obtenidos mediante el método CHAID Exhaustivo, utilizando como variable predictora el curso.

En el primer grupo, constituido por 174 observaciones de primero, segundo y quinto curso, o bien, de alumnos que cursaban asignaturas optativas, se obtuvieron tres funciones discriminantes correlacionadas con las variables R1, R5 y R12. La primera función canónica explicaba el 99,2% de la varianza, clasificándose correctamente el 76,4% de las observaciones originales.

En el segundo grupo, con 21 observaciones pertenecientes a cuarto curso, se obtuvo una sola función discriminante correlacionada con R1, clasificándose correctamente el 95,2% de los datos originales respectivamente.

A la vista de los resultados obtenidos en todos los casos podemos concluir que casi todos los grupos, independientemente de la variable predictora utilizada para su obtención, vienen caracterizados por ítems relacionados con la primera dimensión subyacente del cuestionario extraída mediante el análisis factorial y denominada *Dominio, organización y claridad* y con la segunda, denominada *Interrelación profesor- alumno*.

De todos los ítems correlacionados con la primera dimensión (R1, R2, R3, R4, R5, R6 y R19), el más frecuente en la mayoría de grupos es el ítem R1 (*El profesor explica con orden y claridad*), mientras que los ítems R3 (*El ritmo de las clases es adecuado*), R4 (*Sintetiza y subraya los conceptos más importantes*) y R6 (*Los contenidos más importantes del programa han sido desarrollados durante el curso*) sólo aparecen en uno de los grupos cada uno.

De todos los ítems relacionados con la segunda dimensión (R9, R10, R11 y R12), el más frecuente en casi todos los grupos es el R12 (*Consigue que los alumnos estén motivados por la asignatura*), mientras que el ítem R10 (*Dialoga con los alumnos sobre la marcha de las clases*) únicamente aparece en 3 grupos.

Únicamente en tres de los grupos aparecen ítems relacionados con la tercera y cuarta dimensión (*Recursos e Interrelación con otras asignaturas*). Es decir, que básicamente con dos de las dimensiones del cuestionario podríamos caracterizar los distintos perfiles de tipologías docentes obtenidos, siendo los ítems R1, R2, R5 R11 y R12 los que más frecuentemente encontramos en la mayoría de los grupos.

7.5.3. Identificación de los ítems de mayor poder discriminante entre las tipologías docentes

Por último, se aplicó el método SIMCA utilizando como variables las puntuaciones medias obtenidas en los ítems del cuestionario, con la finalidad de determinar cuales de ellos discriminan mejor entre los distintos grupos obtenidos en cada uno de los nodos de la primera ramificación, utilizando como variables predictoras el departamento, la titulación y el curso en las carreras técnicas, superiores y de segundo ciclo.

Los modelos utilizados para la aplicación del método fueron los obtenidos al realizar un análisis de componentes principales para cada uno de los nodos mencionados anteriormente. A estos modelos se les ha denominado como “*nodo i de la variable predictora j*”, siendo *i* el número de nodo correspondiente en el árbol de clasificación y *j* la variable predictora utilizada en la ramificación de dicho árbol.

Por un lado, se evaluó el poder discriminante de cada una de las variables entre los distintos modelos estimados, que nos informa sobre cuales son los ítems que más contribuyen a la diferenciación entre dos de los nodos obtenidos mediante la técnica de árboles de clasificación. Este estudio se efectuó por un lado, para cada posible combinación de modelos estimados en los distintos departamentos, por otro lado, para cada posible combinación de los modelos estimados para las titulaciones y, finalmente para cada posible combinación entre los modelos de los cursos.

Por otro lado, se evaluó la distancia entre modelos para identificar aquellos que presentan un mayor o menor parecido.

En las figura 7.15 se muestra de nuevo el árbol obtenido para la variable titulación en las carreras superiores y en la figura 7.16 los resultados obtenidos para cada nodo. La totalidad de los resultados del poder de discriminación entre los distintos departamentos, titulaciones y cursos para las titulaciones técnicas y de segundo ciclo obtenidos para la totalidad de los ítems incluidos en el cuestionario, se muestran en el anexo 6.

En este caso, podemos detectar que entre los modelos del nodo 1 y 2 todas las variables presentaban valores de poder de discriminación superiores a tres, es decir, que la diferenciación entre estas clases no es debida a unas pocas variables sino al conjunto de todas ellas.

Por el contrario, entre los modelos 1 y 3, 1 y 4 y 3 y 4 no existe ninguna variable con valor del poder discriminante superior a tres. Entre los modelos 2 y 3 y 3 y 4, los ítems que resultaron con un valor de poder discriminante superior a 3 fueron el R1 y el R19.

En el caso de evaluar el poder discriminante obtenido al trabajar con los modelos de cada uno de los nodos obtenidos mediante la variable predictora departamento, obtuvimos que únicamente entre los modelos 1 y 4, 2 y 4, 3 y 4, 4 y 5, 4 y 6 y 3 y 6, existían variables con valores de poder discriminante superior a 3. En este caso, el ítem R19, suele ser el que aparece con valores de poder discriminante superior a tres entre el modelo 4 y el resto de modelos, mientras que entre las clases nodo 3 y nodo 6 era el ítem R8.

Entre los modelos de los nodos obtenidos con la variable curso no existe ninguna variable con valor del poder discriminante superior a tres.

Resultados similares se obtienen para las titulaciones técnicas y de segundo ciclo, variando los ítems de poder discriminante superior a tres en función de los modelos que se estén considerando. En estos casos no se mantiene que el ítem R19 sea el que más frecuentemente presente valores de poder discriminante entre los modelos superiores a tres.

Identificación y caracterización de categorías docentes

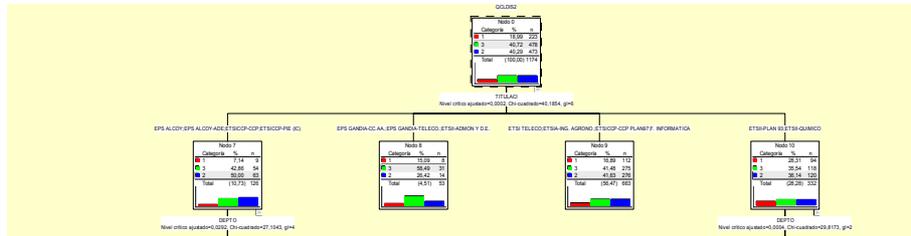


Figura 7.15. Árbol obtenido con variable predictora titulación

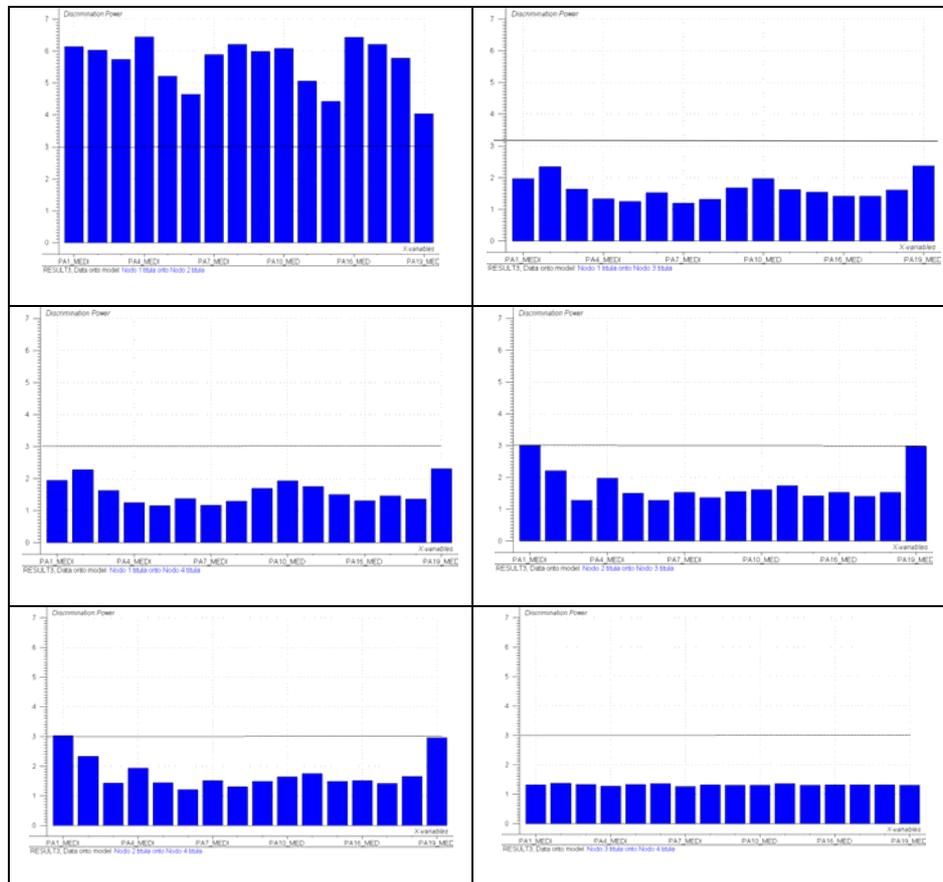


Figura 7.16. Gráficos de poder discriminante entre los modelos de las titulaciones

Esto nos indicaría que sí que existen ítems en el cuestionario que permiten diferenciar entre los distintos perfiles de tipologías docentes y, estos ítems son diferentes en función de las clases que se estén considerando.

Por último se estudian las distancias entre clases de los distintos modelos. En la figura 7.17 se muestra el valor de la distancia entre los modelos de los 6 nodos obtenidos en las carreras superiores mediante la variable predictora departamento. Los valores obtenidos para todos los modelos se muestran en el anexo 6.

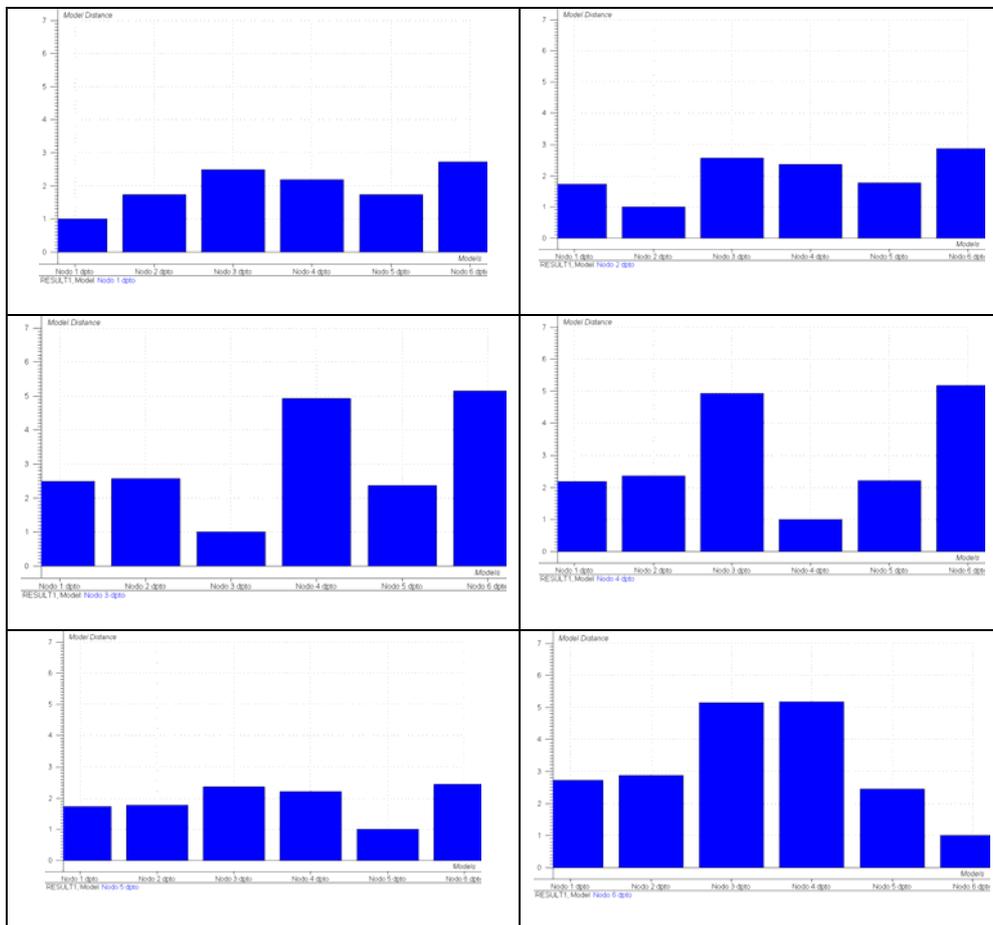


Figura 7.18. Gráficos de distancia entre los modelos obtenidos con la variable departamento

Numéricamente, los valores de distancia obtenidos entre los distintos modelos se muestran en la tabla siguiente.

Tabla 7.28. Valor de la distancia entre modelos

	Nodo 1	Nodo 2	Nodo 3	Nodo 4	Nodo 5	Nodo 6
Nodo 1	1,000	1,725	2,492	2,184	1,735	2,717
Nodo 2	1,725	1,000	2,566	2,371	1,773	2,868
Nodo 3	2,492	2,566	1,000	4,932	2,368	5,142
Nodo 4	2,184	2,371	4,932	1,000	2,220	5,171
Nodo 5	1,735	1,773	2,368	2,220	1,000	5,452
Nodo 6	2,171	2,868	5,142	5,171	2,452	1,000

En este caso concreto, no se detectan valores de distancias muy elevados, pero, por ejemplo entre los modelos nodo 1 y nodo 4 obtenidos con la variable predictora titulaciones, se obtiene un valor de distancia de 28.415, lo que nos indicaría que existe diferencia entre ambos.

Algo similar sucede en las titulaciones técnicas y de segundo ciclo, donde también se obtienen valores de distancia superiores a 5 entre algunos de los modelos, mientras que entre otros muchos los valores son relativamente pequeños. Esto pondría de manifiesto, que algunas de las tipologías sí difieren considerablemente, mientras que otras son más parecidas.

7.6. Conclusiones

En este capítulo se ha presentado una metodología para obtener grupos homogéneos de tipologías docentes en función de distintas variables descriptivas referentes al profesor o asignatura, o por lo menos evidenciar cuales de ellas son realmente influyentes, e identificar posteriormente los ítems del cuestionario que mejor discriminan entre los grupos obtenidos.

Las conclusiones más importantes derivadas del estudio desarrollado en este capítulo son:

- El método de clasificación CHAID se ha mostrado adecuado para identificar tipologías docentes en función de variables referidas al profesor y/o asignatura. Sin embargo, en casi todos los análisis desarrollados, algunos de los grupos obtenidos eran susceptibles de seguir ramificándose, ya que la ramificación es finalizada cuando ningún predictor es encontrado para mejorar la varianza residual explicada dentro de cada grupo. Esto sugiere que la batería de predictores podría ser mejorada. Es decir, que las variables descriptivas utilizadas no son suficientes para estimar la calidad docente, existiendo muchas otras variables no registradas que podrían tener una influencia relevante.
- Existen influencias jerarquizadas de las variables descriptivas o factores utilizados sobre la calidad docente percibida por los alumnos. Es decir, además de percibirse en las diversas variables la distinta capacidad explicativa, ésta está condicionada por los niveles presentados por las variables independientes a las que la metodología asigna un rango en su influencia sobre la variable criterio. En este sentido, los factores con efecto más significativo sobre la pertenencia a un grupo u otro de calidad docente son el departamento y la titulación, frente a otros como sexo, edad, etc.
- El principal inconveniente de esta técnica es que su desarrollo (crecimiento del árbol) está fuertemente influenciado por los datos originales. Ello explica que en la medida en la que en estos datos de partida, se incluyen circunstancias especiales, éstas afectarán a las futuras posibilidades predictivas del método, desviándolo notoriamente. De ello se desprende, que en nuestro caso la clasificación obtenida en los últimos niveles de los árboles presentados no es representativa.
- Se pone de manifiesto que la calidad docente del profesorado valorada mediante el CEDA, no puede ser tratada como un grupo homogéneo sino que está

integrado por subconjuntos diferenciados entre sí y que deben ser tenidos en cuenta a la hora de extraer conclusiones o desarrollar otros estudios globales.

- La identificación de los ítems del cuestionario que mejor caracterizan cada uno de los nodos obtenidos se consiguió mediante la aplicación de un análisis discriminante con la variable criterio categorizada. En la gran mayoría de los casos aparecen ítems correlacionados con la primera y segunda dimensión, especialmente los codificados como R1, R12, R2 y R11, lo que indicaría que las otras dos dimensiones no son consideradas relevantes a la hora de establecer las tipologías docentes, en función de las valoraciones de los estudiantes.
- La aplicación del método SIMCA permitió detectar los ítems de mayor poder discriminante entre los modelos propuestos (correspondientes a cada uno de los nodos) y obtener las distancias obtenidas entre modelos. Este método multivariante se ha propuesto como una técnica muy adecuada para comparar las tipologías obtenidas en la segmentación proporcionada por el método CHAID. Sin embargo, este método por sí solo no proporciona resultados satisfactorios en la clasificación con las variables consideradas.
- La aplicación de la metodología propuesta se ha realizado sobre bases de datos incompletas, lo que ha implicado la utilización con fines descriptivos de las técnicas y, no tanto, con fines predictivos. Sin embargo, la aplicación combinada de los métodos propuestos (Análisis Clúster, CHAID y SIMCA) sobre bases de datos completas permitiría no sólo identificar tipologías docentes, sino modelizar los datos para obtener una adecuada clasificación de observaciones futuras.
- Finalmente, comentar que las metodologías estadísticas utilizadas en este capítulo han puesto de manifiesto su potencial indiscutible en este ámbito. Sin embargo, debe mejorarse sustancialmente la calidad de los datos de partida con el fin de obtener resultados más representativos de su aplicación.

8. Conclusiones y valoración final

8.1. Conclusiones Generales

Si bien en los capítulos precedentes se ha dedicado un apartado donde se resumen los resultados y conclusiones obtenidos en cada uno de ellos, a continuación y a modo de síntesis se recogen las principales conclusiones que se derivan de la aplicación de métodos multivariantes al estudio de análisis del cuestionario de evaluación docente de la UPV mediante opinión de los alumnos.

La primera conclusión sobre la aplicación de la metodología del análisis factorial es que se comprueban y superan satisfactoriamente todos los tipos de análisis sobre la pertinencia, fiabilidad y validez de la matriz de datos.

La aplicación del análisis factorial, tanto con fines exploratorios como con fines confirmatorios permitió definir la dimensionalidad del cuestionario y la estructura en dichas dimensiones. En este sentido, destacar que el análisis factorial confirmatorio, basado en los modelos de ecuaciones estructurales, se ha manifestado como una técnica imprescindible en la determinación de la dimensionalidad de los cuestionarios, ya que permite escoger los modelos que mejor representan la estructura de los datos.

La matriz de datos quedó conformada por 5 dimensiones de calidad docente en los datos del grupo A (*Dominio, organización y claridad de la asignatura y del programa; Evaluación; Interrelación profesor-alumno; Recursos; Interrelación con otras asignaturas*) y 4 en los del grupo B (*Dominio, organización y claridad de la asignatura y del programa; Interrelación profesor-alumno; Recursos; Interrelación con otras asignaturas*).

En ambos casos todos los factores están correlacionados entre sí. Aunque en casi todas las ocasiones se obtienen valores de cargas factoriales superiores a 0,5, los ratios predominantes, para el grupo A, que son los más globales al incluir la dimensión de evaluación, son: el ítem R1 (*“El profesor explica con orden y claridad”*) con la primera dimensión, los ítems R13 (*“El examen se centra en lo fundamental de la asignatura y no en aspectos poco tratados”*) y R14 (*“En general es justo calificando”*) con la segunda, la

Conclusiones y Valoración final

variable R10 (“*Dialoga con los alumnos sobre la marcha de las clases, tomando en cuenta sus opiniones*”) con la tercera, el ítem R16 (“*Los materiales utilizados y/o recomendados han sido de gran ayuda para preparar la asignatura*”) en la cuarta y el ítem R8 (“*Establece conexiones con los contenidos de otras asignaturas*”) en la quinta, lo que coincide con el marco teórico tradicional. En el grupo B, los ítems indicadores serían los mismos, excepto como es lógico, los que corresponden a la dimensión de evaluación que no existen.

La metodología del análisis clúster, ha permitido la identificación de tres tipologías en función del nivel de satisfacción de la docencia recibida. Esta diferenciación se produce en todos los ítems del cuestionario, es paralela y se mantiene en diversos periodos evaluativos, lo que nos confirma la estabilidad del cuestionario utilizado.

De la caracterización de los clusters en base a los ítems, realizada mediante el análisis discriminante, se dedujeron dos resultados globales:

- Los ítems con mayor poder discriminante, entre los clusters encontrados, son la consideración global de buen profesor (R19), motivar a los alumnos por la asignatura (R12), dialogar con los alumnos y considerar sus opiniones (R10), sintetizar los conceptos más importantes (R4), llevar un ritmo de la clase adecuado (R3), tener bien preparadas las clases (R2) y haber preparado y recomendado materiales útiles para la asignatura (R16). En este conjunto de variables quedan representadas las dimensiones extraídas mediante el Análisis Factorial efectuado en el capítulo cinco, para conocer la estructura subyacente del cuestionario, siendo algunas de ellas las variables indicadoras de cada dimensión.
- En segundo lugar, se detectó que el grupo que presenta mayores discrepancias respecto a los otros dos en la globalidad de los ítems es el grupo de los profesores con puntuaciones medias bajas, mientras que el grupo de profesores con puntuaciones medias altas es el que presenta menor variabilidad en todos los ítems.

La aplicación del análisis CHAID Exhaustivo para estudiar las características descriptivas referentes al profesor y asignatura que más influyen en la diferenciación de las

tipologías docentes obtenidas, puso de manifiesto que sí es posible segmentar los datos en función de algunas de ellas, especialmente la variables departamento, titulación y curso, lo cual podría estar sugiriendo que las valoraciones de los alumnos dependen de cada facultad, del departamento que imparte la asignatura y del curso, por la naturaleza de las disciplinas y el interés por las mismas, y se perciben como categorías diferenciadoras entre la opinión manifestada por los alumnos. También se comprobó, especialmente en las titulaciones de segundo ciclo, que otras características como si el profesor es doctor o no pueden resultar significativas.

Por tanto, la calidad docente del profesorado valorada mediante el CEDA no puede ser tratada como un grupo homogéneo sino que está integrado por subconjuntos diferenciados entre sí y que deben ser tenidos en cuenta a la hora de extraer conclusiones o diseñar planes de mejora de la calidad educativa. En este sentido, las técnicas estadísticas propuestas pueden ser de gran utilidad, ya que son capaces de detectar y caracterizar estos grupos.

El estudio de los nodos terminales obtenidos en la clasificación desarrollada con el algoritmo CHAID ha permitido identificar y caracterizar tipologías docentes obtenidas en función de las variables descriptivas utilizadas en la segmentación. No obstante, si en un futuro se amplían el número de variables referidas a la asignatura o al profesor, la caracterización de dichas tipologías puede modificarse, ya que el algoritmo permite detectar interacciones de orden elevado.

En cada una de las tipologías encontradas, puede determinarse qué ítems del cuestionario tienen más relevancia en la valoración global otorgada. El método adecuado para ello es el análisis discriminante con una variable clasificadora categorizada.

En este caso se han encontrado variables que, si bien, en los análisis factoriales Análisis Factoriales constituían los ítems de mayor correlación con alguna de las dimensiones subyacentes, no aparecen incluidos en ninguna de las tipologías, como ítems más relevantes para caracterizarlas, como el ítem R16 o el R8, es decir, que estos ítems no presentan un elevado poder para discernir entre los sujetos que pertenecen a cada uno de los

Conclusiones y Valoración final

grupos. Esto nos indicaría que son ítems esenciales para representar la estructura subyacente de los datos, pero no para discernir entre distintas tipologías.

En los análisis discriminantes realizados, se detectó que aquello que predice en mayor medida la valoración o consideración global de buen profesor, tanto en los datos totales como en las agrupaciones efectuadas en función de distintas variables descriptivas, es el ítem R12, *consigue que los alumnos estén motivados por la asignatura* y el ítem R1, *explica en clase con orden y claridad*. Este dato pone de manifiesto la importancia que dan los alumnos a que el profesor sepa mantenerles motivados y despierte su interés por la asignatura correspondiente, lo que debería ser tenido en cuenta a la hora de elaborar programas y acciones de mejora de la calidad docente.

Finalmente, la aplicación del método SIMCA no se ha efectuado con la finalidad de clasificar y modelizar las observaciones, sino que se presenta como una potente herramienta para obtener distintos resultados prácticos. En primer lugar, el valor del poder discriminante permitió evaluar si existen ítems en el cuestionario que permiten diferenciar entre los distintos perfiles de tipologías docentes obtenidos mediante el algoritmo CHAID, cuáles son éstos y si son diferentes en función de las clases o modelos que se estén considerando. En segundo lugar, el valor de distancia entre los modelos definidos para los distintos departamentos, titulaciones y cursos, pone de manifiesto que algunas de las tipologías sí difieren considerablemente, mientras que otras son más parecidas.

A la vista de los resultados obtenidos, se hace necesario realizar algunas reflexiones a considerar en los futuros procesos de evaluación de la calidad docente:

- Del estudio se concluye, entre otros aspectos, que la opinión de los alumnos en torno a las conductas docentes de sus profesores permiten establecer diferencias sobre el desempeño docente individual entre profesores de una misma titulación y entre distintas titulaciones, y entre profesores de un mismo departamento y entre distintos departamentos. Esto podría ser un elemento que apoya la utilidad del CEDA como instrumento para recoger información que diagnostique la realidad personal de cada profesor y, con ello, se tenga una fuente de datos que permita la

reflexión personal de los profesores y también por parte de la institución respecto de sus profesores. Estas últimas acciones deberían servir de base para establecer actividades, personales e institucionales, de mejoramiento y perfeccionamiento de los profesores en áreas específicas del quehacer docente, finalidad primordial que se persigue con este tipo de instrumento.

- La inexistencia de un perfil docente generalizado supone encontrarse con posiciones muy dispares respecto a las características o cualidades que deben poseer los profesores universitarios. En este sentido, el perfil integral del profesor universitario podría ser definido como el conjunto organizado y coherente de atributos altamente deseables en el mismo, los cuales se expresarían en los conocimientos, destrezas, actitudes y valores que le permiten desempeñarse efectivamente y, con sentido creador y crítico, en las funciones de la docencia, investigación/creación, gestión y servicio que corresponden a su condición académica. En el caso de la función docente, objeto de nuestro estudio, podemos concluir que las características más deseables por los alumnos respecto a sus profesores son: que consiga motivar a los alumnos por la asignatura, que dialogue con los alumnos y considere sus opiniones, que explique en clase con orden y claridad y sintetice los conceptos más importantes, que tenga bien preparadas las clases y lleve un ritmo de las mismas adecuado, y que los materiales recomendados sean útiles para preparar la asignatura. Sin embargo, nos encontramos que estas cualidades pueden variar según la tipología del profesor considerado.
- Si se descubre cómo son los buenos profesores a juicio de sus alumnos se pueden establecer estrategias de perfeccionamiento para lograr dichas actividades o características. La excelencia docente no debe ser una propiedad difusa y utópica para la organización. Si existen criterios de excelencia claramente compartidos y establecidos, existirá un compromiso de la comunidad educativa en el logro efectivo de los niveles de excelencia deseados. Este compromiso pasará a ser un elemento integrante y catalizador del proyecto educativo. Por ello, aunque el marco de referencia pueda ser el mismo, tanto los elementos a valorar y su

ponderación, como los procedimientos a utilizar para evaluar a unos y a otros, no pueden ser los mismos, ya que las situaciones y ámbitos de desarrollo cambian según las distintas características contextuales y personales. Sólo en la medida en la que los procedimientos e instrumentos de recogida de la información se realicen teniendo en cuenta los objetivos, actividades y normas académicas de la titulación, centro, departamento y curso donde el profesor presta sus servicios, percibiremos con mayor claridad la coherencia que debe existir en toda evaluación y las posibles áreas de mejora.

- La metodología propuesta se manifiesta como una potente herramienta, para poder extraer, analizar e interpretar la información contenida en el CEDA y, en consecuencia poder definir tipologías docentes. En relación con la orientación de esta metodología, es conveniente indicar que, no sólo es aplicable a la evaluación de la función docente mediante opinión de los alumnos, sino que puede hacerse extensiva a otros tipos de evaluaciones como la de los colegas o superiores y a otro tipo de funciones del profesor universitario como la investigadora o la de gestión. Sin embargo, debe considerarse que dependiendo de las modalidades en las que se desarrolle, los parámetros, indicadores y variables señaladas en el estudio podrían variar. De esta manera es como la metodología pretende convertirse en una guía orientadora, que sujeta a posibles modificaciones futuras, facilite su aplicación y utilización en las instituciones, siempre en aras de una mejora continua.

8.2. Limitaciones del Estudio

Para una mejor valoración final del trabajo es necesario identificar aquellas restricciones que afectan a su planteamiento y desarrollo, limitando por tanto los resultados y conclusiones obtenidas. Algunas de ellas ya se han ido comentando a lo largo del trabajo.

El trabajo ha planteado desarrollar una metodología estadística para extraer, analizar e interpretar la información contenida en el CEDA, con la finalidad de poder utilizar una herramienta determinante para la evaluación del profesorado con vistas a

conseguir mejorar la calidad docente. Sin embargo, para evaluar la docencia es necesario considerar otros factores que relacionen la actuación docente con el entorno de la universidad y las tareas de investigación y gestión. La labor docente del profesor universitario no debe considerarse de un modo aislado sino inmerso en un entorno concreto, considerando todas sus actividades y recursos en un determinado contexto espacial y temporal. Las relaciones entre las condiciones estructurales del marco externo e interno de la universidad, la calidad docente a partir del conjunto de recursos y capacidades disponibles y los resultados obtenidos son difíciles de analizar, no sólo por el alto número de variables relevantes a considerar y por la complejidad de sus vinculaciones, sino porque dichas interdependencias son consecuencia de un proceso dinámico y acumulativo.

Una primera limitación surge por lo tanto de la propia formulación de la investigación. Con el planteamiento adoptado no se pretende averiguar cuál es el proceso y cualidades por el cuál los profesores son valorados de manera diferente mediante la opinión de los alumnos, sino que se trata de identificar cuáles son las características y las causas que explican un rendimiento superior en unos docentes en un momento determinado, para poder utilizar ésta información con fines de mejora continua.

Por otra parte, pretender y modelizar la calidad docente universitaria en toda su dimensión y complejidad, quizás es imposible. La aproximación al problema considerado en este trabajo, debe considerarse necesariamente parcial. En este sentido una segunda limitación que surge es la que deriva de la selección de los factores o variables introducidos en el cuestionario y de los que se disponía información y han sido incluidos en el estudio. La relación de posibles factores a considerar para valorar la calidad del profesor universitario es evidentemente más amplia, a pesar de que se es consciente de la dificultad de la medición de los mismos.

Por último, una tercera restricción es la relacionada con las bases de datos utilizadas. Las propias limitaciones de los datos han impedido hacer análisis más profundos y detallados, aunque existen muchas posibilidades de ampliar estos resultados con unas bases de información más amplias, representativas y elaboradas.

Por tanto, los resultados obtenidos corresponden a unas muestras reducidas en tamaño y, en consecuencia, estos resultados no pueden utilizarse para obtener conclusiones definitivas respecto al comportamiento medio de los profesores universitarios en la UPV. No obstante, y dado que el objetivo fundamental del estudio era la puesta en marcha un procedimiento de evaluación de la calidad del profesor universitario, consideramos que dicho objetivo si ha cubierto. Sería conveniente elaborar bases de datos amplias, con curriculum lo más completos posibles de profesores. Se debería promover que nuestra universidad y, en general, cualquier universidad española dispusiese de una base de datos con información exhaustiva de todos los profesores adscritos a sus departamentos. Necesariamente esta información, debería incluir las mismas variables en todos los casos, de modo que fuese posible establecer estudios comparativos y globales.

8.3. Líneas Futuras de Investigación

Entre las posibles líneas que quedan abiertas con este trabajo podemos destacar las siguientes:

- Definir que observaciones son aberrantes y en función de que criterios, para efectuar el estudio de las mismas.
- Determinar el punto de corte en las puntuaciones obtenidas en los ítems del cuestionario a partir del cual podríamos considerar a los profesores como excelentes y determinar las características más relevantes de este grupo de profesores.
- Efectuar el reescalado de las puntuaciones obtenidas, definiendo una función que nos permita transformar a una nueva escala cualquier posible valoración, ya que se ha puesto de manifiesto que los alumnos tienden a valorar al alza y utilizan en muy pocas ocasiones los valores extremos de la escala actual.
- Aplicar de manera combinada el método SIMCA y el método CHAID sobre bases de datos completas para identificar tipologías docentes y modelizar los datos para obtener una adecuada clasificación de observaciones futuras.

9. Bibliografía

AAVV. (2002). Marco general de la evaluación del profesorado. *Agència per a la Qualitat del Sistema Universitari a Catalunya*. Barcelona.

Abalde, E.; De Salvador, X.; Gonzalez Cabanach, R. y Muñoz Cantero, J.M. (1995). Análisis de la evaluación de la docencia universitaria por los alumnos en la Universidad de La Coruña (1993-1994). En *A. (Comp) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagogía*, 289-292. Valencia: AIDIPE.

Aleamoni, L.M. (1974). The usefulness of student evaluations in improving college teaching. *Urbana Measurement and Research Division. Office of Instructional resources, University of Illinois*.

Aleamoni, L.M. (1978b). Development and factorial validation of the Arizona Course/Instructor Evaluation Questionnaire. *Educational and Psychological Measurement*, 38, 1063 - 1067.

Aleamoni, L.M. (1981). Student ratings of Instruction. En *J. Millman (Eds.), Handbook of Teacher Evaluation*. Beverly Hills, C.A: Sage.

Altschuid, J.W y Zheng, H.Y. (1995). Assessing the effectiveness of research organizations. *Evaluation Review*, 19, (2), 197 - 216.

Aparicio, J.J. (1991). Criterios de calidad en investigación y enseñanza universitaria. *I Congreso Internacional sobre calidad de la educación universitaria*. Puerto de Santa Maria, 6 - 8 de Marzo.

Aparicio, J.J.; Tejedor, F.J. y Sanmartin, R. (1982). La enseñanza universitaria vista por sus alumnos: un estudio para la evaluación de los cursos de enseñanza superior. *ICE-Universidad Autónoma de Madrid*. Madrid.

Bibliografía

Aparicio Izquierdo, F. y González Tirados, R.M. (1994). Calidad de la enseñanza superior y otros temas universitarios. *ICE-Universidad Politécnica de Madrid. Madrid.*

Apodaka, P.; Arbizu, F.; Grao, J.; Hornilla, T. y Olalde, C. (1990). Experiencias evaluativas en la Universidad del País Vasco: consideraciones en torno a la puesta en marcha del proceso. *Revista Española de Pedagogía, 48, (186), 327-336.*

Arbuckle, J.L. (1994b). A permutation test for analysis of covariance structures. *Presented at the annual meeting of the Psychometric Society, University of Illinois, Champaign, IL.*

Arreola, R.A. (1987a). Faculty evaluation model for community and Junior Colleges. *New Directions for teaching and learning. Techniques for evaluation and improving instructions, 31, 65 - 74.*

Arreola, R.A. (1987b). The role of student government in faculty evaluation. *New Directions for teaching and learning. Techniques for evaluating and improving instructions, 31, 39 - 46.*

Arubayi, E. (1986). Student's evaluations of instruction in higher education: A review. *Higher Education, 11, 1-10.*

Avi-Itzhak, T. y Kremer, L. (1986). An investigation into the relationship between university faculty attitudes toward student rating and organizational and background factors. *Educational Research Quaterly, 10, (2), 31 - 38.*

Barber, L.W. (1990). Seft-assessment. *En J. Millman y L. Daling Hammond (Eds.) The new handbook of teacher evaluation, 216-228. Newbury Park, CA: Sage Publications.*

Barr, A.S. (1948). The measurement and prediction of teaching efficiency: A summary of investigations. *Journal of Experimental Education, 16, 203 - 283.*

Batista Foguet, J.M. y Coenders Gallart, G. (2000). Modelos de Ecuaciones Estructurales. *Cuadernos de Estadística. (Eds.) La Muralla, S.A.*

Benedito, V.; Cabrera, F.; Hernández, F.; Mercadé, F. y Rodríguez, S. (1989). La evaluación del profesorado universitario. *Revista de Educación*, 290, 279-291.

Bentler, P.M. y Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Biggs, D., Ville, B. y Suen, E. (1991). A method of choosing multiway partions for classification and decision tress. *Journal of Applied Statistics*, 18, (1), 49 - 60.

BOE 25-08-1983. Ley Orgánica 11/1983, de 25 de Agosto, de Reforma Universitaria. Jefatura de Estado.

BOE 3-07-85. Ley Orgánica 8/1985, Reguladora del Derecho a la Educación (LODE).

BOE 9-9-1989. Real decreto 1086/1989, de 28 de agosto, sobre Retribuciones del Profesorado Universitario.

BOE 30-06-1990. Resolución de 20 de junio de 1990, del Consejo de Universidades, por la que se establecen los criterios generales para la evaluación de la actividad docente del profesorado universitario prevista en el artículo 2. 3. c) del Real Decreto 1086/1989, de 28 de agosto, sobre retribuciones del profesorado universitario.

BOE 3-10-90. Ley Orgánica 1/1990, de Ordenación General del Sistema Educativo (LOGSE).

BOE 21-12-95. Ley Orgánica 9/1995, de la Participación, la Evaluación y el Gobierno de los centros docentes (LOPEG).

BOE 9-12-95. Real decreto 1947/1995, de 1 de diciembre, por el que se establece el Plan Nacional de Evaluación de la Calidad de las Universidades.

BOE 21-12-2001. Ley Orgánica 6/2001, de 21 de Diciembre, de Universidades. Jefatura de Estado.

Bibliografía

BOE 21-04-01. Real Decreto 408/2001, de 20 de abril, por el que se establece el II Plan de la Calidad de las Universidades.

BOE 20-06-2002. Ley Orgánica 5/2002, de 19 de junio, de las Cualificaciones y de la Formación Profesional.

BOE 24-12-2002. Ley Orgánica 10/2002, de 23 de diciembre, de Calidad de la Educación.

Bollen, K.A. y Long, J.S. (1993). Testing structural equation models. (Eds.) *Newbury Park, California: Sage*.

Bonetti, S. (1995). On the use of student questionnaires. *Higher Education Review*, 26, (3), 57 - 64.

Borich, G.D. y Madden, S.K. (1997). Evaluating classroom instruction: A sourcebook of instruments. (Eds.) *Reading, Massachusetts: Addison-Wesley*.

Borrell, N. (1995). Modelos para la evaluación externa e interna de los centros docentes. En Medina, A. y Villar, L. M. (Coord.), *Evaluación de Programas Educativos, Centros y Profesores.*, 275-302. (Eds.) *Universitas, S.A. Madrid*.

Boyce, A.C. (1915). Method for measuring teacher's efficiency. *Bloomington, Illinois: Public School Publishing Co.*

Bradburd, R.M. y Mann, D.P. (1993). Wealth in higher education institutions. *Journal of Higher Education*, 64, (4), 472 - 493.

Braskamp, L.A.; Brandenburg, D.C. y Ory, J.C. (1984). Evaluating teaching effectiveness: A practical guide. (Eds.) *California: Sage*.

Brown, M.W y Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. and Long, J.S. (Eds.) *Testing structural equation models*, 136-162. *Newbury Park, California: Sage*.

Cajide, J. (1994). Análisis factorial confirmatorio de las características de calidad docente universitaria (solución LISREL). *Bordón*, 46, (4), 389-405.

Cashin, W.E. (1983). Concerns about using student ratings in community colleges. *New Directions for Community Colleges*, 11, (1), 57 - 65.

Cattell, R.B. (1972). El análisis científico de la personalidad. Los tests de personalidad. (Eds.) Fontanella. Barcelona.

Centra, J.A. (1974). The relationship between student and alumni rating of teachers. *Educational and Psychological Measurement*, 34, 321-326.

Centra, J.A. (1988). Determining faculty effectiveness. (Eds.) San Francisco-Londres: Jossey-Bass.

Centra, J.A. y Creech, F.R. (1976). The relationship between student, teacher and course characteristics and student rating of teacher effectiveness. (Eds.) Princeton, N. J., Educational Testing Service.

Chiu, T.; Fang, D.; Chen, J.; Wang, Y. y Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of Seventh ACM SIGKDD International Conference on knowledge discovery and data mining*, 263-268. San Francisco.

Clogg, C.C. y Eliasin, S.R. (1987). Some problems in log-linear analysis. *Sociological Methods and Research*, 16, (1), 8 - 44.

Cohen, P.A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 13, 321 - 341.

Comisión de Coordinación Técnica. (2001). Guía de Evaluación. Plan Nacional de Evaluación de la Calidad de las Universidades. (Eds.) Ministerio de Educación y Cultura. Consejo de Universidades. Secretaría General. Madrid: MEC.

Bibliografía

Cook, T.D. (1989). Improving the quality of student ratings of instruction: a look at two strategies. *Research in Higher Education*, 30, (1), 31 - 45.

Coomans, D.; Broeckart, I.; Derde, M. P.; Tassin, A.; Massart, D. L. y Wold, S. (1984). Use of a Microcomputer for the Definition of Multivariate Confidence Regions in Medical Diagnosis Based on Clinical Laboratory Profiles. *Computers and Biomedical Research*, 17, (1), 1-14.

Costin, F., Greenough, W. T. y Menges, R. T. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research*. 41, 511-535.

Creager, J.A. (1950). A multiple factor analysis of the Purdue rating scale for instructors. *Purdue University Studies in Higher Education*, 70, 75 - 96.

Cronbach, L.J. (1984). Essentials of psychological testing. (Eds.) New York: Harper & Row.

Cruse, D.B. (1987). Student evaluations and the university professor. *Higher Education*, 15, (6), 723 - 737.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317-327.

Davey, K.B. y Sell, G.R. (1985). Instructional evaluation for development/improvement: Fact or fiction based on a case study of faculty practices? *Annual Meeting of the American Educational Research Association, Division J, 69th, Chicago, IL, March 31 - April 4.*

De la Orden, A. (1987). Evaluación, selección y promoción del profesorado universitario. *Bordón*, 39, (1), 5 - 29.

De la Orden, A. (1992). Calidad y Evaluación de la Enseñanza Universitaria. *Congreso Internacional de las Universidades. Madrid.*

De la Orden, A. (1993). La escuela en la perspectiva del producto educativo. Reflexiones sobre evaluación de centros docentes. *Bordón*, 45, (3), 263-270.

De Miguel, M. (1988a). Aspectos Metodológicos de la Investigación Educativa. En *Dendaluze, I (Coord): Aspectos Metodológicos de la Investigación Educativa. (Eds.) Narcea. Madrid.*

De Miguel, M. (1988b). Paradigmas de la Investigación Educativa Española. En *Dendaluze, I (Coord): Aspectos Metodológicos de la Investigación Educativa. (Eds.) Narcea. Madrid.*

De Miguel, M. (1989a). Metodología de la Investigación Participativa y Desarrollo Comunitario. *Actas de la II Jornada sobre la Educación Permanente. UNED. Centro Asociado de Asturias.*

De Miguel, M. (1989b). Modelos de Investigación sobre organizaciones educativas. *Revista de Investigación Educativa*, 7, (13), 21-56.

De Miguel, M. (1991). Indicadores de Calidad de la Docencia Universitaria. *I Congreso Internacional sobre calidad de la educación universitaria. Puerto de Santamaría (Cádiz), 6 - 8 de marzo.*

De Miguel, M. (1997). La Evaluación de los centros educativos. Una aproximación a un enfoque sistémico. *Revista de Investigación Educativa*, 15, (2), 145-178.

De Miguel, M. (1998). La evaluación del profesorado: criterios y propuestas para mejorar la función docente. *Revista de Educación*, 315, 67-83.

De Miguel, M; Mora, J y Rodríguez, S. (1991). La Evaluación de las Instituciones Universitarias. (Eds.) *Consejo de Universidades, Secretaría General. Madrid: MEC.*

Bibliografía

De Nevé, H.M.F. (1991). University teacher's thinking about lecturing: student evaluation of lecturing as an improvement perspective for the lecturer. *Higher Education*, 22, (1), 63 - 91.

Dowell, D.A. y Neal, J.A. (1982). The validity and accuracy of student ratings of instruction: A reply to Peter A. Cohen. *Journal of Higher Education*, 54, 459 - 463.

Doyle, K.O. (1975). Student evaluation of instruction. (Eds.) *Lexington, M.A.: Lexington Books*.

Doyle, K.O. (1983). Evaluating teaching. (Eds.) *Lexington, M.A.: Lexington Books*.

Doyle, P. y Fenwick, I. (1975). The pitfalls of AID analysis. *Journal of Marketing Research*, 12, 408 - 413.

Drucker, A.J.A. y Remmers, H.H. (1950). Do alumni and students differ in their attitudes toward instructors? *Purdue University Studies in Higher Education*, 70, 62 - 64.

Duncan, O. D. (1975). Introduction to structural equation models. (Eds.) *New York: Academic Press*.

Einhorn, H.J. (1972). Alchemy in the behavioural sciences. *Public Opinion Quarterly*, 36, 367 -378.

El-Hassan, K. (1995). Student's ratings of instruction: generalizability of findings. *Studies in Educational Evaluation*, 21, 411 - 429.

Elliot, D.N. (1950). Characteristics and relationships of various criteria of college and university teaching. *Purdue University Studies in Higher Education*, 70, 5 - 61.

Esbensen, K. and Geladi, P. (1990). The start and early history of chemometrics: selected interviews. Part 2. *Journal of Chemometrics*, 4, 389-412.

Escudero, T. (1979). La Evaluación de los centros educativos. Una aproximación a un enfoque sistémico. *Tesis Doctoral. ICE-Universidad de Zaragoza. Zaragoza.*

Escudero, T. (1986). Algunos criterios y evidencias del rendimiento universitario. *En M. Latiesa (Eds.), Demanda de la educación superior y rendimiento académico en la universidad, 187 - 204. Madrid: CIDE - Consejo de Universidades, MEC.*

Escudero, T. (1987). Experiencias evaluativas en la Universidad de Zaragoza. *En Varios (Eds.), Consideraciones Metodológicas sobre Evaluación y Mejora de la Docencia Universitaria. Servei de formació permanent. Universitat de València. Valencia.*

Escudero, T. (1988). Evaluación del rendimiento institucional universitario. Entre la utopía y el positivismo. *IV Seminario de Modelos de Investigación Educativa. La evaluación de centros: nuevas perspectivas en la investigación sobre organizaciones educativas. Santiago de Compostela.*

Escudero, T. (1989a). Aproximación pragmática a la evaluación de la universidad. *Revista de Investigación Educativa, 7, (13), 93-112.*

Escudero, T. (1989b). Experiencias Evaluativas en la Universidad de Zaragoza. *Informes de Investigación evaluativa nº 1. Consideraciones Metodológicas sobre la Evaluación y Mejora de la Docencia Universitaria, 89-132. Servei de formació permanent. Universitat de València. Valencia.*

Escudero, T. (1991). Enfoques Modélicos en la Evaluación de la Enseñanza Universitaria. *III Jornadas Nacionales de Didáctica Universitaria. Evaluación y Desarrollo Profesional. Universidad de las Palmas, Servicio de Publicaciones. Las Palmas de Gran Canarias, 23 - 26 de Septiembre.*

Escudero, T. (1996). Proyecto docente de cátedra de Universidad no publicado. *Universidad de Zaragoza. Zaragoza.*

Bibliografía

Escudero, T. (1999). Indicadores de rendimiento académico: Una experiencia en la Universidad de Zaragoza. *En Consejo de Universidades (Comp.), Indicadores en la universidad: Información y decisiones*, 251-262. Madrid: MEC

Escudero, T.; Garcés, R.; Palacián, E.; Blanco, F. y Bueno, C. (1996). Bases para un plan estratégico de la Universidad de Zaragoza: *Encuesta sobre acciones estratégicas*. ICE-Universidad de Zaragoza. Zaragoza.

Escudero, T.; Morales, J.; Gastaminza, R.; Barral, M.J.; Lamata, F.; Romero, S.; Conde, B.; Garcés, R. y Palacián, E. (2000). Evaluación de las prácticas en la licenciatura de Medicina. ICE. Universidad de Zaragoza.

Everitt, B. (1980). Cluster Analysis. (Eds.) New York: Halsted Press.

Feldman, K.A. (1976). The superior college teacher from the student's view. *Research in Higher Education*, 5, 243-288.

Feldman, K.A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*, 6, 223-274.

Feldman, K.A. (1978). Course characteristics and college student's rating of their teachers: what we know and what we don't. *Research and Higher Education*, 9, 199-242.

Feldman, K.A. (1979). The significance of circumstances for college student's ratings of their teachers and courses. *Research in Higher Education*, 10, 149 - 172.

Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, (1), 3 - 124.

Feldman, K.A. (1984). Class size and college student's evaluations of teachers and courses: A closer look. *Research in Higher Education*, 21, 45 - 116.

Fernández, A.; Maiques, J.M. y San Onofre, V. (1991). Evaluación de la calidad de la docencia: la experiencia de la U.P.V. *Actas de I Congreso Internacional sobre Calidad de la Enseñanza Universitaria*, 116-125. ICE-Universidad de Cádiz. Cádiz

Fernández Ballesteros, R. (1995). Evaluación de programas: Una guía práctica en ámbitos sociales, educativos y de salud. (Eds.) *Síntesis*. Madrid.

Fernández Díaz, M.J. (1988). Investigación evaluativa de instituciones universitarias. *Tesis Doctoral*. Universidad Complutense de Madrid.

Fernández Sánchez, J. (1988). La evaluación de la enseñanza universitaria: la experiencia de la Complutense. *Studia Paedagogica*, 20, 135 - 147.

Fernández Sánchez, J. (1991). Evaluación de la calidad docente de la universidad complutense (1989 -1990). *Gaceta de la Complutense*, 74, 12-15. Universidad Complutense. Madrid.

Fernández Sánchez, J. (1992). Evaluation and decision making in the Complutense. *University of Madrid. Higher Education Management*, 4, (3), 336 - 345.

Fernández Sánchez, J.; Lillo, B. y Martínez, M.R. (1987). Proyecto de evaluación de la educación superior española. *Informes de Investigación Evaluativa nº 1. Consideraciones Metodológicas sobre Evaluación y Mejora de la Docencia Universitaria*, 51-88. Universidad de Valencia. Valencia.

Fernández Sánchez, J. y Mateo, M.A. (1991). Student evaluation of university teaching quality: analysis of a questionnaire. *Educational and Psychological Measurement*, 51, (3), 776 – 192-

Fernández Sánchez, J. y Mateo, M.A. (1994). The circular evaluative process of higher education: A Spanish sample. En *OECD (Eds.), Evaluation and decision making process in higher education: French, German and Spanish experiences*. Paris.

Bibliografía

Fernández Sánchez, J.; Mateo, M.A. y Muñoz, J. (1995). Evaluation of the academic setting in Spain. *European Journal of Psychological Assessment*, 11, 134 - 138.

Fernández Sánchez, J.; Mateo, M.A. y Muñoz, J. (1996). Valoración por parte del profesorado de la evaluación docente realizada por los alumnos. *Psicothema*, 8, (1), 167 - 172.

Ferrández, M.R. (1991). Sistemas de evaluación de la docencia práctica: el caso de la facultad de medicina de la universidad de València. *Tesis de Licenciatura. Universitat de València*.

Ferrández, M.R.; Jornet, J.M.; Pérez Carbonell, A. y González Such, J. (1995a). Estudio del Cuestionario de Evaluación de la Docencia a partir de Opiniones de Estudiantes en la Universitat Jaume I: (I) El Cuestionario y el Planteamiento Metodológico *En A. (Comp) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagógica*, 296-300. Valencia: AIDIPE.

Ferrández, M.R.; Jornet, J.M.; Pérez Carbonell, A. y González Such, J. (1995b). Estudio del Cuestionario de Evaluación de la Docencia a partir de Opiniones de Estudiantes en la Universitat Jaume I: (II) Estudio Factorial. *En A. (Comp) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagógica*, .304-308. Valencia: AIDIPE.

Frank, I. E. y Friedman, J. H. (1989). Classification: oltimers and newcomers. *Journal of Chemometrics*, 3, 463-475.

Franke-Wikberg, S. (1990). Evaluating education quality on the institutional level. *Higher Education Management*, 2, (3), 271 - 292.

French-Lazovik, G. y Gibson, C.L. (1984). Effects of verbally labeled anchor points an the distributional parameters of rating measures. *Applied Psychological Measurement*, 8, (1), 49 - 57.

García Jiménez, E.; Gil Flores, J. y Rodríguez Gómez, G. (2000). Análisis Factorial. *Cuadernos de Estadística. (Eds.) La Muralla, S.A.*

García Ramos, J.M. (1987). El análisis factorial confirmatorio aplicado a la investigación pedagógica no experimental. *Bordón*, 267, 245-267.

García Ramos, J.M. (1988). Modelos Exploratorios y Confirmatorios en la Investigación Pedagógica no experimental. *Revista de las Ciencias de la Educación*, 136, 423-450.

García Ramos, J.M. (1989). Extensión de la evaluación. En *Pérez Juste, R. y García Ramos, J.M., Diagnóstico, evaluación y toma de decisiones*, 315-385. (Eds.) Rialp. Madrid.

García Ramos, J.M. (1997). El análisis factorial confirmatorio en la validación del constructo Competencia Docente del Profesor Universitario. *Bordón*, 49, (4), 361-391.

García Ramos, J.M. y Congosto, E. (2000). Evaluación y calidad del profesorado. En *González Ramirez, T. (Coord.). Evaluación y Gestión de la Calidad Educativa. Un Enfoque Metodológico*, 127-157. (Eds.) Aljibe. Málaga.

García Ramos, J.M.; Fernandez Diaz, M.J. y Congosto Luna, E. (1995a). Evaluación docente del profesorado. En *A. (Comp.) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagógica*, 271-275. Valencia: AIDIPE.

García Ramos, J.M.; Fernández Díaz, M.J. y Congosto Luna, E. (1995b). La evaluación de un modelo de tutoría académica en la Universidad. En *A. (Comp.) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagógica*, 268-271. Valencia: AIDIPE.

García Valcacer, A.; Salvador, L. y Zubieta, J.C. (1991). Elementos para análisis evaluativo de la universidad. El caso de la Universidad de Cantabria. En *(Eds.), La Investigación Educativa sobre la Universidad*, 281-323. Madrid: CIDE-Servicio de Publicaciones del MEC.

Bibliografía

Geladi, P. y Esbensen, K. (1990). The start and early history of chemometrics: selected interviews. Part 1. *Journal of Chemometrics*, 4, 337-354.

Geladi, P. y Kowalski, B. (1986). Partial least-squares regression: a tutorial. *Analytica Chemical Act*, 185, 1-32.

Gels, G. L. (1977). Evaluation: definitions, problems and strategies. *En C.K. Knapper et al. (Eds.), If teaching is important. The evaluation of instruction in Higher Education. Toronto, Canada: Clarke, Irwin.*

Gil flores, J.; García Jiménez, E. y Rodríguez Gómez, G. (2001). *Análisis Discriminante. (Eds.) La Muralla, S.A.*

Gillmore, G. M. (1984). Student ratings as a factor in faculty employment decisions and periodic review. *Journal of College and University Law*, 10, (4), 557 - 576.

Gillmore, G. M.; Kane, M.T. y Naccarato, R.W. (1978). The generalizability of student ratings of instruction: Estimating the teacher and course components. *Journal of Educational Measurement*, 15, 1-13.

Goldberger, A.S. (1964). *Econometric theory. (Eds.) New York: John Wiley & Sons.*

Goldberger, A.S. y Duncan, O.D. (1973). *Structural equation models in the social sciences. (Eds.) New York: Academic Press.*

Goldschmid, M.L. (1976). The role of institutional management in improving instruction in higher education. *Third General Conference on Institutional Management in Higer Education*, 13 - 16. Paris.

González González, D.; López Fuentes, R.; Hidalgo Díez, E.; García Lupión, B. y Pegalajar, M. (1999). Evaluación del profesorado universitario por los alumnos. *En A. (Comp.) (Eds.) Nuevas realidades educativas, nuevas necesidades metodológicas*, 223-227. *Facultad de CCEE. Universidad de Málaga: AIDIPE. CEDMA.*

González Such, J. (1997). Estudio de un instrumento para la Evaluación del Profesorado Universitario. *Tesis Doctoral. Universitat de València. Valencia.*

González Such, J.; Jornet, J.M. y Pérez Carbonell, A. (1995). Análisis de un cuestionario de evaluación de clases prácticas experimentales. *En A. (Com.p) (Eds.), Estudios de Investigación Educativa en Intervención Psicopedagogía, 296-300. Valencia: AIDIPE.*

González Such, J.; Jornet, J.M.; Pérez Carbonell, A. y Ferrandez, M.R. (1993). Factores intervinientes en la valoración del profesor por parte del estudiante. *VI Seminario sobre Modelos de Investigación Educativa, 23-25. Madrid.*

González Such, J.; Jornet, J.M.; Suárez, J.M. y Benlloch, C. (1990). Análisis de la estabilidad de niveles y dimensiones observadas en el Cuestionario de Evaluación de la Docencia a partir de las opiniones de estudiantes en dos aplicaciones anuales sucesivas en centros de áreas de Ciencias Psicopedagógicas de la Universitat de València. *Revista de Investigación Educativa, 8, (16), 689-696.*

González Such, J.; Jornet, J.M.; Suarez, J.M. y Pérez Carbonell, A. (1990). Análisis de la estabilidad de niveles y dimensiones observadas en el Cuestionario de Evaluación de la Docencia a partir de las opiniones de estudiantes en dos aplicaciones anuales sucesivas en centros de áreas de Ciencias Humanas de la Universitat de València. *Revista de Investigación Educativa, 8, (16), 681-688.*

González Such, J.; Jornet, J.M.; Suárez, J.M. y Pérez Carbonell, A. (1999). Análisis de tipologías de calidad docente a partir de un Cuestionario de Evaluación del Profesorado universitario. *Bordón, 51, (1), 95-113*

González Such, J.; Jornet, J.M.; Pérez Carbonell, A. y Villanueva, P. (1999). ¿Qué les parece a los estudiantes importante en los cuestionarios de evaluación de la docencia?. *En A. (Comp.) (Eds.) Nuevas realidades educativas, nuevas necesidades metodológicas, 223-227. Facultad de CCEE. Universidad de Málaga: AIDIPE. CEDMA.*

Bibliografía

González Such, J.; Suárez, J.M.; Jornet, J.M. y Ferrández, M.R. (1990). Análisis de la estabilidad de niveles y dimensiones observadas en el Cuestionario de Evaluación de la Docencia a partir de las opiniones de estudiantes en dos aplicaciones anuales sucesivas en centros de áreas de Ciencias de la Salud de la Universitat de València. *Revista de Investigación Educativa*, 8, (16), 643-650.

Good, T.L. y Brophy, J.E. (1986). School effects. En T.L. Good y J.E. Brophy (Eds.), *Hand book of Research on Teaching*, 570 - 602.

Good, T.L. y Murray, C. (1990). Teacher ratings: A call for teacher control and self-evaluation. En Millman y L. Darling Hammond (Eds.), *The New Handbook of Teacher Evaluation*, 191 - 215. Newbury Park, Cal.: Sage Publications.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.

Groth, R. (1997). Data mining: a hands-on approach for business professionals. (Eds.) *Upper Saddle River, N.J.: Prentice Hall*.

Guha, S.; Rastogi, R. y Shim, K. (1999). ROCK: A robust clustering algorithm for categorical attributes. *Proceedings of the 15th IEEE Conference on Data Engineering*, 512-521, 23-26 March.

Hair, J. F.; Anderson, R. E.; Tatham, R. L. y Black, W. C. (1995). Multivariate data analysis. (Eds.) *New York, Prentice Hall International, Inc.*

Hoffman, J.E. y Kremer, L. (1983). Course evaluation and attitudes toward college teaching. *Higher Education*, 12, 681 - 690.

Hoyt, D.P. y Cashin, W.E. (1977). Development of the IDEA system. *Idea report N° 1. Manhattan: Center for Faculty Evaluation and Development, Kansas State University.*

Ibañez-Martín, J.A. (1990). Dimensiones de la competencia profesional del profesor de universidad. *Revista Española de Pedagogía*, 186, 239-257.

Instituto de las Ciencias de la Educación. (1994). Informe sobre la evaluación del profesorado. *ICE -Universidad Politécnica de Valencia*.

Jackson, J.E. (1991). A user's guide to Principal Components. (Eds.) *New York: Wiley*.

Jiménez, C. (1985). La eficacia docente de la UNED: aproximación a un modelo. *Revista de Investigación Educativa*, 3, (6), 401 - 406.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects and materials. (Eds.) *New York: McGraw Hill*.

Jonhston, J. (1972). Econometric methods. (Eds.) *New York: McGraw-Hill*.

Jöreskog, K.G. (1967). Some contributions to Maximum Likelihood Factor Analysis. *Psychometrika*, 34, 183 - 202.

Jöreskog, K.G. (1969). A general approach to confirmatory factor analysis. *Psychometrika*, 34, 183-202.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.

Jöreskog, K.G. y Lawley, D.N. (1968). New methods in Maximum Likelihood Factor Analysis. *British Journal of Mathematical and Statistical Psychology*, 85-96.

Jöreskog, K.G. y Sörbom, D. (1981). Lisrel V. Analysis of linear structural relations by the method of Maximum Likelihood. (Eds.) *Chicago, Illinois: SPSS Publications. International Educational Services*.

Bibliografía

Jöreskog, K.G. y Sörbom, D. (1986). Lisrel VI. Analysis of linear structural relationships by Maximum Likelihood, Instrumental Variable, and Least Squares methods. (Eds.) Moresville, Indiana. *Scientific Software*.

Jöreskog, K. G. y Sörbom, D. (1989). Lisrel 7. A guide to the program and applications. (Eds.) Chicago, Illinois: *SPSS Publications*.

Jornet, J. (1991). Enfoques en la evaluación universitaria. *III Jornadas Nacionales de Didáctica Universitaria: Evaluación y Desarrollo Profesional*. Universidad de las Palmas, Servicio de Publicaciones. Las Palmas de Gran Canarias, 23 - 26 de Septiembre.

Jornet, J.M.; González Such, J. y Suárez, J.M. (1989). Estudio de validación de un cuestionario de valoración de la docencia universitaria para los estudiantes en un conjunto homogéneo de centros de la Universitat de València. *Revista de Investigación Educativa*, 7, (13), 57-92.

Jornet, J.M.; González Such, J.; Suárez, J.M. y Pérez Carbonell, A. (1995). Análisis de la consistencia de los cuestionarios de opinión de estudiantes para evaluaciones docentes. En A. (Comp.) (Eds.), *Estudios de Investigación Educativa en Intervención Psicopedagógica*, 260-263. Valencia: *AIDIPE*.

Jornet, J.M.; González Such, J.; Suárez, J.M. y Pérez Carbonell, A. (1996). Evaluación de la actividad universitaria. En G. *Quintás* (Eds.), *Reforma y Evaluación de la Universidad*, 182-244. Servei de publicacions de la Universitat de València. Valencia.

Jornet, J.M.; González Such, J.; Suárez, J.M.; Pérez Carbonell, A.; Ferrández, M.R. y Villanueva, P. (1993). Problemas metodológicos en la evaluación de la docencia universitaria a partir de opiniones de estudiantes. *Seminario sobre las reformas educativas actuales en España. Su evaluación: metodologías y resultados*. Madrid.

Jornet, J.M. y Suárez, J.M. (1989). Revisión de modelos y métodos en la determinación de estándares y en el establecimiento de un punto de corte en la Evaluación Referida al Criterio (ECR). *Bordón*, 41, 277-301.

Jornet, J.M. y Suárez, J.M. (1997). Aplicaciones del análisis cluster en estudios evaluativos. *Documento Doctorado. Departamento MIDE*.

Jornet, J.M.; Villanueva, P.; Suárez, J. M. y Alfaro, I. (1989). Proyecto de implantación de un sistema de evaluación del profesorado en la universidad de Valencia. *Informes de investigación evaluativa nº 1. Consideraciones metodológicas sobre la evaluación y mejora de la docencia universitaria. Universidad de Valencia. Valencia*.

Kaiser, H.G. y Caffrey, J. (1965). Alpha Factor Analysis. *Psychometrika*, 30, 1-14.

Kass, G.V. (1975). Significance testing in automatic interaction detection (AID). *Applied Statistics*, 24, 178 -189.

Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, (2), 119-127.

Kim, J.O. y Mueller, C.W. (1978). Factor analysis. (Eds.) *Statistics Methods and Practical Issues. London: Sage*.

Kmenta, J. (1971). Elements of econometrics. (Eds.) *New York: Macmillan*.

Kratz, H.E. (1889). Characteristics of the best teachers as recognized by children. *Pedagogical Seminary*, 3, 413 - 418.

Kulik, J.A. y Kulik, C.C. (1974). Students ratings of instruction. *Teaching and Psychology*, 1, 51-57.

Lawley, D.N. y Maxwell, A.E. (1971). Factor analysis as a statistical method. (Eds.) *London: Butterwoths*.

Bibliografía

Levin, H.M. (1991). Raising productivity in higher education. *Journal of Higher Education*, 62, (3), 241 - 262.

L'Hommedieu, R.; Menges, R.J. y Brinko, K.T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology*, 82, (2), 232 - 241.

Linsky, A.S. y Straus, M.A. (1973). Student evaluation of teaching. *Teaching Sociology*, 1, (1), 103 - 118.

Luxán, J.M. (1998). La Evaluación de la Universidad de España. *Revista de Educación*, 315, 11-28.

Magdison, J. (1992). Chi-squared analysis of a scalable dependent variable. *In proceedings of the Annual Meetings of the American Statistical Association, Educational Statistics Section.*

Magnusen, K.O. (1987). Faculty evaluation, performance and pay: Application and issues. *Journal of Higher Education*, 58, (5), 516 - 529.

Mahmoud, M.M. (1991). Descriptive models of student decision behaviour in evaluation of higher education. *Assessment and Evaluation in Higher Education*, 16, (2), 133 - 148.

Marsh, H.W. (1980). The influence of student, course and instructor characteristics in evaluations of university teaching. *American Educational Research Journal* 17, 219 - 237.

Marsh, H.W. (1982a). Validity of student's evaluations of college teaching: A multirait multimethod analysis. *Journal of Educational Psychology*, 2, 264 - 297.

Marsh, H.W. (1982b). SEEQ: A reliable, valid and useful instrument for collecting student's evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77-95.

Marsh, H.W. (1984). Student's evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.

Marsh, H.W. (1987a). Students evaluations of university teaching: Research, findings, methodological issues and directions for future research. *International Journal of Education Research*, 11, (3), 253-388.

Marsh, H.W. (1987b). Student's evaluations of teaching. En M.J. Dunkin (Eds.), *encyclopaedia of teaching and teacher education*, 181-187. Oxford: Pergamo. 1.

Marsh, H.W.; Fleiner, H. y Thomas, C. S. (1975). Validity and usefulness of student's evaluations of instructional quality. *Journal of Educational Psychology*, 67, 833-839.

Marsh, H.W. y Overall, J. V. (1981). The relative influence of course level, course type and instructor on student's evaluations of college teaching. *American Educational Research Journal*, 18, 103-112.

Marsh, H.W. y Roche, L. (1993). The use of students evaluations and an individually structure intervention to enhance university teaching effectiveness. *Educational Research Journal*, 30, 217-251.

Martens, H. y Martens, M. (2001). Multivariate analysis of quality. (Eds.) Baffin Lane, Chinchester, John Wiley & Sons Ltd.

Mateo, J. (1987). La evaluación del profesorado universitario. Algunas reflexiones en torno al estado de la cuestión. En Varios (Eds.) *Consideraciones Metodológicas sobre Evaluación y Mejora de la Docencia Universitaria*, 133-203. Servei de formacio permanent, Universitat de València. Valencia.

Mateo, J. (1990). Función docente y demanda discente. *Revista Española de Pedagogía*. 48, (186), 319-326.

Bibliografía

Mateo García, M.A. y Fernández Sánchez, J. (1991). La dimensionalidad de los conceptos de masculinidad y feminidad. *Investigaciones Psicológicas*, 9, 95 – 116.

Mateo García, M.A. y Fernández Sánchez, J. (1992a). Dimensiones de la calidad de la enseñanza universitaria. *Psicothema*, 5, 2, 265-275.

Mateo García, M.A.; Escudero, T.; De Miguel, M.; Mora, J.G. y Rodríguez Espinar, S. (1996a). La evaluación del profesorado. Un tema a debate. *Revista de Investigación Educativa*, 14, (2), 73-94.

Mateo García, M.A.; Escudero, T.; De Miguel, M.; Mora, J.G. y Rodríguez Espinar, S. (1996b). Evaluación del Profesorado de Secundaria. (Eds.) *Cedecs: Barcelona*.

Mateo García, M.A. y Fernández Sánchez, J. (1992b). Análisis confirmatorio de la estructura dimensional de un cuestionario para la evaluación de la calidad de la enseñanza. *Investigaciones Psicológicas*, 11, 73-82. (Eds.) *Universidad Complutense. Madrid*.

McBean, E.A. y Lennox, W.C. (1985). Effect of survey size on student ratings of teaching. *Higher Education*, 14, (2), 117 - 125.

McKeachie, W.J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384 - 397.

McKeachie, W.J. (1986). Teaching tips: A guidebook for the beginning college teacher. (Eds.) *Lexington: Mass.: Heath*.

McKeachie, W.J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology*, 82, (2), 189 - 200.

McKeachie, W.J.; Lin, Y.G.; Daugherty, M.; Neigler, C.; Nork, J.; Walz, M. y Baldwin, R. (1980). Using student ratings and consultation to improve instruction. *British Journal of Educational Psychology*, 50, 168 - 174.

McLachlan, G.J. (1992). Discriminant analysis and statistical Pattern Recognition. (Eds.) New York.: Wiley-Interscience.

McQueen, J.B. (1967). Some methods for classification and analysis of multivariate observations. In *proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.

Miller, M.T. (1971). Instructor attitudes toward, and their use of, student ratings of teachers. *Journal of Educational Psychology*, 62, (3), 235 - 239.

Miller, R.I. (1981). Simultaneous Statistical Inference, 6 - 8. (Eds.) Springer, New York.

Miller, R.I. (1987). Evaluating faculty for promotion and tenure. (Eds.) Jossey-Bass Publishers. San Francisco.

Mora Ruiz, J. G. (1995). La evaluación de las instituciones universitarias. *Revista de Investigación Educativa*, 9, (17), 27-48.

Mora Ruiz, J. G. (1998). La Evaluación institucional de la universidad. *Revista de Educación*, 315, 29-44.

Mora Ruiz, J. G. (1999). Indicadores y decisiones en las universidades. En *Consejo de Universidades (Comp.) (Eds.), Indicadores en la Universidad: Información y decisiones*, 19-30. Madrid: MEC.

Morgan J.A. and Sonquist, J.N. (1963a). Problems in the analysis of survey data: an a proposal. *Journal of American Statistical Association*, 58, 415 - 434.

Moses, I. (1986). Self and student evaluation of academic staff. *Assessment and Evaluation in Higher Education*, 76 - 78.

Muñoz, F. (1997). Indicadores educativos: modelos y ejemplos. *Cuadernos de Pedagogía*, 256, 54-61. (Eds.) Práxis. Barcelona

Bibliografía

Muñoz, J.; García, A. y Virgos, J.M. (1991). Escala de la Universidad de Oviedo para la evaluación del profesorado. *Psicothema*, 3, (2), 269 - 281.

Muñoz Cantero, J.M.; Rios de Deus, M.P. y Abalde, E. (2002). Evaluación docente vs. evaluación de la calidad. *Revista Electrónica de Investigación Educativa (RELIEVE)*, 8, (2), 102-134.

Murray, H.G. (1980). Evaluating university teaching: A review of research. (Eds.) *Toronto: Ontario Confederation of University Faculty Association.*

Murray, H.G. (1984). The impact of formative and summative evaluation of teaching in North American Universities. *Assessment and Evaluation in Higher Education*, 9, (2), 117 - 132.

Muthén, B. y Kaplan, D. (1995). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical & Statistical Psychology*, 38, 171-189.

Nimmer, J.G. y Stone, E.F. (1991). Effects of grading practices and time of rating on student ratings of faculty performance and student learning. *Research in Higher Education*, 32, (2), 195 - 215.

Norusis, M.J. (1986). SPSS/PC+. Advanced statistics for the IBM PC/TXT/AT. (Eds.) *Chicago: SPSS inc.*

Orr, L. (1972). The dependence of transition proportions in the education system on observed social factors and school characteristics. *Journal of the Royal Statistical Society*, 135, 74-95.

Ory, J.C. y Parker, S.A. (1989). Assessment activities at large, research universities. *Research in Higher Education*, 30, 375-385.

Overall, J.V. y Marsh, H.W. (1980). Student's evaluation of instructors: A longitudinal study of their stability. *Journal of Educational Psychology*, 72, 181 - 185.

Overall, J.V. y Marsh, H.W. (1982). Student's evaluation of teaching: An update. *American Association for Higher Education Bulletin*, 35, (4), 9 - 13.

Peña, D. (1997). La mejora de la calidad de la educación: reflexiones y experiencias. *Boletín de estudios económicos*, 5, (161), 202 - 227.

Pérez-Gil, J.A. y Moreno, R. (1991). Una validación de la interpretación no causal del análisis factorial. *Currículum, extra 1/2*, 225-229.

Pérez Juste, R. (1994). Investigación evaluativa. En V. García Hoz (Eds.), *Problemas y métodos de investigación en educación personalizada*, 405 - 418. Madrid: Rialp.

Pérez Juste, R. (2000). La evaluación de programas educativos: conceptos básicos, planteamientos generales y problemática. *Revista de Investigación Educativa*. 18, (2), 261-287.

Pérez Juste, R.; García Llamas, J.L. y Martínez Mediano, C. (1995). Evaluación de programas y centros educativos. (Eds.) *Cinzel. Madrid*.

Pérez Juste, R. y Martínez Aragón, L. (1989). Evaluación de centros y calidad educativa. (Eds.) *Cinzel. Madrid*.

Prieto, G. (1982). La universidad a través de sus alumnos. Estudio empírico del sistema docente y del éxito académico. *ICE- Universidad de Salamanca. Salamanca*.

Pung, G. y Stewart, D.W. (1983). Cluster analysis in marketing research: Review and suggestion for application. *Journal of Marketing Research*, 20 de Mayo, 134 - 148.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1, 81 -106.

Bibliografía

Remmers, H.H. (1928). The relationships between student's marks and student attitudes towards instructors. *School and Society*, 28, 759 - 760.

Remmers, H.H. (1931). The equivalence of judgements and test items in the sense of the Spearman-Brown formula. *Journal of Educational Psychology*, 22, 66 - 71.

Remmers, H.H. (1934). Reliability and halo effect on high school and college student's judgement of their teachers. *Journal of Applied Psychology*, 18, 619 - 630.

Remmers, H.H. (1963). Teaching methods in research on teaching. *En N.L. Gage (Eds.), Handbook on Teaching. Chicago: Hand McNally.*

Remmers, H.H.; Martin, F.D. y Elliot, D.N. (1949). Are student ratings of their instructors related to their grades? *Purdue University Studies in Higher Education*, 44, 17 - 26.

Renner, R.R. (1981). Comparing professors: How student ratings contribute to decline in quality higher education. *Phi Delta Kappan*, Octubre, 128 - 130.

Rippey, R.M. (1981). The evaluation of teaching in medical schools. (Eds.) *New York: Springer Publishing Company.*

Rodríguez Espinar, S. (1986). Predicción y explicación del rendimiento relativo en función de los factores de personalidad del H.S.P.Q. *Investigación presentada para el concurso para la Cátedra de la Universidad de Barcelona. Barcelona.*

Rodríguez Espinar, S. (1987). La evaluación del profesorado universitario. Análisis de la Opinión del Profesorado de la Universidad de Barcelona. *Informes de Investigación Evaluativa nº 1. Consideraciones Metodológicas sobre Evaluación y Mejora de la Docencia Universitaria, 211-235. Servei de formacio permanent, Universitat de València. Valencia.*

Rodríguez Espinar, S. (1991). Dimensiones de la calidad universitaria. *I Congreso Internacional sobre calidad de la educación universitaria. Puerto de Santa María (Cádiz), 6-8 de Marzo.*

Rodríguez Espinar, S. (1996). La evaluación institucional universitaria. *Revista de Investigación Educativa, 15, (2), 179-216.*

Rodríguez Osuna, J. (1991). Métodos en el muestreo. Cuadernos Metodológicos. *Centro de Investigaciones Sociológicas (CIS).* Madrid.

Rotem, A. y Glasman, N. (1977). Evaluation of university instructors in the United States. The context. *Higher Education, 6, 75 - 92.*

Rotger, B. (1990). Educación y futuro. Evaluación Formativa. (Eds.) *Cincael.* Madrid.

Roux, M. (1993). Classification des données d'enquêtes. *En Grangé, D. y Lebart, L. (Eds.), Traitement statistiques des enquêtes.* Dunod, Paris.

Salthouse, T.; McKeachie, W. y Lin, Y.G. (1978). An experimental investigation of factors affecting university promotion decisions. *Journal of Higher Education, 49, 177 - 183.*

Salvador, L. y García, A. (1989). Evaluación de la docencia: recogida y tratamiento informático de la opinión de los alumnos. *ICE-Universidad de Cantabria. Cantabria.*

Salvador, L. y Sanz, J. (1987). Evaluación de la docencia mediante cuestionario de alumnos en la ETS de Caminos, Canales y Puertos de la Universidad de Cantabria. *II Congreso Mundial Vasco: Congreso de Educación. Bilbao.*

Salvador, L. y Sanz, J. (1988a). Evaluación de la docencia mediante cuestionario de alumnos: Universidad de Cantabria (curso 86-87). *Studia Paedagogica, 20, 41-71.*

Bibliografía

Salvador, L. y Sanz, J. (1988b). Evaluación de la docencia 1987-1988. Opiniones del Alumno. *Vicerrectorado de Ordenación Académica. Universidad de Cantabria.*

Scheck, D.C. (1978). The use and abuse of student evaluations of teaching effectiveness in higher education. *College Student Journal Monograph, 12, 1 - 13.*

Scriven, J. (1967). The methodology of evaluation. En R. Stake E. (Eds.), *Handbook of Teacher Evaluation. Beverly Hills, C.A: Sage.*

Seldin, P. (1989). How colleges evaluate professors, 1988 vs. 1983. *Ahe Bulletin, 41, (7), 3 - 7.*

Sharma, S. (1996). Applied multivariate techniques. (Eds.) *John Wiley & Sons, Inc. New Cork.*

Smalzried, N.T. y Remmers, H.H. (1943). A factor analysis of the Purdue rating scale for instructors. *American Educational Research Journal, 12, 417 - 433.*

Sobrado Fernández, L.M. (1991). Evaluación de la docencia universitaria. *Revista de las ciencias de la educación, 199, 153-169.*

Sonquist, J.N.; Baker, E.L. y Morgan, J.A. (1971). Searching for structure (Alias-AID-III). (Eds.) *Michigan: Institute for Social Research, University of Michigan.*

Sonquist, J.N. y Morgan, J.A. (1964). The detection of interactions effects. *Monografía, 35. Survey Research Centre, Institute for Social Research, University of Michigan.*

Stevens, J.J. y Aleamoni, L.M. (1985). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science, 13, 285 - 304.*

Tejedor, F.J. (1985). Problemática de la enseñanza universitaria. *Revista de Investigación Educativa, 3, (6), 322 – 337.*

Tejedor, F.J. (1990). La evaluación del profesorado en la Universidad de Santiago. *Revista Española de Pedagogía*, 186, 337-361.

Tejedor, F.J. (1991). Experiencias españolas de evaluación de la enseñanza universitaria y nuevas perspectivas. *III Jornadas Nacionales de Didáctica Universitaria. Evaluación y desarrollo profesional. Universidad de las Palmas. Servicio de Publicaciones. Las Palmas de Gran Canaria.*

Tejedor, F.J.; Jato, E. y Míguez, C. (1988). Evaluación del profesorado universitario por los alumnos de la Universidad de Santiago. *Studia Paedagogica*, 20, 73 -134.

Tejedor, F.J. y García Valcárcer, A. (1996). Evaluación institucional en la Universidad. *Revista Galega de Psicopedagogía*. 6, 101-146.

Tejedor, F.J. y Montero, M.L. (1990). Indicadores de calidad de la docencia para evaluación del profesor universitario. *Revista Española de Pedagogía*, 186, 259 - 279.

Thurstone, L.L. (1935). The measurement of attitudes: A psychological method and some experiments with a scale for measuring attitudes toward the church. (Eds.) *The University of Chicago Press. Chicago.*

Thurstone, L.L. (1947). Multiple factor analysis. A development and expansion of the vectors of the mind. (Eds.) *Chicago: The University of Chicago Press.*

Torrens-Ibern, J. (1972). Modèles et méthodes de l'analyse factorielle. (Eds.) *Paris: Dunod.*

Tourón, J. (1989). La validación del Constructo: su aplicación al CEED (Cuestionario evaluación eficacia docente). *Bordón*, 41, (4), 735-756.

Turker, L. y Lewis, C. (1973). A reliability coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*, 38, 1-10.

Bibliografía

Valera, A. y López, J.A. (2000). Metodología de la evaluación de la enseñanza universitaria. *Psicothema*, 12, (suplemento 2), 553-556.

Vidal, J. (Comisión de Coordinación Técnica). (1999). Indicadores en la Universidad: información y decisiones. Plan nacional de evaluación de la calidad de las universidades. (Eds.) Ministerio de Educación y Cultura. Consejo de Universidades. Secretaría General. Madrid: MEC.

Villar Angulo, L.M. (1983a). Diagnóstico instruccional en la enseñanza superior. ICE-Universidad de Sevilla. Sevilla.

Villar Angulo, L.M. (1983b). Evaluación de la enseñanza universitaria por los estudiantes. *Enseñanza*, 1, 263 - 298.

Villar Angulo, L.M. (1987). Evaluación del ambiente de aprendizaje en el aula universitaria. *Revista de Investigación e Innovación Educativa*, 2, 53 - 65.

Villar Angulo, L.M. (1990). El profesor como profesional: formación y desarrollo personal. *Universidad de Granada*.

Villar Angulo, L.M. (1991). Modelos de desarrollo profesional del profesorado universitario. *III Jornadas Nacionales de Didáctica Universitaria. Evaluación y Desarrollo Profesional. Universidad de las Palmas. Servicio de Publicaciones. Las Palmas de Gran Canaria, 23 - 26 de Septiembre*.

Watkins, D. y Akande, A. (1992). Student evaluations of teaching effectiveness: A Nigerian Investigation. *Higher Education*, 24, (4), 453 - 463.

Wold, S. (1976). Pattern recognition by means of disjoint Principal Components models. *Pattern Recognition*, 8, 127-139.

Wold, S.; Esbensen, K. y Geladi, P. (1987). Principal Component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.

Wolf, R.M. (1990a). Evaluation in education. . (Eds.) *New York: Praeger.*

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In *Bollen, K.A. and Long, J.S. (Eds.), Testing structural equations models, 256-293. Newbury Park, California: Sage.*

Zhang, T.; Ramakrishnan, R. y Livny M. (1996). Birch: an efficient data clustering method for very large databases. *Proceedings of the ACM SIGMOD. Conference of Management Data. Montreal, Canada.*

Bibliografia

Anexos

Anexo 1. Cuestionario de Evaluación Docente de la UPV



UNIVERSITAT POLITÈCNICA DE VALÈNCIA /
UNIVERSIDAD POLITÉCNICA DE VALÈNCIA

ENQUESTA D'OPINIÓ DE L'ALUMNAT
SOBRE L'ACTUACIÓ DOCENT DEL PROFESSORAT /
ENCUESTA DE OPINIÓN DEL ALUMNADO
SOBRE LA ACTUACIÓN DOCENTE DEL PROFESORADO



INSTITUT DE CIÈNCIES
DE L'EDUCACIÓ
INSTITUTO DE CIENCIAS
DE LA EDUCACIÓN

CONTESTA AMB SINCERITAT, SI NO TENS
PROU INFORMACIÓ, NO OPINES /
CONTESTA CON SINCERIDAD, SI NO TIENES
SUFICIENTE INFORMACIÓN, NO OPINES

OMPLI L'ENQUESTA NOMÉS AMB LLAPIS /
RELLENA LA ENCUESTA SÓLO CON LÁPIZ

marqueu així /
marca así

així no marqueu /
así no marques

DADES IDENTIFICATIUS / DATOS IDENTIFICATIVOS

NOM / NOMBRE: _____ CENTRE / CENTRO: _____

TITULACIÓ / TITULACIÓN: _____ CURS / CURSO: _____

ASSIGNATURA / ASIGNATURA: _____ GRUP / GRUPO: _____

CODIS DE L'ALUMNE / CÓDIGOS DEL ALUMNO

SEXE / SEXO: D R BECARI / BECARIO: SI NO EDAT / EDAD: _____

Desenes / Decenas: 1 2 3 4 5
Unitats / Unidades: 0 1 2 3 4 5 6 7 8 9

CONVOCATÒRIA EN L'ASSIGNATURA / CONVOCATORIA EN LA ASIGNATURA: 1 2 3 4 5 6 7 8 9 0
AQUÍ PERCENTATGE DE LES CLASSES D'AQUEST PROFESSOR EN AQUESTA ASSIGNATURA HAS ASSISTIT? / LA QUE PORCENTAJE DE LAS CLASES DE ESTE PROFESOR Y EN ESTA ASIGNATURA HAS ASISTIDO? 0 a 20%
21 a 40%
41 a 60%
61 a 80%
81 a 100%

CURSOS ANTERIORS AMB EL MATEM PROFESSOR / ANTERIORES CURSOS CON EL MISMO PROFESOR: 1 2 3 4 5

ESCALA DE VALORACIÓ / ESCALA DE VALORACIÓN

1	2	3	4	5	6
TOTALMENT EN DESACORD / TOTALMENTE EN DESACUERDO	MÉS AVIAT EN DESACORD / MÁS BIEN EN DESACUERDO	TERME MITJÀ / TÉRMINO MEDIO	MÉS AVIAT D'ACORD / MÁS BIEN DE ACUERDO	TOTALMENT D'ACORD / TOTALMENTE DE ACUERDO	NO DISPONE D'ELEMENTS DE JUDICI PER A OPINAR / NO DISPONGO DE ELEMENTOS DE JUDICIO PARA OPINAR

1	El professor explica a classe amb ordre i claredat.	1 2 3 4 5 6	1	El profesor explica en clase con orden y claridad.	1 2 3 4 5 6
2	Fa la sensació que té les classes ben preparades.	1 2 3 4 5 6	2	Da la sensación de tener bien preparadas las clases.	1 2 3 4 5 6
3	El ritme de les classes permet seguir el fil de les explicacions.	1 2 3 4 5 6	3	El ritmo de las clases permite seguir el hilo de las explicaciones.	1 2 3 4 5 6
4	Sintetiza i subratlla els conceptes que considera importants.	1 2 3 4 5 6	4	Sintetiza y subraya los conceptos que considera importantes.	1 2 3 4 5 6
5	Quan respon a una pregunta ho fa realment al que se li ha preguntat, amb claredat i precisió.	1 2 3 4 5 6	5	Cuando responde a una pregunta, lo hace realmente a lo preguntado con claridad y precisión.	1 2 3 4 5 6
6	Els continguts més importants del programa s'han desenvolupat suficientment en el curs.	1 2 3 4 5 6	6	Los contenidos más importantes del programa se han desarrollado suficientemente en el curso.	1 2 3 4 5 6
7	Dóna a l'assignatura un enfocament aplicat, oferint exemples, demostracions i transferències a la vida real i professional.	1 2 3 4 5 6	7	Da a la asignatura un enfoque aplicado, ofreciendo ejemplos, demostraciones y transferencias a la vida real y profesional.	1 2 3 4 5 6
8	Estableix connexions amb els continguts d'altres assignatures.	1 2 3 4 5 6	8	Establece conexiones con los contenidos de otras asignaturas.	1 2 3 4 5 6
9	Anima els alumnes a plantejar problemes i dubtes a classe.	1 2 3 4 5 6	9	Anima a los alumnos a plantear problemas y dudas en clase.	1 2 3 4 5 6
10	Dialoga amb els alumnes sobre la marxa de les classes, tene en compte les seues opinions.	1 2 3 4 5 6	10	Dialoga con los alumnos sobre la marcha de las clases, tomando en cuenta sus opiniones.	1 2 3 4 5 6
11	Té una actitud receptiva amb els alumnes.	1 2 3 4 5 6	11	Tiene una actitud receptiva hacia los alumnos.	1 2 3 4 5 6
12	Aconseguir que els alumnes estiguen motivats per l'assignatura.	1 2 3 4 5 6	12	Consigue que los alumnos estén motivados por la asignatura.	1 2 3 4 5 6
LES QÜESTIONS 13, 14 I 15 NOMÉS S'HAN DE RESPONDRE POSAT CAS QUE S'HAJA FET ALGUNA AVALUACIÓ DINS D'AQUEST CURS / LAS CUESTIONES 13, 14 Y 15 SÓLO DEBEN RESPONDERSE EN CASO DE HABER REALIZADO ALGUNA EVALUACIÓN DENTRO DE ESTE CURSO					
13	Els exàmens se centren en els aspectes fonamentals de l'assignatura i no en els que s'han tractat poc.	1 2 3 4 5 6	13	Los exámenes se centran en lo fundamental de la asignatura y no en aspectos poco tratados.	1 2 3 4 5 6
14	En general, es justificant.	1 2 3 4 5 6	14	En general, es justo calificando.	1 2 3 4 5 6
15	L'alumne té la possibilitat de comentar i revisar els resultats dels exàmens.	1 2 3 4 5 6	15	El alumno tiene la posibilidad de comentar y revisar los resultados de los exámenes.	1 2 3 4 5 6
16	Els materials emprats i/o recomanats (apuntes, bibliografia, etc.) han sigut una gran ajuda per a preparar l'assignatura.	1 2 3 4 5 6	16	Los materiales utilizados y/o recomendados (apuntes, bibliografía, etc.) han sido una gran ayuda para preparar la asignatura.	1 2 3 4 5 6
17	Utilitza adequadament els mitjans didàctics disponibles (pizarra, transparents, dispositius, vídeos, màquines, etc.) per ajudar a comprendre millor les explicacions.	1 2 3 4 5 6	17	Utiliza adecuadamente los medios didácticos disponibles (pizarra, transparencias, dispositivos, vídeos, máquinas, etc.) para ayudar a comprender mejor las explicaciones.	1 2 3 4 5 6
18	Els continguts teòrics s'apliquen adequadament en exercicis, problemes, treballs, etc.	1 2 3 4 5 6	18	Los contenidos teóricos se aplican adecuadamente en ejercicios, problemas, trabajos, etc.	1 2 3 4 5 6
19	Tenint en compte les limitacions, penso que el professor que imparteix aquesta assignatura s'ha de considerar com un bon professor.	1 2 3 4 5 6	19	Teniendo en cuenta las limitaciones, pienso que el profesor que imparte esta asignatura debe considerarse como un buen profesor.	1 2 3 4 5 6

Enquesta aprovada per la Junta de Govern de la Universitat Politècnica de València, amb data 7 de juliol de 1994 /
 Encuesta aprobada por la Junta de Gobierno de la Universidad Politécnica de Valencia el 7 de julio de 1994

**Gràcies per la vostra col·laboració /
Gracias por tu colaboración**

Anexos

Anexo 2. Posibles tipos de relaciones causales entre variables

Los posibles tipos de relaciones causales que pueden conducir a que dos variables, a las que denominaremos X_1 y X_2 , covaríen son los que se detallan a continuación:

- X_1 y X_2 pueden covariar si una causa a la otra, lo que implicaría asumir un modelo de regresión de una de ellas sobre la otra, como se muestra en los esquemas a y b. En ambos casos la relación es de tipo directo. Estas relaciones directas pueden ser también recíprocas como se muestra en el esquema c.

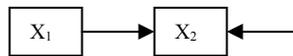


Figura 2.1.a



Figura 2.2.b

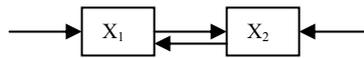


Figura 2.3c

- X_1 y X_2 también covariarán si ambas tienen una causa común X_3 , como se muestra en la figura 2.4.d. En este caso al tipo de relación entre las variables se le denomina espúrea.

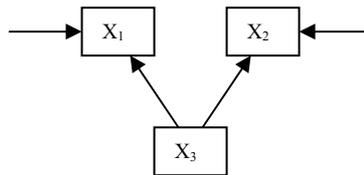


Figura 2.4.d

- X_1 y X_2 también covariarán si ambas están relacionadas a través de una variable intermitente X_3 , como se muestra en la figura 2.5.e. Este tipo de relación es llamada indirecta.

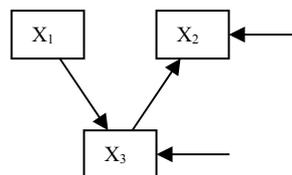


Figura 2.5.e

- Finalmente, puede existir otro tipo de relación que puede emparentarse con la relación indirecta y con la relación espúrea y que se representa en la figura 2.6.f. La diferencia en este tipo de relación es que en este caso, X_1 y X_3 , son ambas exógenas y carecen de mecanismo causal explícito que las relaciones entre sí, dejándose la covariación entre ambas como no explicada. Como no es posible discernir si X_3 contribuye a la covariación entre X_1 y X_2 por vía indirecta o espúrea, a este tipo de efecto se le denomina conjunto.

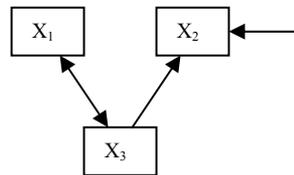


Figura 2.6.f

Anexo 3. Escalado de datos

- **Escalado de datos o normalización.** Es un tratamiento que modifica la unidad de medida de las variables mediante el desplazamiento del origen del sistema de coordenadas o expandiendo o contrayendo los ejes de coordenadas. Los distintos procedimientos de escalado que se han utilizado son:

- **Centrado o normalización con la media.** A cada valor se le resta el valor medio de la columna (o fila) en la que está situado, neutralizando así la influencia de factores ocultos. Con esta transformación se desplaza el origen de coordenadas al centroide de la nube de puntos. Es equivalente a trabajar con la matriz de covarianzas. Es el caso más clásico de normalización. Es equivalente a remplazar las variables originales centrándolas en 0. De esta forma, para describir una muestra se utilizan sólo los valores relativos de las variables, perdiendo la información que pueden aportar los valores absolutos. Esta transformación es recomendable en el caso particular que todas las variables estén medidas en la mismas unidades y se asuma que sus valores sean proporcionales a un factor que no puede ser directamente tenido en cuenta en el análisis. Una propiedad de las muestras normalizadas con la media es que el área bajo la curva es la misma para todas las muestras. La transformación de la media no es adecuada si los valores de la curva no tienen el mismo signo. Ésta fue ideada originalmente para valores positivos sólo, pero puede ser fácilmente aplicada a muestras con todos los valores negativos dividiendo por el valor absoluto de la media en lugar de la media original. De esta forma se mantiene el signo original.

$$x'_{ij} = x_{ij} - \bar{x}_{i(j)} \quad \text{o} \quad x'_{ij} = x_{ij} - \bar{x}_{(i)j}$$

- **Normalización con el máximo.** Es una transformación alternativa a la normalización clásica. Aquí se divide cada fila por el máximo absoluto en lugar de por la media. La eficacia de esta transformación es dudosa si todos los valores de la curva no tienen el mismo signo. Esta transformación convierte el máximo en 1 si todos los datos son positivos, en -1 si todos los datos son negativos.

$$x'_{ij} = x_{ij} - \max_i(x_{ij})$$

- **Estandarización.** Si se realiza por columna, todas las variables están a la misma escala eliminándose el efecto debido a las diversas varianzas de las variables (la desviación típica para todas ellas es 1).

$$x_{ij}' = \frac{x_{ij}}{s_{i(j)}}$$

- **Autoescalado** Este pretratamiento es la combinación del centrado y la estandarización. Equivale al uso de una transformada t de student. Si se realiza este pretratamiento por fila en espectrofotometría NIR (Near Infrared) se conoce como SNV (*Standard normal variable*).

$$x_{ij}' = \frac{x_{ij} - \bar{x}_{i(j)}}{s_{i(j)}}$$

- **Escalado por rango de columna.** Las variables se escalan de modo que el mínimo de cada variable es cero y el máximo 1 (o 100)

$$x_{ij}' = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}$$

- **Escalado por rango de fila.** Esta transformación se divide cada fila por su rango. Así el recorrido de las muestras normalizadas con el rango de fila es 1.

$$x_{ij}' = \frac{x_{ij} - \min_i(x_{ij})}{\max_i(x_{ij}) - \min_i(x_{ij})}$$

Anexo 4. Resultados del AF Exploratorio y Confirmatorio

Tabla 4.1. Matriz de configuración de los componentes (5.3.1.1.1 a)

	Componente	
	1	2
R1	,942	-,137
R2	,872	-7,573E-02
R3	,836	-4,769E-02
R4	,831	8,768E-03
R5	,882	-7,143E-02
R6	,606	,250
R7	,783	-1,697E-02
R8	,788	-4,973E-02
R9	,793	2,561E-02
R10	,735	8,106E-02
R11	,728	,124
R12	,802	7,547E-02
R13	2,941E-02	,839
R14	2,983E-02	,839
R15	2,665E-03	,826
R16	,436	,414
R17	,546	,300
R18	,501	,394
R19	,806	,123

Tabla 4.2. Matriz de configuración de los componentes (5.3.1.1.2a)

	Componente				
	1	2	3	4	5
R1	,929	-5,02E-02	2,41E-02	4,93E-02	-5,66E-02
R2	,860	-3,71E-02	-2,09E-02	,139	-6,16E-02
R3	,758	9,08E-02	4,11E-02	-7,78E-02	9,37E-02
R4	,656	,115	7,23E-02	-1,72E-02	,154
R5	,656	-1,82E-02	,223	5,40E-02	2,25E-02
R6	,458	,314	-,116	6,26E-02	,279
R7	8,74E-02	1,66E-02	2,95E-02	,184	,703
R8	-4,52E-02	1,78E-02	,205	7,65E-02	,767
R9	,112	3,42E-03	,642	8,39E-02	,178
R10	1,87E-02	5,52E-02	,787	3,03E-02	,124
R11	,171	8,92E-02	,734	4,47E-02	-2,25E-02
R12	,311	,109	,424	2,83E-02	,190
R13	,103	,833	-,117	2,05E-02	,112
R14	1,03E-02	,826	,144	-4,48E-02	2,01E-02
R15	-7,55E-02	,630	,205	,291	-,170
R16	2,25E-02	6,31E-02	2,34E-03	,812	6,18E-02
R17	6,25E-02	-4,39E-02	4,32E-02	,822	8,81E-02
R18	,254	,239	-6,46E-03	,431	9,33E-02
R19	,548	,118	,240	,140	1,76E-02

Tabla 4.3. Matriz de configuración de los componentes (5.3.1.1.2 b)

	Componente			
	1	2	3	4
R1	,920	-1,556E-02	8,196E-03	-2,122E-02
R2	,885	-7,988E-02	3,113E-02	1,605E-02
R3	,705	9,700E-02	1,870E-02	2,720E-02
R4	,643	9,382E-02	1,667E-02	,155
R5	,663	,214	-2,309E-02	4,153E-02
R6	,435	-7,705E-02	,356	,155
R7	,134	-7,360E-02	2,732E-02	,833
R8	-8,052E-02	,126	1,171E-02	,865
R9	5,375E-02	,712	1,332E-02	,165
R10	-5,925E-02	,874	5,343E-02	6,242E-02
R11	,137	,819	4,775E-02	-7,691E-02
R12	,269	,455	,154	,109
R16	-8,251E-02	1,204E-02	,944	-6,877E-03
R17	9,223E-05	3,436E-02	,807	5,706E-02
R18	,183	5,450E-02	,668	-2,321E-02
R19	,594	,250	,143	2,268E-02

Tabla 4.4. Matriz de configuración de los componentes (5.3.1.1.2 b)

	Componente			
	1	2	3	4
R1	,920	-1,556E-02	8,196E-03	-2,122E-02
R2	,885	-7,988E-02	3,113E-02	1,605E-02
R3	,705	9,700E-02	1,870E-02	2,720E-02
R4	,643	9,382E-02	1,667E-02	,155
R5	,663	,214	-2,309E-02	4,153E-02
R6	,435	-7,705E-02	,356	,155
R7	,134	-7,360E-02	2,732E-02	,833
R8	-8,052E-02	,126	1,171E-02	,865
R9	5,375E-02	,712	1,332E-02	,165
R10	-5,925E-02	,874	5,343E-02	6,242E-02
R11	,137	,819	4,775E-02	-7,691E-02
R12	,269	,455	,154	,109
R16	-8,251E-02	1,204E-02	,944	-6,877E-03
R17	9,223E-05	3,436E-02	,807	5,706E-02
R18	,183	5,450E-02	,668	-2,321E-02
R19	,594	,250	,143	2,268E-02

Tabla 4.5. Matriz de configuración de los componentes (5.3.1.3)

	Bruta					Reescalada				
	Componente					Componente				
	1	2	3	4	5	1	2	3	4	5
R1	1,136	-,050	-,020	-,069	,040	,943	-,041	-,016	-,057	,034
R2	,961	-,046	,020	-,064	,169	,849	-,040	,017	-,057	,150
R3	,915	,134	-,047	,106	-,125	,770	,113	-,039	,089	-,105
R4	,731	,133	-,094	,163	,000	,642	,117	-,083	,143	,000
R5	,747	-,031	-,254	,017	,088	,647	-,027	-,220	,015	,076
R6	,521	,349	,102	,310	,104	,450	,301	,088	,268	,090
R7	,101	,039	-,066	,784	,238	,087	,034	-,057	,675	,205
R8	-,048	,059	-,264	,838	,092	-,041	,051	-,230	,729	,080
R9	,124	-,013	-, 776	,201	,103	,105	-,011	-,655	,170	,087
R10	-,004	,043	-, 1,071	,137	,013	-,004	,034	-,845	,108	,010
R11	,197	,068	-, 939	-,052	,054	,161	,056	-,768	-,042	,044
R12	,404	,158	-, 570	,246	-,026	,317	,124	-,448	,193	-,021
R13	,127	1,003	,135	,134	,058	,104	,824	,111	,110	,048
R14	,010	1,089	-,194	,002	-,060	,008	,851	-,152	,001	-,047
R15	-,092	,601	-,266	-,255	,515	-,078	,507	-,225	-,215	,435
R16	,080	,057	-,018	,130	,851	,071	,050	-,016	,114	,749
R17	,128	-,047	-,064	,152	,811	,116	-,043	-,058	,138	,736
R18	,310	,245	-,015	,132	,478	,272	,215	-,013	,116	,419
R19	,672	,132	-,300	,019	,158	,557	,110	-,249	,016	,131

Tabla 4.6. Matriz configuración (5.3.1.3)

	Bruta				Reescalada			
	Componente				Componente			
	1	2	3	4	1	2	3	4
R1	1,034	,026	,004	-,024	,932	,024	,004	-,022
R2	,892	,074	,043	,026	,856	,071	,042	,025
R3	,847	-,092	-,003	,016	,752	-,082	-,003	,014
R4	,662	-,096	,033	,159	,631	-,092	,031	,152
R5	,690	-,205	-,010	,059	,653	-,194	-,010	,056
R6	,442	,064	,405	,146	,412	,060	,378	,137
R7	,141	,086	,024	,934	,129	,078	,022	,851
R8	-,084	-,142	,022	,910	-,078	-,132	,020	,846
R9	,066	-,779	,012	,196	,059	-,697	,010	,176
R10	-,082	-,1,073	,056	,058	-,070	-,905	,048	,049
R11	,169	-,898	,048	-,083	,151	-,804	,043	-,075
R12	,332	-,546	,185	,117	,279	-,458	,155	,098
R16	-,102	-,010	1,045	-,007	-,093	-,009	,957	-,007
R17	,018	-,036	,794	,077	,017	-,035	,766	,075
R18	,181	-,056	,767	-,036	,165	-,051	,700	-,033
R19	,658	-,262	,160	,027	,600	-,238	,146	,024

Tabla 4.7. Matriz de Correlación entre Componentes

Componentes	1	2	3	4
1	1,000	-,647	,690	,605
2	-,647	1,000	-,591	-,598
3	,690	-,591	1,000	,585
4	,605	-,598	,585	1,000

Tabla 4.8. Matriz configuración datos 00/01

	Componente			
	1	2	3	4
MEDIA R1	1,001	-,079	,005	-,019
MEDIA R2	,901	-,206	,109	-,133
MEDIA R3	,762	,238	-,059	,012
MEDIA R4	,790	,217	-,073	-,058
MEDIA R5	,686	,151	,055	-,128
MEDIA R6	,714	,015	,215	-,002
MEDIA R7	,100	,027	,174	-,756
MEDIA R8	,140	,240	,026	-,639
MEDIA R9	,088	,712	,065	-,190
MEDIA R10	,014	,835	,122	-,078
MEDIA R11	,091	,798	,102	-,043
MEDIA R12	,451	,355	,073	-,211
MEDIA R16	,062	,079	,807	-,052
MEDIA R17	-,012	,042	,846	-,133
MEDIA R18	,483	,175	,446	,210
MEDIA R19	,681	,193	,095	-,110

Tabla 4.9. Matriz configuración datos 99/00

	Componente			
	1	2	3	4
MEDIA R1	,994	-,063	,007	-,020
MEDIA R2	,923	-,177	,115	-,064
MEDIA R3	,633	,311	-,069	-,072
MEDIA R4	,642	,253	,071	-,076
MEDIA R5	,699	,204	,025	-,096
MEDIA R6	,537	,044	,391	-,001
MEDIA R7	,118	,039	,070	-,810
MEDIA R8	,078	,181	,100	-,689
MEDIA R9	,116	,727	,085	-,127

MEDIA R10	-,067	,837	,130	-,145
MEDIA R11	,157	,751	,078	-,062
MEDIA R12	,445	,405	,065	-,175
MEDIA R16	,004	,034	,844	-,107
MEDIA R17	,013	-,023	,777	-,241
MEDIA R18	,224	,242	,672	,240
MEDIA R19	,716	,200	,086	-,076

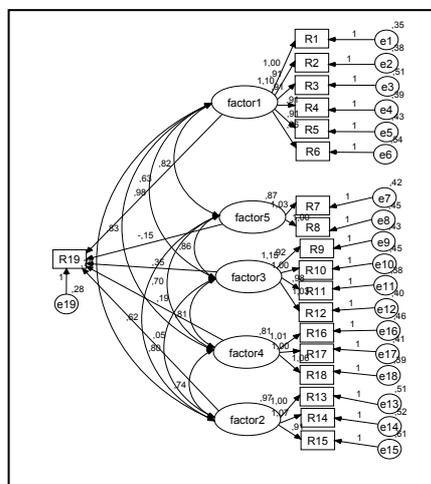


Figura 4.1. Solución no estandarizada del modelo 5

Tabla 4.10. Coeficientes de regresión del modelo 5

Coef. Regresión	Estimado	S.E.
R14 ← factor2	1,073	0,008
R15 ← factor2	0,905	0,008
R13 ← factor2	1,0	
R6 ← factor1	0,857	0,006
R5 ← factor1	0,909	0,006
R4 ← factor1	0,908	0,006
R3 ← factor1	0,907	0,006
R2 ← factor1	0,907	0,006
R1 ← factor1	1,0	
R9 ← factor3	0,916	0,006
R10 ← factor3	1,0	
R11 ← factor3	0,984	0,006
R12 ← factor3	1,027	0,006
R7 ← factor5	1,029	0,008

R8←-factor5	1,0	
R16←-factor4	1,013	0,008
R18←-factor4	1,060	0,007
R17←-factor4	1,0	
R19←-factor5	-0,148	0,015
R19←-factor3	0,349	0,013
R19←-factor4	0,194	0,019
R19←-factor2	0,053	0,011
R19←-factor1	0,631	0,013

Tabla 4.11. Covarianzas estimadas para el modelo 5

COVARIANZA	ESTIMADO	S.E.
factor 2 ↔ factor 1	0,763	0,010
factor 1 ↔ factor 5	0,819	0,010
factor 1 ↔ factor 3	0,977	0,012
factor 3 ↔ factor 5	0,863	0,011
factor 1 ↔ factor 4	0,825	0,010
factor 3 ↔ factor 4	0,812	0,010
factor 5 ↔ factor 4	0,699	0,009
factor 2 ↔ factor 4	0,739	0,010
factor 2 ↔ factor 3	0,803	0,011
factor 2 ↔ factor 5	0,622	0,009

Tabla 4.12. Varianzas estimadas para el modelo 5

VARIANZA	ESTIMADA	S.E.
Factor2	0,971	0,014
Factor1	1,096	0,014
Factor3	1,153	0,015
Factor5	0,874	0,013
Factor4	0,807	0,011
e1	0,355	0,004
e2	0,379	0,004
e3	0,510	0,005
e4	0,392	0,004
e5	0,427	0,005
e6	0,538	0,006
e7	0,423	0,006
e8	0,451	0,006
e9	0,435	0,005
e10	0,454	0,005
e11	0,378	0,005
e12	0,403	0,005
e16	0,464	0,005
e17	0,406	0,005
e18	0,391	0,005
e13	0,511	0,007

e14	0,523	0,007
e15	0,606	0,007
e19	0,277	0,003

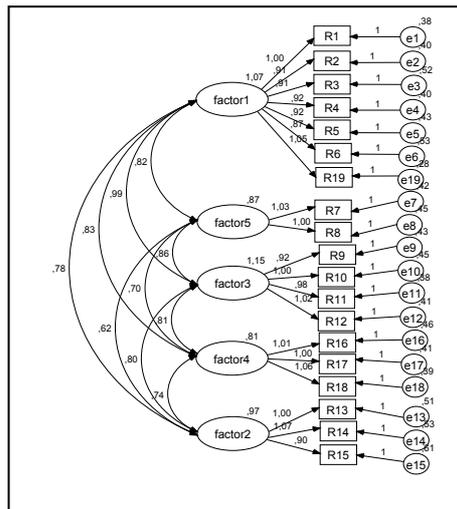


Figura 4.2. Solución no estandarizada para el modelo 4

Tabla 4.13. Coeficientes de regresión del modelo 4

Coef. Regresión	Estimado	S.E.
R14 ← factor2	1,070	0,008
R15 ← factor2	0,904	0,008
R13 ← factor2	1,0	
R6 ← factor1	0,870	0,006
R5 ← factor1	0,918	0,006
R4 ← factor1	0,916	0,006
R3 ← factor1	0,912	0,006
R2 ← factor1	0,908	0,006
R1 ← factor1	1,0	
R9 ← factor3	0,918	0,006
R10 ← factor3	1,0	
R11 ← factor3	0,982	0,006
R12 ← factor3	1,025	0,006
R7 ← factor5	1,030	0,008
R8 ← factor5	1,0	
R19 ← factor1	1,048	0,006
R16 ← factor4	1,014	0,008
R18 ← factor4	1,060	0,007
R17 ← factor4	1,000	

Tabla 4.14. Varianzas obtenidas en el modelo 4

COVARIANZA	ESTIMADO	S.E.
factor 2 ↔ factor 1	0,778	0,010
factor 1 ↔ factor 5	0,816	0,010
factor 1 ↔ factor 3	0,991	0,012
factor 3 ↔ factor 5	0,863	0,011
factor 1 ↔ factor 4	0,830	0,010
factor 3 ↔ factor 4	0,812	0,010
factor 5 ↔ factor 4	0,698	0,009
factor 2 ↔ factor 4	0,740	0,010
factor 2 ↔ factor 3	0,805	0,011
factor 2 ↔ factor 5	0,623	0,009

Tabla 4.15. Varianzas obtenidas en el modelo 4

VARIANZA	ESTIMADA	S.E.
Factor2	0,973	0,014
Factor1	1,071	0,014
Factor3	1,155	0,015
Factor5	0,872	0,013
Factor4	0,806	0,011
e1	0,380	0,004
e2	0,399	0,004
e3	0,521	0,005
e4	0,399	0,004
e5	0,432	0,005
e6	0,533	0,006
e7	0,422	0,006
e8	0,452	0,006
e9	0,430	0,005
e10	0,452	0,005
e11	0,380	0,005
e12	0,407	0,005
e16	0,463	0,005
e17	0,407	0,005
e18	0,392	0,005
e13	0,508	0,007
e14	0,526	0,007
e15	0,606	0,007
e19	0,277	0,003

Anexo 5. Resultados del análisis clúster y de la caracterización de los mismos

Tabla 5.1. Tabla de contingencia para las variables titulación y clúster

TITULACIÓN	Número de conglomerados en dos fases			Total
	1	2	3	
EPS ALCOY	4	21	3	28
	,7%	2,2%	,5%	1,3%
EPS ALCOY-ADE	5	7	2	14
	,9%	,7%	,3%	,7%
EPS ALCOY-IT DISEÑ	16	9	2	27
	2,9%	,9%	,3%	1,3%
EPS ALCOY-IT TELECO	9	9	8	26
	1,6%	,9%	1,3%	1,2%
EPS ALCOY-ITI E.IN	1	7	9	17
	,2%	,7%	1,5%	,8%
EPS ALCOY-ITI ELEC	2	6	6	14
	,4%	,6%	1,0%	,7%
EPS ALCOY-ITI MEC.	3	5	9	17
	,5%	,5%	1,5%	,8%
EPS ALCOY-ITI QUIM	1	2	5	8
	,2%	,2%	,8%	,4%
EPS ALCOY-ITI TEXT	1	2	4	7
	,2%	,2%	,7%	,3%
EPS ALCOY-ORG.IND.	6	7	5	18
	1,1%	,7%	,8%	,8%
EPS GANDIA-CC.AA.	3	6	5	14
	,5%	,6%	,8%	,7%
EPS GANDIA-ELECTR.	4	8	9	21
	,7%	,8%	1,5%	1,0%
EPS GANDIA-FOREST.	4	16	11	31
	,7%	1,7%	1,8%	1,5%
EPS GANDIA-SONIDO	2	12	5	19
	,4%	1,2%	,8%	,9%
EPS GANDIA-TELECO.	4	15	4	23
	,7%	1,6%	,7%	1,1%
EPS GANDIA-TURISMO	2	16	4	22
	,4%	1,7%	,7%	1,0%
ETSI GEODESICA C.T	9	11	4	24
	1,6%	1,1%	,7%	1,1%
ETSI TELECO	54	73	48	175
	9,6%	7,6%	7,9%	8,2%

ETSIA-ING. AGRONO.	108	168	90	366
	19,3%	17,4%	14,8%	17,1%
ETSIA-TECNO. ALIM.	11	25	8	44
	2,0%	2,6%	1,3%	2,1%
ETSICCP-CCAA	12	10	2	24
	2,1%	1,0%	,3%	1,1%
ETSICCP-CCP	3	4	0	7
	,5%	,4%	,0%	,3%
ETSICCP-CCP PLAN97	4	6	7	17
	,7%	,6%	1,1%	,8%
ETSICCP-OOPP	3	3	1	7
	,5%	,3%	,2%	,3%
ETSICCP-OOPP P97	12	14	5	31
	2,1%	1,5%	,8%	1,5%
ETSICCP-PIE (IC)	28	36	13	77
	5,0%	3,7%	2,1%	3,6%
ETSIGCT-TOPOGRAFIA	11	39	42	92
	2,0%	4,0%	6,9%	4,3%
ETSII-ADMON Y D.E.	1	10	5	16
	,2%	1,0%	,8%	,7%
ETSII-AUTOMATICA	0	12	9	21
	,0%	1,2%	1,5%	1,0%
ETSII-MATERIALES	1	1	2	4
	,2%	,1%	,3%	,2%
ETSII-ORG. INDUSTRI	12	21	12	45
	2,1%	2,2%	2,0%	2,1%
ETSII-PLAN 93	69	120	96	285
	12,3%	12,4%	15,7%	13,3%
ETSII-QUIMICO	12	16	19	47
	2,1%	1,7%	3,1%	2,2%
EU ARQUIT. TECNICA	8	14	5	27
	1,4%	1,5%	,8%	1,3%
EU INFORMATICA	29	54	32	115
	5,2%	5,6%	5,2%	5,4%
EUIT INDUSTRIAL	40	68	40	148
	7,1%	7,1%	6,6%	6,9%
EUITA-EXPL. AGROP.	6	18	13	37
	1,1%	1,9%	2,1%	1,7%
EUITA-HORTO Y JARD	11	26	15	52
	2,0%	2,7%	2,5%	2,4%
EUITA-IND. AGR.Y A	10	20	14	44
	1,8%	2,1%	2,3%	2,1%

EUITA-MECAN. Y C.R	0	3	2	5
	,0%	,3%	,3%	,2%
F. INFORMATICA	34	42	28	104
	6,1%	4,4%	4,6%	4,9%
FI-DOCUMENTACION	6	2	7	15
	1,1%	,2%	1,1%	,7%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.2. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
214.88	164	0,0047

Tabla 5.3. Tabla de contingencia para las variables departamento y clúster

DEPTO	Número de conglomerados en dos fases			Total
	1	2	3	
1	7	14	6	27
	1,2%	1,5%	1,0%	1,3%
2	17	26	17	60
	3,0%	2,7%	2,8%	2,8%
3	12	20	14	46
	2,1%	2,1%	2,3%	2,2%
5	0	1	2	3
	,0%	,1%	,3%	,1%
6	0	1	1	2
	,0%	,1%	,2%	,1%
7	14	46	18	78
	2,5%	4,8%	3,0%	3,7%
9	17	10	16	43
	3,0%	1,0%	2,6%	2,0%
10	3	5	4	12
	,5%	,5%	,7%	,6%
11	33	43	19	95
	5,9%	4,5%	3,1%	4,4%
12	24	40	24	88
	4,3%	4,1%	3,9%	4,1%
13	17	42	22	81
	3,0%	4,4%	3,6%	3,8%

14	9	32	18	59
	1,6%	3,3%	3,0%	2,8%
15	18	42	37	97
	3,2%	4,4%	6,1%	4,5%
16	23	28	16	67
	4,1%	2,9%	2,6%	3,1%
17	22	33	19	74
	3,9%	3,4%	3,1%	3,5%
18	9	9	4	22
	1,6%	,9%	,7%	1,0%
19	8	20	14	42
	1,4%	2,1%	2,3%	2,0%
20	26	48	25	99
	4,6%	5,0%	4,1%	4,6%
21	10	16	16	42
	1,8%	1,7%	2,6%	2,0%
22	18	45	27	90
	3,2%	4,7%	4,4%	4,2%
23	11	20	12	43
	2,0%	2,1%	2,0%	2,0%
24	4	11	7	22
	,7%	1,1%	1,1%	1,0%
25	4	5	8	17
	,7%	,5%	1,3%	,8%
26	48	80	48	176
	8,6%	8,3%	7,9%	8,2%
27	3	8	6	17
	,5%	,8%	1,0%	,8%
28	28	42	27	97
	5,0%	4,4%	4,4%	4,5%
30	28	40	22	90
	5,0%	4,1%	3,6%	4,2%
31	20	17	9	46
	3,6%	1,8%	1,5%	2,2%
32	40	65	43	148
	7,1%	6,7%	7,0%	6,9%

33	25	44	21	90
	4,5%	4,6%	3,4%	4,2%
34	4	7	3	14
	,7%	,7%	,5%	,7%
35	3	3	5	11
	,5%	,3%	,8%	,5%
37	0	0	1	1
	,0%	,0%	,2%	,0%
38	4	18	9	31
	,7%	1,9%	1,5%	1,5%
39	32	48	43	123
	5,7%	5,0%	7,0%	5,8%
40	5	9	2	16
	,9%	,9%	,3%	,7%
41	10	4	6	20
	1,8%	,4%	1,0%	,9%
42	5	22	19	46
	,9%	2,3%	3,1%	2,2%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.4. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
1025.62	152	0,0000

Tabla 5.5. Tabla de contingencia para las variables edad y clúster

Edad	Número de conglomerados en dos fases			Total
	1	2	3	
27	0	1	2	3
	,0%	,1%	,3%	,1%
28	2	0	2	4
	,4%	,0%	,3%	,2%
29	3	10	6	19
	,5%	1,0%	1,0%	,9%
30	6	14	7	27
	1,1%	1,5%	1,2%	1,3%
31	16	31	23	70

	2,9%	3,2%	3,8%	3,3%
32	21	45	33	99
	3,8%	4,7%	5,5%	4,7%
33	35	55	30	120
	6,3%	5,8%	5,0%	5,7%
34	19	54	34	107
	3,4%	5,7%	5,6%	5,1%
35	26	54	34	114
	4,7%	5,7%	5,6%	5,4%
36	28	52	41	121
	5,0%	5,5%	6,8%	5,7%
37	18	38	25	81
	3,2%	4,0%	4,1%	3,8%
38	26	37	24	87
	4,7%	3,9%	4,0%	4,1%
39	22	27	12	61
	3,9%	2,8%	2,0%	2,9%
40	14	28	24	66
	2,5%	2,9%	4,0%	3,1%
41	29	23	23	75
	5,2%	2,4%	3,8%	3,5%
42	19	34	23	76
	3,4%	3,6%	3,8%	3,6%
43	14	26	18	58
	2,5%	2,7%	3,0%	2,7%
44	16	36	17	69
	2,9%	3,8%	2,8%	3,3%
45	19	24	22	65
	3,4%	2,5%	3,6%	3,1%
46	25	27	18	70
	4,5%	2,8%	3,0%	3,3%
47	21	15	13	49
	3,8%	1,6%	2,1%	2,3%
48	16	29	11	56
	2,9%	3,0%	1,8%	2,6%
49	14	30	14	58

	2,5%	3,1%	2,3%	2,7%
50	11	24	8	43
	2,0%	2,5%	1,3%	2,0%
51	13	18	10	41
	2,3%	1,9%	1,7%	1,9%
52	13	27	14	54
	2,3%	2,8%	2,3%	2,6%
53	12	23	16	51
	2,2%	2,4%	2,6%	2,4%
54	12	23	13	48
	2,2%	2,4%	2,1%	2,3%
55	14	25	11	50
	2,5%	2,6%	1,8%	2,4%
56	1	8	6	15
	,2%	,8%	1,0%	,7%
57	5	20	12	37
	,9%	2,1%	2,0%	1,7%
58	6	12	5	23
	1,1%	1,3%	,8%	1,1%
59	2	16	5	23
	,4%	1,7%	,8%	1,1%
60	9	4	4	17
	1,6%	,4%	,7%	,8%
61	9	11	6	26
	1,6%	1,2%	1,0%	1,2%
62	5	10	6	21
	,9%	1,0%	1,0%	1,0%
63	6	17	16	39
	1,1%	1,8%	2,6%	1,8%
64	1	7	6	14
	,2%	,7%	1,0%	,7%
65	6	2	1	9
	1,1%	,2%	,2%	,4%
66	8	1	0	9
	1,4%	,1%	,0%	,4%
67	0	4	1	5

	,0%	,4%	,2%	,2%
68	1	2	0	3
	,2%	,2%	,0%	,1%
69	8	0	3	11
	1,4%	,0%	,5%	,5%
70	3	4	3	10
	,5%	,4%	,5%	,5%
71	1	1	0	2
	,2%	,1%	,0%	,1%
73	0	1	1	2
	,0%	,1%	,2%	,1%
74	0	3	0	3
	,0%	,3%	,0%	,1%
76	1	0	0	1
	,2%	,0%	,0%	,0%
77	0	1	0	1
	,0%	,1%	,0%	,0%
83	0	0	1	1
	,0%	,0%	,2%	,0%
85	2	0	1	3
	,4%	,0%	,2%	,1%
Total	558	954	605	2117
	100,0%	100,0%	100,0%	100,0%

Tabla 5.6. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
212,91	208	0,3929

Tabla 5.7. Tabla de contingencia para las variables categoría docente y clúster

Categoría	Número de conglomerados en dos fases			Total
	1	2	3	
ASOL-P1	3	10	5	18
	,5%	1,0%	,8%	,8%
ASOL-P2	0	2	0	2
	,0%	,2%	,0%	,1%
ASO-P3	6	11	8	25

	1,1%	1,1%	1,3%	1,2%
ASO-P4	7	2	3	12
	1,2%	,2%	,5%	,6%
ASO-P5	1	2	0	3
	,2%	,2%	,0%	,1%
ASO-P6	30	51	43	124
	5,3%	5,3%	7,0%	5,8%
ASO-TC	16	21	17	54
	2,9%	2,2%	2,8%	2,5%
AYEU	2	4	0	6
	,4%	,4%	,0%	,3%
CEU	13	63	34	110
	2,3%	6,5%	5,6%	5,2%
CU	116	153	90	359
	20,7%	15,9%	14,8%	16,8%
CU-P3	1	0	1	2
	,2%	,0%	,2%	,1%
EME-P3	0	0	1	1
	,0%	,0%	,2%	,0%
L1447-P03	0	1	0	1
	,0%	,1%	,0%	,0%
L1447-P06	1	4	0	5
	,2%	,4%	,0%	,2%
L1447-TC	3	1	1	5
	,5%	,1%	,2%	,2%
MTL	0	1	0	1
	,0%	,1%	,0%	,0%
TEU	222	396	250	868
	39,6%	41,1%	41,0%	40,7%
TEUI	3	10	3	16
	,5%	1,0%	,5%	,7%
TEU-P6	6	3	0	9
	1,1%	,3%	,0%	,4%
TU	131	228	154	513
	23,4%	23,7%	25,2%	24,0%
TU-P6	0	1	0	1

	,0%	,1%	,0%	,0%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.8. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
74,98	80	0,6470

Tabla 5.9. Tabla de contingencia para las variables doctor y clúster

Doctor	Número de conglomerados en dos fases			Total
	1	2	3	
	3	10	5	18
NS/NC	,5%	1,0%	,8%	,8%
N	216	351	217	784
	38,5%	36,4%	35,6%	36,7%
S	342	603	388	1333
	61,0%	62,6%	63,6%	62,4%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.10. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
3,69	8	0,8841

Tabla 5.11. Tabla de contingencia para las variables tiempo de dedicación y clúster

Tiempo Dedicación	Número de conglomerados en dos fases			Total
	1	2	3	
	3	10	5	18
NS/NC	,5%	1,0%	,8%	,8%
TC	506	877	549	1932
	90,2%	91,0%	90,0%	90,5%
TP	52	77	56	185
	9,3%	8,0%	9,2%	8,7%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.12. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
4,58	8	0,8010

Tabla 5.13. Tabla de contingencia para las variables curso y clúster

Curso	Número de conglomerados en dos fases			Total
	1	2	3	
0	27	53	31	111
	4,8%	5,5%	5,1%	5,2%
1	145	254	148	547
	25,8%	26,3%	24,3%	25,6%
2	128	236	146	510
	22,8%	24,5%	23,9%	23,9%
3	128	207	157	492
	22,8%	21,5%	25,7%	23,0%
4	60	94	60	214
	10,7%	9,8%	9,8%	10,0%
5	65	100	63	228
	11,6%	10,4%	10,3%	10,7%
6	8	20	5	33
	1,4%	2,1%	,8%	1,5%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.14. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
18,3	24	0,7881

Tabla 5.15. Tabla de contingencia para las variables tipo de titulación y clúster

Tipo Titulación	Número de conglomerados en dos fases			Total
	1	2	3	
CICLO2	57	89	49	195
	10,2%	9,2%	8,0%	9,1%
SUPERIOR	329	523	320	1172
	58,6%	54,3%	52,5%	54,9%
TÉCNICA	175	352	241	768
	31,2%	36,5%	39,5%	36,0%
Total	561	964	610	2135

Tabla 5.16. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
10,8	8	0,2131

Tabla 5.17. Tabla de contingencia para las variables semestre y clúster

Semestre	Número de conglomerados en dos fases			Total
	1	2	3	
1	468	836	528	1832
	83,4%	86,7%	86,6%	85,8%
2	93	128	82	303
	16,6%	13,3%	13,4%	14,2%
Total	561	964	610	2135
	100,0%	100,0%	100,0%	100,0%

Tabla 5.18. Tabla de significación estadística test chi-cuadrado

CHI-SQUARE	DF	P-VALUE
9,45	4	0,0508

Anexo 6. Resultados análisis discriminante efectuado en cada nodo, poder discriminante y distancia entre modelos

Tabla 6.1. Autovalores del nodo 1 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalores	% de varianza	% acumulado	Correlación canónica
1	6,395	99,5	99,5	,930
2	,028	,4	100,0	,165
3	,002	,0	100,0	,045

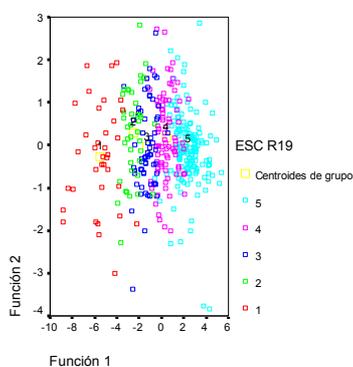


Figura 6.1. Funciones canónicas discriminantes del nodo 1 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.2. Autovalores del nodo 2 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,332(a)	98,3	98,3	,918
2	,078(a)	1,4	99,7	,270
3	,010(a)	,2	99,9	,099
4	,006(a)	,1	100,0	,076

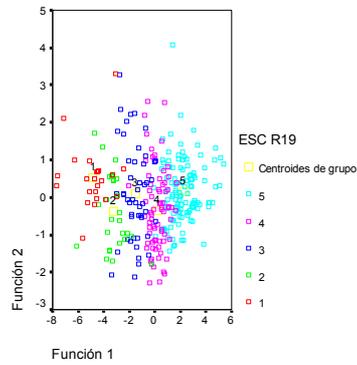


Figura 6.2. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.3. Autovalores del nodo 3 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,209(a)	95,8	95,8	,899
2	,187(a)	4,2	100,0	,397

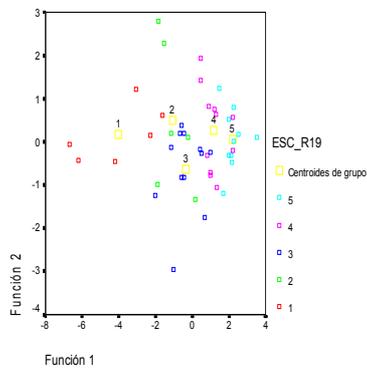


Figura 6.3. Funciones canónicas discriminantes del nodo 3 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.4. Autovalores del nodo 4 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,215(a)	99,8	99,8	,899
2	,008(a)	,2	100,0	,089

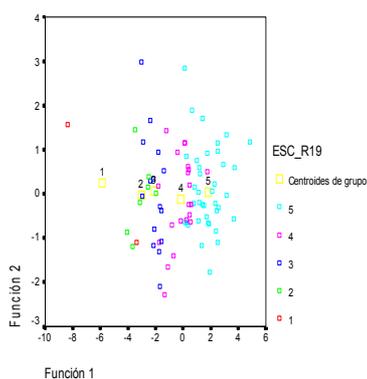


Figura 6.4. Funciones canónicas discriminantes del nodo 4 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.5. Autovalores del nodo 5 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	7,564(a)	96,5	96,5	,940
2	,181(a)	2,3	98,9	,392
3	,072(a)	,9	99,8	,260
4	,018(a)	,2	100,0	,131

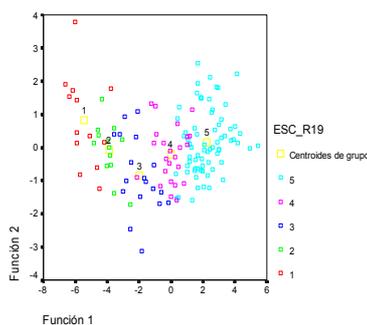


Figura 6.5. Funciones canónicas discriminantes del nodo 5 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.6. Autovalores del nodo 6 siendo la variable de ramificación departamento en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	12,736(a)	98,2	98,2	,963
2	,228(a)	1,8	100,0	,431

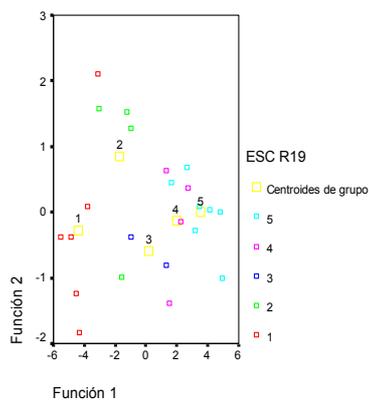


Figura 6.6. Funciones canónicas discriminantes del nodo 6 siendo la variable de ramificación departamento en las carreras superiores

Tabla 6.7. Autovalores del nodo 1 siendo la variable de ramificación titulación en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	3,490(a)	96,9	96,9	,882
2	,065(a)	1,8	98,7	,247
3	,046(a)	1,3	100,0	,209

Tabla 6.8. Matriz de Estructuras del nodo 1 siendo la variable de ramificación titulación en las carreras superiores

	Función		
	1	2	3
MEDIA R1	,764(*)	-,637	,102
MEDIA R12	,687(*)	,465	-,559
MEDIA R11	,601	,707(*)	,373

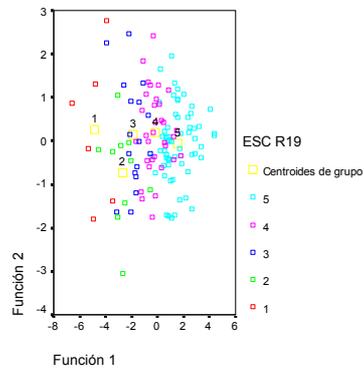


Figura 6.7. Funciones canónicas discriminantes del nodo 1 siendo la variable de ramificación titulación en las carreras superiores

Tabla 6.9. Autovalores del nodo 2 siendo la variable de ramificación titulación en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,262(a)	95,9	95,9	,929
2	,267(a)	4,1	100,0	,459

Tabla 6.10. Matriz de estructura del nodo 2 siendo la variable de ramificación titulación en las carreras superiores

	Función	
	1	2
MEDIA R1	,824(*)	-,567
MEDIA R11	,463	,887(*)

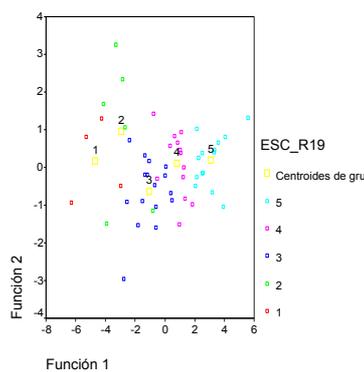


Figura 6.8. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación titulación en las carreras superiores

Tabla 6.11. Autovalores del nodo 3 siendo la variable de ramificación titulación en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,581(a)	99,4	99,4	,932
2	,025(a)	,4	99,8	,157
3	,007(a)	,1	99,9	,085
4	,005(a)	,1	100,0	,069

Tabla 6.12. Matriz de estructuras del nodo 3 siendo la variable de ramificación titulación en las carreras superiores

	Función			
	1	2	3	4
MEDIA R1	,742(*)	-,203	-,054	,336
MEDIA R5	,688(*)	-,330	,329	-,162
MEDIA R12	,681(*)	,303	-,281	,228
MEDIA R4	,652(*)	,238	,612	,091
MEDIA R2	,625(*)	,192	-,032	-,473
MEDIA R16	,359	,618(*)	,278	,261

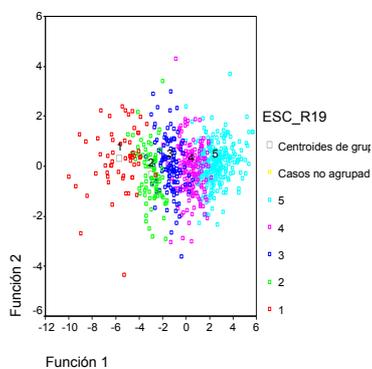


Figura 6.9. Funciones canónicas discriminantes del nodo 3 siendo la variable de ramificación titulación en las carreras superiores

Tabla 6.13. Autovalores del nodo 4 siendo la variable de ramificación titulación en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,933(a)	98,5	98,5	,935
2	,075(a)	1,1	99,5	,264

3	,027(a)	,4	99,9	,163
4	,005(a)	,1	100,0	,071

Tabla 6.14. Matriz de estructuras del nodo 4 siendo la variable de ramificación titulación en las carreras superiores

	Función			
	1	2	3	4
MEDIA R1	,751(*)	-,454	,052	-,476
MEDIA R5	,714(*)	-,223	-,053	,662
MEDIA R12	,687(*)	,606	-,398	-,036
MEDIA R2	,565	-,152	,750(*)	-,310

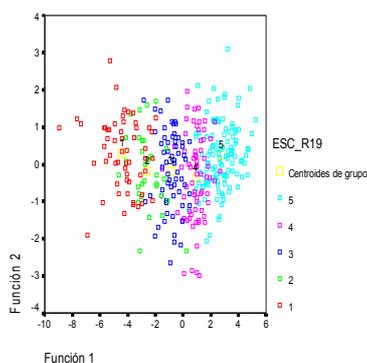


Figura 6.10. Funciones canónicas discriminantes del nodo 4 siendo la variable de ramificación titulación en las carreras superiores

Tabla 6.15. Autovalores del nodo 1 siendo la variable de ramificación curso en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,779(a)	99,5	99,5	,923
2	,027(a)	,5	99,9	,161
3	,004(a)	,1	100,0	,066
4	,001(a)	,0	100,0	,030

Tabla 6.16. Matriz de estructuras del nodo 1 siendo la variable de ramificación curso en las carreras superiores

	Función			
	1	2	3	4
MEDIA R1	,747(*)	-,342	-,459	-,325
MEDIA R5	,717(*)	-,273	,614	,000

MEDIA R12	,666(*)	,566	-,020	,317
MEDIA R2	,609(*)	,041	-,090	-,567
MEDIA R11	,531	,054	,192	,804(*)

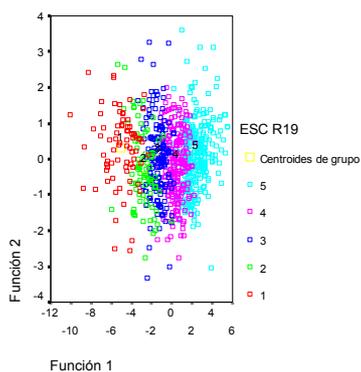


Figura 6.11. Funciones canónicas discriminantes del nodo 1 siendo la variable de ramificación curso en las carreras superiores

Tabla 6.17. Autovalores del nodo 2 siendo la variable de ramificación curso en las carreras superiores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	7,667(a)	99,0	99,0	,941
2	,054(a)	,7	99,7	,226
3	,023(a)	,3	100,0	,149
4	,003(a)	,0	100,0	,054

Tabla 6.18. Matriz de estructuras del nodo 2 siendo la variable de ramificación curso en las carreras superiores

	Función			
	1	2	3	4
MEDIA R1	,730(*)	-,049	,323	-,527
MEDIA R12	,700(*)	-,341	-,385	,197
MEDIA R2	,602(*)	,413	-,195	-,148
MEDIA R7	,403	,647(*)	-,069	,052
MEDIA R11	,495	-,044	,219	,819(*)

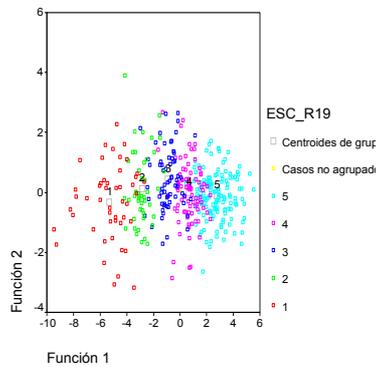


Figura 6.12. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación curso en las carreras superiores

Tabla 6.19. Autovalores del nodo 1 siendo la variable de ramificación titulación en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,156(a)	97,4	97,4	,928
2	,113(a)	1,8	99,1	,319
3	,054(a)	,9	100,0	,227

Tabla 6.20. matriz de estructuras del nodo 1 siendo la variable de ramificación titulación en las carreras técnicas

	Función		
	1	2	3
MEDIA R5	,754(*)	-,610	,243
MEDIA R2	,678	,173	-,714(*)
MEDIA R12	,630	,440	,640(*)

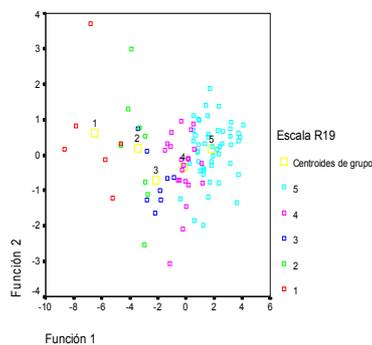


Figura 6.13. Funciones canónicas discriminantes del nodo 1 siendo la variable de ramificación titulaciones en las carreras técnicas

Tabla 6.21. Autovalores del nodo 2 siendo la variable de ramificación titulación en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,260(a)	98,3	98,3	,917
2	,082(a)	1,5	99,9	,275
3	,007(a)	,1	100,0	,086

Tabla 6.22. Matriz de estructura del nodo 2 siendo la variable de ramificación titulación en las carreras técnicas

	Función		
	1	2	3
MEDIA R12	,781(*)	-,623	,047
MEDIA R1	,776(*)	,399	-,487
MEDIA R11	,547	-,047	,836(*)

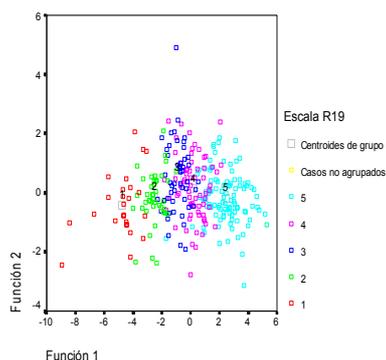


Figura 6.14. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación titulaciones en las carreras técnicas

Tabla 6.23. Autovalores del nodo 3 siendo la variable de ramificación titulación en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,458(a)	99,1	99,1	,904
2	,042(a)	,9	100,0	,200

Tabla 6.24. Matriz de estructura del nodo 3 siendo la variable de ramificación titulación en las carreras técnicas

	Función	
	1	2
MEDIA R1	,804(*)	-,595
MEDIA R12	,775(*)	,632

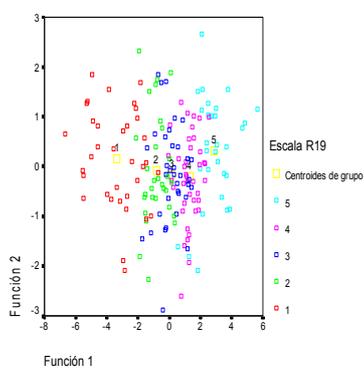


Figura 6.15. Funciones canónicas discriminantes del nodo 3 siendo la variable de ramificación titulaciones en las carreras técnicas

Tabla 6.25. Autovalores del nodo 4 siendo la variable de ramificación titulación en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,357(a)	99,0	99,0	,930
2	,036(a)	,6	99,5	,186
3	,022(a)	,3	99,8	,145
4	,010(a)	,2	100,0	,099

Tabla 6.26. Matriz de estructura del nodo 4 siendo la variable de ramificación titulación en las carreras técnicas

	Función			
	1	2	3	4
MEDIA R1	,691(*)	-,238	-,125	,383
MEDIA R2	,586	-,690(*)	-,363	-,048
MEDIA R18	,408	,627(*)	-,227	,369
MEDIA R11	,465	,041	,815(*)	-,260
MEDIA R12	,620	,364	,057	-,650(*)

Tabla 6.27. Autovalores del nodo 5 siendo la variable de ramificación titulación en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,792(a)	99,4	99,4	,923
2	,038(a)	,6	100,0	,191

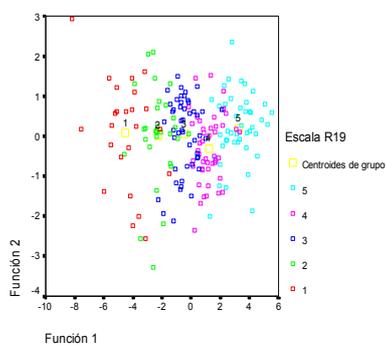


Figura 6.16. Funciones canónicas discriminantes del nodo 4 siendo la variable de ramificación titulaciones en las carreras técnicas

Tabla 6.28. Matriz de estructura del nodo 5 siendo la variable de ramificación titulación en las carreras técnicas

	Función	
	1	2
MEDIA R1	,783(*)	-,622
MEDIA R12	,517	,856(*)

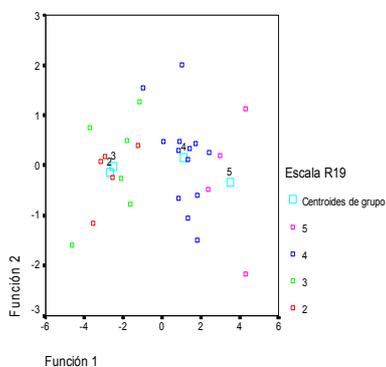


Figura 6.17. Funciones canónicas discriminantes del nodo 5 siendo la variable de ramificación titulaciones en las carreras técnicas

Tabla 6.29. Autovalores del nodo 1 siendo la variable de ramificación departamento en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,793(a)	90,8	90,8	,910
2	,399(a)	7,6	98,3	,534
3	,070(a)	1,3	99,6	,255
4	,020(a)	,4	100,0	,139

Tabla 6.30. Matriz de estructura del nodo 1 siendo la variable de ramificación departamento en las carreras técnicas

	Función			
	1	2	3	4
MEDIA R12	,793(*)	-,246	,165	,405
MEDIA R1	,782(*)	,165	-,383	,192
MEDIA R11	,628	,217	,630(*)	-,189
MEDIA R5	,415	-,425	,544(*)	,185
MEDIA R3	,541	,450	,301	,596(*)

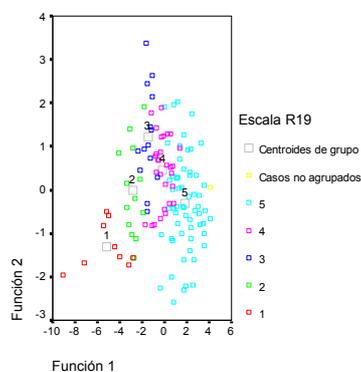


Figura 6.18. Funciones canónicas discriminantes del nodo siendo la variable de ramificación departamento en las carreras técnicas

Tabla 6.31. Autovalores del nodo 2 siendo la variable de ramificación departamento en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,662(a)	99,2	99,2	,922
2	,039(a)	,7	99,8	,193
3	,009(a)	,2	100,0	,094
4	,000(a)	,0	100,0	,013

Tabla 6.32. Matriz de estructura del nodo 1 siendo la variable de ramificación departamento en las carreras técnicas

	Función			
	1	2	3	4
MEDIA R1	,773(*)	,433	-,294	-,357
MEDIA R12	,686(*)	-,665	,250	-,156
MEDIA R2	,627(*)	,160	-,626	,436
MEDIA R11	,471	-,057	,783(*)	,401

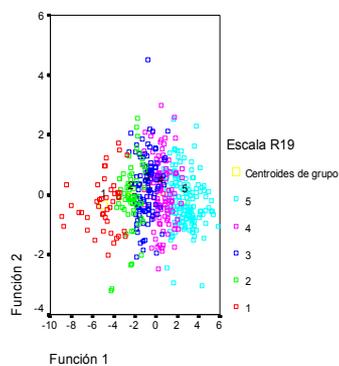


Figura 6.19. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación departamento en las carreras técnicas

Tabla 6.33. Autovalores del nodo 3 siendo la variable de ramificación departamento en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,333(a)	99,6	99,6	,901
2	,019(a)	,4	100,0	,138

Tabla 6.34. Matriz de estructura del nodo 3 siendo la variable de ramificación departamento en las carreras técnicas

	Función	
	1	2
MEDIA R12	,817(*)	-,577
MEDIA R2	,673	,740(*)

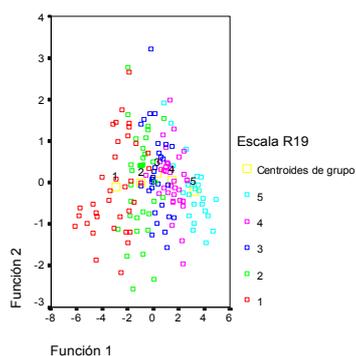


Figura 6.20. Funciones canónicas discriminantes del nodo 3 siendo la variable de ramificación departamento en las carreras técnicas

Tabla 6.35. Autovalores del nodo 4 siendo la variable de ramificación departamento en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,129(a)	86,7	86,7	,927
2	,941(a)	13,3	100,0	,696

Tabla 6.36. Matriz de estructura del nodo 4 siendo la variable de ramificación departamento en las carreras técnicas

	Función	
	1	2
MEDIA R1	,864(*)	-,503
MEDIA R12	,850(*)	,527

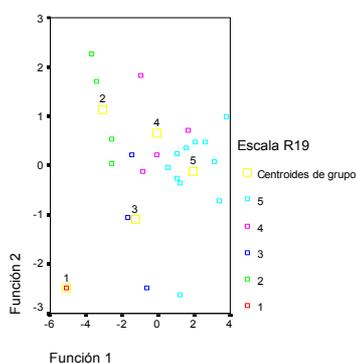


Figura 6.21. Funciones canónicas discriminantes del nodo 4 siendo la variable de ramificación departamento en las carreras técnicas

Tabla 6.37. Autovalores del nodo 1 siendo la variable de ramificación curso en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,996(a)	99,1	99,1	,913
2	,032(a)	,6	99,7	,176
3	,014(a)	,3	100,0	,118
4	,000(a)	,0	100,0	,004

Tabla 6.38. Matriz de estructura del nodo 1 siendo la variable de ramificación curso en las carreras técnicas

	Función			
	1	2	3	4
MEDIA R1	,769(*)	,224	,143	-,581

MEDIA R12	,742(*)	-,614	-,253	,095
MEDIA R5	,732(*)	,486	-,435	,199
MEDIA R18	,575	,100	,695(*)	,419

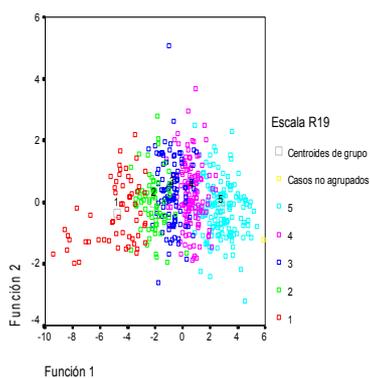


Figura 6.21. Funciones canónicas discriminantes del nodo 1 siendo la variable de ramificación curso en las carreras técnicas

Tabla 6.39. Autovalores del nodo 2 siendo la variable de ramificación curso en las carreras técnicas

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,933(a)	98,8	98,8	,925
2	,067(a)	1,1	99,9	,250
3	,007(a)	,1	100,0	,085
4	,000(a)	,0	100,0	,021

Tabla 6.40. Matriz de estructura del nodo 2 siendo la variable de ramificación curso en las carreras técnicas

	Función			
	1	2	3	4
MEDIA R1	,752(*)	,331	-,477	-,311
MEDIA R12	,719(*)	-,626	,269	-,138
MEDIA R2	,652(*)	,261	-,460	,543
MEDIA R11	,497	,138	,854(*)	,070

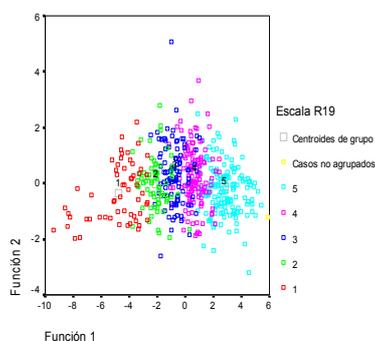


Figura 6.22. Funciones canónicas discriminantes del nodo 2 siendo la variable de ramificación curso en las carreras técnicas

Tabla 6.41. Autovalores del nodo 1 siendo la variable de ramificación departamento en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	14,588(a)	93,6	93,6	,967
2	,997(a)	6,4	100,0	,707

Tabla 6.42. Matriz de estructura del nodo 1 siendo la variable de ramificación departamento en el segundo ciclo

	Función	
	1	2
MEDIA R1	,925(*)	-,381
MEDIA R4	,708(*)	,707

Tabla 6.43. Autovalores del nodo 2 siendo la variable de ramificación departamento en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2,612(a)	95,0	95,0	,850
2	,111(a)	4,0	99,1	,316
3	,026(a)	,9	100,0	,158

Tabla 6.44. Matriz de estructura del nodo 2 siendo la variable de ramificación departamento en el segundo ciclo

	Función		
	1	2	3
MEDIA R2	,761(*)	-,128	-,636
MEDIA R5	,745(*)	-,506	,435
MEDIA R8	,451	,756(*)	,474

Tabla 6.45. Autovalores del nodo 3 siendo la variable de ramificación departamento en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	8,725(a)	99,8	99,8	,947
2	,015(a)	,2	100,0	,120

Tabla 6.46. Matriz de estructura del nodo 3 siendo la variable de ramificación departamento en el segundo ciclo

	Función	
	1	2
MEDIA R1	,885(*)	-,465
MEDIA R10	,403	,915(*)

Tabla 6.47. Autovalores del nodo 4 siendo la variable de ramificación departamento en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	3,751(a)	100,0	100,0	,889

Tabla 6.48. Matriz de estructura del nodo 4 siendo la variable de ramificación departamento en el segundo ciclo

	Función
	1
MEDIA R1	1,000

Tabla 6.49. Autovalores del nodo 1 siendo la variable de ramificación titulación en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,491(a)	98,1	98,1	,931
2	,103(a)	1,6	99,7	,306
3	,021(a)	,3	100,0	,142

Tabla 6.50. Matriz de estructura del nodo 1 siendo la variable de ramificación titulación en el segundo ciclo

	Función		
	1	2	3
MEDIA R1	,819(*)	-,083	-,567
MEDIA R5	,726(*)	-,329	,604
MEDIA R10	,373	,757(*)	,536

Tabla 6.51. Autovalores del nodo 2 siendo la variable de ramificación titulación en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	4,145(a)	100,0	100,0	,898

Tabla 6.52. Matriz de estructura del nodo 2 siendo la variable de ramificación titulación en el segundo ciclo

	Función
	1
MEDIA R1	1,000

Tabla 6.53. Autovalores del nodo 3 siendo la variable de ramificación titulación en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	11,042(a)	98,3	98,3	,958
2	,193(a)	1,7	100,0	,402

Tabla 6.54. Matriz de estructura del nodo 3 siendo la variable de ramificación titulación en el segundo ciclo

	Función	
	1	2
MEDIA R10	,366	,931(*)
MEDIA R5	,650	-,760(*)

Tabla 6.55. Autovalores del nodo 1 siendo la variable de ramificación curso en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	5,833(a)	99,2	99,2	,924
2	,036(a)	,6	99,8	,186
3	,011(a)	,2	100,0	,102

Tabla 6.56. Autovalores del nodo 1 siendo la variable de ramificación curso en el segundo ciclo

	Función		
	1	2	3
MEDIA R1	,812(*)	-,028	-,583
MEDIA R5	,761(*)	-,438	,479
MEDIA R12	,648	,665(*)	,373

Tabla 6.57. Autovalores del nodo 2 siendo la variable de ramificación curso en el segundo ciclo

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,868(a)	100,0	100,0	,934

Tabla 6.58. Matriz de estructura del nodo 2 siendo la variable de ramificación curso en el segundo ciclo

	Función
	1
MEDIA R1	1,000

Tabla 6.59. Poder discriminante de los datos de los modelos departamento 1, 2, 3, 4 y 5 proyectados sobre los modelos del resto de departamentos en las titulaciones superiores

	Departamentos														
	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
MEDIA R1	1.327	1.484	1.487	1.356	1.362	1.490	1.432	1.302	1.451	1.625	1.432	1.634	1.364	1.553	1.360
MEDIA R2	1.281	1.612	1.323	1.299	1.625	1.620	1.361	1.284	1.784	1.791	1.556	1.909	1.488	1.788	1.706
MEDIA R3	1.329	1.586	1.530	1.446	1.692	1.554	1.647	1.416	1.714	2.034	1.410	1.906	1.758	2.395	1.712
MEDIA R4	1.346	1.486	1.290	1.321	1.455	1.473	1.219	1.335	1.402	1.532	1.380	1.562	1.197	1.357	1.294
MEDIA R5	1.298	1.590	1.116	1.259	1.838	1.616	1.288	1.290	2.026	1.840	1.572	2.431	1.208	1.953	1.973
MEDIA R6	1.351	1.821	1.355	1.405	1.500	1.733	1.401	1.371	1.367	2.180	1.699	2.521	1.478	2.424	1.439
MEDIA R7	1.285	1.510	1.204	1.262	1.800	1.460	1.275	1.296	1.754	2.379	1.351	3.192	1.276	2.512	1.819
MEDIA R8	1.293	1.602	1.240	1.265	1.971	1.657	1.276	1.309	2.079	2.343	1.594	3.255	1.263	2.995	1.808
MEDIA R9	1.340	1.752	1.224	1.282	1.775	1.686	1.223	1.307	1.762	2.020	1.553	2.302	1.237	1.722	1.523
MEDIA R10	1.326	1.447	1.219	1.296	1.753	1.506	1.306	1.302	2.040	1.907	1.509	2.786	1.435	3.305	1.694
MEDIA R11	1.297	1.554	1.507	1.277	1.516	1.499	1.399	1.255	1.577	2.054	1.593	1.962	1.543	2.913	1.454
MEDIA R12	1.295	1.439	1.528	1.396	1.351	1.515	1.658	1.432	1.340	2.331	1.469	1.682	1.645	2.128	1.484
MEDIA R16	1.276	1.543	1.203	1.310	1.468	1.723	1.300	1.341	1.604	1.918	1.548	2.434	1.108	1.921	1.513
MEDIA R17	1.316	1.605	1.191	1.296	1.610	1.811	1.360	1.327	1.633	1.894	1.590	2.119	1.324	2.272	1.406
MEDIA R18	1.377	1.430	1.396	1.337	2.077	1.530	1.542	1.392	2.457	2.603	1.538	2.918	1.129	2.422	1.465
MEDIA R19	1.306	1.598	3.540	1.323	1.374	1.642	3.446	1.279	1.396	4.182	1.657	1.783	3.676	3.598	1.517

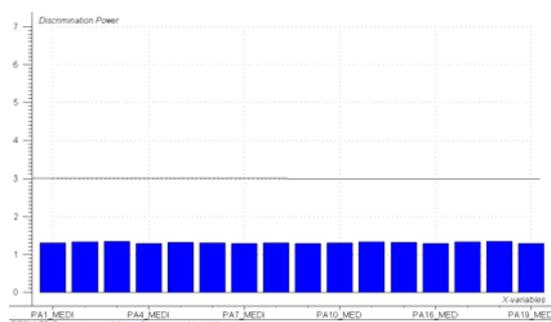


Figura 6.23. Poder discriminante de los datos de los modelos curso 1 y 2 proyectados sobre los modelos del resto de departamentos en las titulaciones superiores

Tabla 6.60. Poder discriminante de los datos de los modelos departamento 1, 2 y 3 proyectados sobre los modelos del resto de departamentos en las titulaciones técnicas

Departamento						
	1-2	1-3	1-4	2-3	2-4	3-4
MEDIA R1	3.277	2.399	2.365	2.5	2.801	2.037
MEDIA R2	3.568	2.701	2.551	2.2	2.855	2.123
MEDIA R3	3.440	2.571	2.085	2.5	2.418	2.301
MEDIA R4	4.217	2.174	2.010	2.0	2.155	1.711
MEDIA R5	3.357	1.593	2.428	1.9	2.670	2.238
MEDIA R6	3.943	2.436	2.296	2.0	2.357	2.154
MEDIA R7	3.578	2.076	1.757	2.2	2.174	2.291
MEDIA R8	4.798	2.213	1.916	2.2	2.038	2.232
MEDIA R9	4.255	2.351	2.524	1.9	2.476	1.760
MEDIA R10	4.311	2.212	1.830	1.7	1.962	2.036
MEDIA R11	3.429	2.050	2.336	2.1	2.724	1.797
MEDIA R12	3.674	2.493	2.352	2.3	2.671	2.540
MEDIA R16	2.649	2.061	1.999	2.1	2.171	2.274
MEDIA R17	3.560	2.044	2.581	1.9	2.972	2.572
MEDIA R18	3.394	1.907	2.325	2.0	2.920	2.522
MEDIA R19	3.147	2.991	2.193	2.5	2.305	1.677

Tabla 6.61. Poder discriminante de los datos de los modelos titulaciones 1, 2, 3 y 4 proyectados sobre los modelos del resto de departamentos en las titulaciones técnicas

Departamento										
	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
MEDIA R1	2.145	3.231	4.129	2.687	2.004	1.963	1.547	3.798	2.187	2.510
MEDIA R2	1.857	3.062	4.273	3.284	2.001	1.802	1.678	3.572	3.035	3.071
MEDIA R3	2.066	3.204	3.968	2.822	2.030	1.790	1.693	3.553	2.461	2.345
MEDIA R4	1.334	2.918	3.962	3.023	1.556	1.355	1.639	3.523	3.513	3.322
MEDIA R5	1.643	3.899	3.791	3.415	1.743	1.498	1.742	3.592	3.091	3.019
MEDIA R6	1.325	3.444	4.164	3.085	1.541	1.337	1.627	3.452	3.108	2.626
MEDIA R7	1.659	3.757	4.174	2.634	1.480	1.573	1.929	3.526	2.428	2.295
MEDIA R8	1.406	2.822	2.628	3.008	1.502	1.522	2.000	3.411	2.841	2.953
MEDIA R9	1.685	4.550	4.309	2.650	1.322	1.233	1.507	4.327	2.009	1.877
MEDIA R10	1.520	3.752	4.563	3.047	1.670	1.385	1.608	3.644	2.530	2.534
MEDIA R11	1.782	3.756	3.633	3.175	1.666	1.536	1.831	3.664	2.964	2.762
MEDIA R12	1.414	3.266	4.006	2.635	1.644	1.466	1.387	3.525	3.064	2.604
MEDIA R16	1.772	4.534	4.351	3.285	1.604	1.374	1.978	3.278	2.896	2.892
MEDIA R17	1.886	4.447	4.332	3.187	1.551	1.340	1.719	3.180	2.825	2.648
MEDIA R18	1.917	4.390	4.952	2.739	1.510	1.525	2.017	4.699	2.528	2.618
MEDIA R19	2.204	3.201	4.228	2.952	2.337	2.945	1.908	3.927	3.412	3.574

MEDIA R9	1.638	1.519	3.330
MEDIA R10	2.113	2.219	5.813
MEDIA R11	1.831	1.837	4.681
MEDIA R12	1.402	2.805	5.315
MEDIA R16	1.354	1.466	6.041
MEDIA R17	1.368	2.516	6.583
MEDIA R18	1.562	1.447	4.761
MEDIA R19	1.718	2.746	4.197

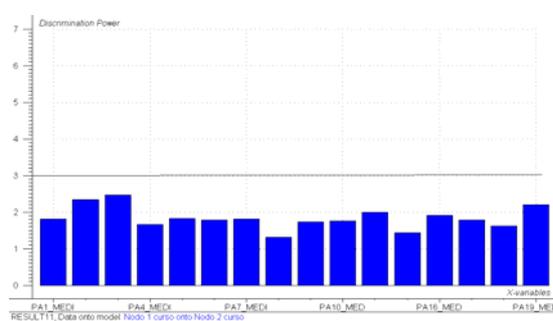


Figura 6.25. Poder discriminante de los datos de los modelos curso 1 y 2 proyectados sobre los modelos del resto de departamentos en las titulaciones de segundo ciclo

Tabla 6.65. Distancia entre modelos de las titulaciones en las titulaciones superiores

	Nodo 1	Nodo 2	Nodo 3	Nodo 4
Nodo 1 titula	1.000	28.415	2.656	2.511
Nodo 2 titula	28.415	1.000	3.067	3.084
Nodo 3 titula	2.656	3.067	1.000	1.732
Nodo 4 titula	2.511	3.084	1.732	1.000

Tabla 6.66. Distancia entre modelos de los cursos en las titulaciones superiores

	Nodo 1
Nodo 1	1.000
Nodo 2	1.706

Tabla 6.67. Distancia entre modelos de los departamentos en las titulaciones técnicas

	Nodo 1	Nodo 2	Nodo 3	Nodo 4
Nodo 1	1.000	11,618	5.400	4.740
Nodo 2	11.618	1.000	4.931	5.939
Nodo 3	5.400	4.931	1.000	4.454
Nodo 4	4.740	5.939	4.454	1.000

Tabla 6.68. Distancia entre modelos de las titulaciones en las titulaciones técnicas

	Nodo 1	Nodo 2	Nodo 3	Nodo 4	Nodo 5
Nodo 1 titula	1.000	2.780	12.189	17.237	8.778

Nodo 2 titula	2.780	1.000	2.848	2.598	2.910
Nodo 3 titula	12.189	2.848	1.000	14.370	8.177
Nodo 4 titula	17.237	2.598	14.370	1.000	7.672
Nodo t titula	8.778	2.910	8.177	7.672	1.000

Tabla 6.69. Distancia entre modelos de los cursos en las titulaciones técnicas

	Nodo 2
Nodo 1	1.000
Nodo 2	9.842

Tabla 6.70. Distancia entre modelos de los departamentos en las titulaciones de segundo ciclo

	Nodo 1	Nodo 2	Nodo 3	Nodo 4
Nodo 1	1.000	2.312	2.318	2.659
Nodo 2	2.312	1.000	2.237	4.1143
Nodo 3	2.318	2.237	1.000	4.588
Nodo 4	2.659	4.143	4.588	1.000

Tabla 6.71. Distancia entre modelos de las titulaciones en las titulaciones de segundo ciclo

	Nodo 1	Nodo 2	Nodo 3
Nodo 1 titula	1.000	2.590	4.280
Nodo 2 titula	2.590	1.000	20.986
Nodo 3 titula	4.280	25.986	1.000

Tabla 6.72. Distancia entre modelos de los cursos en las titulaciones de segundo ciclo

	Nodo 2
Nodo 1	1.000
Nodo 2	3.209

Notación y Abreviaturas

Con la finalidad de simplificar la escritura y debido a la gran cantidad de métodos mencionados se utilizarán algunas abreviaturas que se encuentran a continuación.

Asimismo, se incluyen las abreviaturas utilizadas en el texto para referirnos a las distintas titulaciones.

ACP	Análisis de Componentes Principales
AF	Análisis Factorial
AFC	Análisis del Factor Común
AGFI	Adjusted Goodness Fit Index
CEDA	Cuestionario de Evaluación Docente mediante Opinión de los Alumnos
CF	Cluster Feature
CFA	Análisis Factorial Confirmatorio
CHAID	Chi Squared Automatic Interaction Detector
EE	Encuesta de Evaluación
EPS ALCOY	Escuela Politécnica Superior de Alcoy
EPS ALCOY-ADE	Escuela Politécnica Superior de Alcoy Administración y Dirección de Empresas
EPS ALCOY-I.MATER.	Escuela Politécnica Superior de Alcoy
EPS ALCOY-INF.GES.	Escuela Politécnica Superior de Alcoy. (Especialidad Informática de Gestión)
EPS ALCOY-IT DISEÑO	Escuela Politécnica Superior de Alcoy Ingeniería Técnica Diseño
EPS ALCOY-IT TELECOM	Escuela Politécnica Superior de Alcoy Ingeniería Técnica Telecomunicaciones
EPS ALCOY-ITI E.IN	Escuela Politécnica Superior de Alcoy. Ingeniería Técnica Industrial (Especialidad Electrónica Industrial)
EPS ALCOY-ITI ELEC	Escuela Politécnica Superior de Alcoy Ingeniería Técnica Industrial (Especialidad Electricidad)
EPS ALCOY-ITI MEC.	Escuela Politécnica Superior de Alcoy Ingeniería Técnica Industrial (Especialidad Mecánica)
EPS ALCOY-ITI QUIM	Escuela Politécnica Superior de Alcoy. Ingeniería Técnica

	Industrial (Especialidad Química)
EPS ALCOY-ITI TEXT	Escuela Politécnica Superior de Alcoy. Ingeniería Técnica Industrial (Especialidad Textil)
EPS ALCOY-ORG.IND.	Escuela Politécnica Superior de Alcoy Organización Industrial
EPS GANDIA-AUDIOV.	Escuela Politécnica Superior de Gandia. Licenciado en Comunicación Audiovisual.
EPS GANDIA-CC.AA.	Escuela Politécnica Superior de Gandía. Licenciatura en Ciencias Ambientales
EPS GANDIA-ELECTR.	Escuela Politécnica Superior de Gandia (Especialidad Sistemas Electrónicos)
EPS GANDIA-FOREST.	Escuela Politécnica Superior de Gandia. Ingeniero Técnico Forestal
EPS GANDIA-SONIDO	Escuela Politécnica Superior de Gandia (Especialidad Sonido e Imagen)
EPS GANDIA-TELECO.	Escuela Politécnica Superior de Gandia (Especialista Sistemas Telecomunicaciones)
EPS GANDIA-TURISMO	Escuela Politécnica Superior de Gandía. (Especialidad Sonido e Imagen)
ER	Rango Expandido
ETS ARQUITECTURA	Escuela Técnica Superior de Arquitectura
ETSI GEODESICA C.T	Escuela Técnica Superior de Geodésica, Cartografía y Topografía
ETSI INDUSTRIALES	Escuela Técnica Superior de Ingenieros Industriales
ETSI TELECO	Escuela Técnica Superior de Ingenieros de Telecomunicaciones
ETSIA-ING. AGRONO	Escuela Técnica Superior de Ingenieros Agrónomos
ETSIA-TECNO. ALIM.	Licenciatura en Ciencia y Tecnología de los Alimentos
ETSICCP-CCAA	Escuela Técnica Superior de Ingenieros de Caminos Canales y Puertos. Ciencias Ambientales
ETSICCP-CCP	Escuela Técnica Superior de Ingenieros de Caminos Canales y Puertos
ETSICCP-CCP PLAN97	Escuela Técnica Superior de Ingenieros de Caminos Canales y Puertos (Plan 97)
ETSICCP-OOPP	Escuela Técnica Superior de Ingenieros de Caminos Canales

	y Puertos (Obras Públicas)
ETSICCP-OOPP P97	Escuela Técnica Superior de Ingenieros de Caminos Canales y Puertos (Obras Públicas)
ETSICCP-PIE (IC)	Escuela Técnica Superior de Ingenieros de Caminos Canales y Puertos (PIE)
ETSIGCT-TOPOGRAFIA	Escuela Técnica Superior de Geodésica, Cartografía y Topografía (Ingeniería Técnica de Topografía)
ETSII-ADMÓN Y D.E.	Escuela Técnica Superior de Ingenieros Industriales en Administración y Dirección de Empresas
ETSII-AUTOMÁTICA	Escuela Técnica Superior de Ingenieros Industriales (Especialidad Automática)
ETSII-MATERIALES	Escuela Técnica Superior de Ingenieros Industriales (Ingeniería de los Materiales)
ETSII-ORG. INDUSTRIAL	Escuela Técnica Superior de Ingenieros Industriales (Especialidad Organización Industrial)
ETSII-PLAN 93	Escuela Técnica Superior de Ingenieros Industriales. Plan 93
ETSII-QUÍMICO	Escuela Técnica Superior de Ingenieros Industriales (Especialidad Química)
EU ARQUIT. TÉCNICA	Escuela de Arquitectura Técnica
EU INFORMÁTICA	Ingeniería Técnica Informática
EUIT AGRÍCOLA	Escuela Técnica Superior del Medio Rural y Enología (Especialidad Agrícolas)
EUIT INDUSTRIAL	Ingeniería Técnica Industrial
EUITA-EXPL. AGROP.	Escuela Técnica Superior del Medio Rural y Enología (Especialidad explotaciones Agropecuarias)
EUITA-HORTO Y JARD	Escuela Técnica Superior del Medio Rural y Enología (Especialidad Horto y Jardinería)
EUITA-IND. AGR. Y A	Escuela Técnica Superior del Medio Rural y Enología (Especialidad Industrias Agroalimentarias)
EUITA-MECAN. Y C.R	Escuela Técnica Superior del Medio Rural y Enología (Especialidad Mecánica y Caminos Rurales)
EUITI-DISEÑO INDU.	Escuela Técnica Industrial. (Especialidad Diseño Industrial)
EUITI-ELECTRICIDAD	Escuela Técnica Industrial. (Especialidad Electricidad)
EUITI-ELECTRÓNICA	Escuela Técnica Industrial. (Especialidad Electrónica)

EUITI-MECANICA	Escuela Técnica Industrial. (Especialidad Mecánica)
EUITI-QUIMICA IND.	Escuela Técnica Industrial. (Especialidad Química Industrial)
FI	Facultad de Informática
FI-DOCUMENTACION	Facultad Informática. Licenciado en Documentación
GFI	Godness of Fit Index
HCA	Análisis Cluster jerárquico
ICE	Instituto de las Ciencias de la Educación
LDA	Análisis Discriminante Lineal
MV	Máxima Versimilitud
NFI	Normal Fit Index
NIPALS	Nonlinear Iterative Partial Least Squares
NNFI	No Normal Fit Index
PC	II Plan de Calidad de las Universidades
PNECU	Plan Nacional de la Calidad de la Universidades Españolas
RMSEA	Approximation Root Mean Square Residual
RMSR	Root Mean Square Residual
SEM	Modelos de Ecuaciones Estructurales
SIMCA	Soft Independent Modelling of Class Analogy
UPV	Universidad Politécnica de Valencia