

Estrategias metodológicas para la construcción de indicadores compuestos en la gestión universitaria

MARÍA DEL CARMEN BAS CERDÁ

**EDITORIAL
UNIVERSITAT POLITÈCNICA DE VALÈNCIA**

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
Departamento de Estadística e Investigación Operativa Aplicadas y Calidad



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

**ESTRATEGIAS METODOLÓGICAS PARA LA
CONSTRUCCIÓN DE INDICADORES COMPUESTOS EN
LA GESTIÓN UNIVERSITARIA**

TESIS DOCTORAL

Presentada por:

María del Carmen Bas Cerdá

Dirigida por:

José Miguel Carot Sierra

José Jabaloyes Vivas

Valencia, enero 2014

Colección Tesis Doctorales

© María del Carmen Bas Cerdá

© 2014, de la presente edición: Editorial Universitat Politècnica de València
Telf.: 963 877 012 / www.lalibreria.upv.es

ISBN: 978-84-9048-194-3 (versión CD)

Queda prohibida la reproducción, distribución, comercialización, transformación, y en general, cualquier otra forma de explotación, por cualquier procedimiento, de todo o parte de los contenidos de esta obra sin autorización expresa y por escrito de sus autores.

Agradecimientos

Muchos científicos encuentran sus mejores ideas mientras se toman un café o se beben unas copas con amigos. Indudablemente el café me ha acompañado mucho a lo largo de estos cuatro años de arduo trabajo, pero mi principal fuente de inspiración la he encontrado en las personas que me rodean.

En primer lugar quiero dar las gracias a mis directores de tesis, José Miguel Carot y José Jabaloyes, por la labor de dirección de la misma. En particular, a José Miguel Carot por la confianza puesta en mí para la realización de este trabajo, por su gran apoyo y por formarme profesionalmente durante todos estos años.

Muy especialmente quiero agradecer la ayuda incondicional y desinteresada de Stefano Tarantola y de todo su equipo de investigación del Joint Research Centre de la Comisión Europea de Ispira, y sus excelentes orientaciones en los fundamentos metodológicos de la construcción de Indicadores Compuestos.

Asimismo quiero dar las gracias al profesor y amigo David Jornet por su apoyo y ánimo en los últimos momentos de la realización de la tesis con sus grandes consejos matemáticos. Y a mis compañeros/as del Centro de Gestión de la Calidad y del Cambio por su apoyo constante en el día a día.

Por último, pero no menos importante, quiero agradecer con mucho cariño a mi padre, a mi madre y a mi hermano por estar ahí siempre y animarme a cumplir mis objetivos, tanto profesionales como personales. Y a Javi, por su apoyo incondicional y su infinita paciencia en los momentos más duros de la realización de esta Tesis Doctoral.

Sin todos ellos no hubiese sido posible la elaboración de este trabajo de investigación. Muchas gracias a todos.

*Essentially, all models are wrong,
but some are useful*
George E.P. Box, 1979

Resumen

En los últimos años, las instituciones de educación superior se han sometido a diversos cambios condicionados, en gran medida, por las exigencias de la sociedad actual. El cambio de una universidad de élites a una universidad de masas, la creciente demanda en la mejora de la calidad, la globalización del conocimiento, la competitividad a nivel nacional e internacional, la elaboración de procedimientos de mejora internos, los cambios en las relaciones con la administración y con la sociedad en general; son algunos de los factores ineludibles y de creciente importancia en el ámbito de la educación superior en Europa y en todo el mundo.

La rendición de cuentas, para aumentar la confianza entre sociedad y universidad, y la exigencia de transparencia, para el diseño y ejecución de planes de mejora y de toma de decisiones, son aspectos clave para responder a las necesidades de la sociedad. Sin embargo, es un hecho que las universidades siguen siendo más opacas de lo que es deseable, a pesar de que en los últimos años haya habido un esfuerzo común por cambiar esta situación.

La universidad española ha desarrollado diversas iniciativas para evaluar la calidad de las actividades universitarias. En particular, uno de los retos relevantes para garantizar la calidad en el sistema universitario español es la evaluación de la actividad del profesorado universitario, observándose la importancia cada vez mayor de las actividades de evaluación que aseguren la calidad de los procesos formativos que afecten a los estudiantes, a la comunidad universitaria y a la sociedad en general. Los resultados de la evaluación permitirán a las universidades el diseño y la ejecución de planes de mejora de la actuación del profesorado. Por tanto, es importante disponer de herramientas de medición que permitan evaluar la actividad del profesorado universitario en sus tres dimensiones: docencia, investigación y gestión. Sin embargo, muchas veces el instrumento de evaluación utilizado no responde a la necesidad de transparencia exigida por la sociedad.

El objetivo de este trabajo es diseñar estrategias metodológicas para la construcción de indicadores compuestos que permitan evaluar y medir conceptos multidimensionales relacionados con la gestión de la educación superior. Para ello, se realiza una revisión exhaustiva de las diferentes técnicas estadísticas que se pueden aplicar en el desarrollo de indicadores compuestos. Asimismo, se estudia la importancia de los indicadores como herramienta básica para una correcta gestión de la calidad y para la rendición de cuentas en las universidades.

Una vez realizada una revisión sobre el panorama de los indicadores en la educación superior, este trabajo se centra en la evaluación de las actividades académicas del personal docente e investigador de la Universitat Politècnica de València. En primer lugar, se realiza un análisis exploratorio del conjunto de indicadores propuestos por la universidad para la evaluación docente del profesorado, en el que se aplican técnicas estadísticas de reducción de la

dimensionalidad y técnicas de clasificación. Posteriormente, se proponen diferentes metodologías de construcción de un indicador compuesto que mida la actividad docente del profesorado, según el tipo de normalización y ponderación aplicado. Finalmente, se estudian las características analíticas de cada una de las metodologías propuestas mediante técnicas de análisis de incertidumbre y de sensibilidad global para evaluar la calidad y robustez de los modelos propuestos, según los objetivos planteados por la universidad. El análisis de incertidumbre se realiza mediante un procedimiento de simulación cuasialeatoria y se visualiza aplicando una técnica de agregación no compensatoria sobre los resultados, conocida como método de Copeland. Para realizar el análisis de sensibilidad global se emplea una metodología no paramétrica conocida como *State Dependent Parameter* basada en el Filtro de Kalman y en técnicas de suavizado (*smoothing*). El análisis de incertidumbre y de sensibilidad se consideran técnicas necesarias que aportan validez y sustento metodológico a los indicadores compuestos planteados.

Finalmente, se estudia la generalización del indicador compuesto referido a la actividad docente y se diseña un indicador compuesto en el que se aplica una técnica de ponderación y agregación no paramétrica conocida como *Benefit of the Doubt* para evaluar la actividad académica de cada docente con respecto al resto, en base a las tres dimensiones principales del profesorado universitario: docencia, investigación y gestión universitaria. Así pues, esta técnica permite medir la eficacia relativa en el desarrollo de la actividad académica de cada docente según su perfil académico.

Las técnicas metodológicas propuestas en este trabajo resultan aplicables en cualquier ámbito de estudio y ofrecen un esquema de construcción de indicadores compuestos robusto y transparente.

Resum

Els últims anys, les institucions d'educació superior s'han sotmès a diversos canvis condicionats, en gran mesura, per les exigències de la societat actual. El canvi d'una universitat d'elits a una universitat de masses, la creixent demanda de millora de la qualitat, la globalització del coneixement, la competitivitat a nivell nacional i internacional, l'elaboració de procediments de millora interns, els canvis en les relacions amb l'administració i amb la societat en general; són alguns dels factors ineludibles i de creixent importància en l'àmbit de l'educació superior a Europa i arreu del món.

La rendició de comptes, per a augmentar la confiança entre societat i universitat, i l'exigència de transparència, en el disseny i l'execució de plans de millora i en la presa de decisions, són aspectes clau per a respondre a les necessitats de la societat. No obstant això, és un fet que les universitats continuen sent més opaques del que fóra desitjable, malgrat que els últims anys s'haja fet un esforç comú per canviar aquesta situació.

La universitat espanyola ha desenvolupat diverses iniciatives per a avaluar la qualitat de les activitats universitàries. En particular, un dels reptes rellevants per a garantir la qualitat en el sistema universitari espanyol és l'avaluació de l'activitat del professorat universitari, atesa la importància cada vegada major de les activitats d'avaluació que asseguren la qualitat dels processos formatius que afecten els estudiants, la comunitat universitària i la societat en general. Els resultats de l'avaluació permetran a les universitats el disseny i l'execució de plans de millora de l'actuació del professorat. Per tant, és important disposar d'eines de mesurament que permeten avaluar l'activitat del professorat universitari en les tres dimensions que comprèn aquesta: docència, investigació i gestió. No obstant això, moltes vegades l'instrument d'avaluació utilitzat no respon a la transparència exigida per la societat.

L'objectiu d'aquest treball és dissenyar estratègies metodològiques per a la construcció d'indicadors compostos que permeten avaluar i mesurar conceptes multidimensionals relacionats amb la gestió de l'educació superior. Per a fer-ho, es porta a terme una revisió exhaustiva de les diferents tècniques estadístiques que es poden aplicar al desenvolupament d'indicadors compostos. Així mateix, s'estudia la importància dels indicadors com a eina bàsica per a una correcta gestió de la qualitat i per a la rendició de comptes a les universitats.

Una vegada realitzada la revisió del panorama dels indicadors en l'educació superior, aquest treball se centra en l'avaluació de les activitats acadèmiques del personal docent i investigador de la Universitat Politècnica de València. En primer lloc, es fa una anàlisi exploratòria del conjunt d'indicadors proposats per la universitat per a l'avaluació docent del professorat, en què s'apliquen tècniques estadístiques de reducció de la dimensionalitat i tècniques de classificació. Posteriorment, es proposen diferents metodologies de construcció d'un indicador compost que mesure l'activitat docent del professorat, segons el tipus de normalització i ponderació aplicat.

Finalment, s'estudien les característiques analítiques de cadascuna de les metodologies proposades mitjançant tècniques d'anàlisi d'incertesa i de sensibilitat global, a fi d'avaluar la qualitat i robustesa dels models proposats, segons els objectius plantejats per la Universitat. L'anàlisi d'incertesa es porta a terme mitjançant un procediment de simulació quasialeatòria, i es visualitza aplicant una tècnica de agregació no compensatòria sobre els resultats, coneguda com mètode de Copeland. Per a realitzar l'anàlisi de sensibilitat global s'empra una metodologia no paramètrica coneguda com *State Dependent Parameter* i basada en el filtre de Kalman i en tècniques d'allisatge (*smoothing*). L'anàlisi d'incertesa i la de sensibilitat es consideren com tècniques necessàries que aporten validesa i fonament metodològic als indicadors compostos plantejats.

Finalment, s'estudia la generalització de l'indicador compost referit a l'activitat docent, i es dissenya un indicador compost en el qual s'aplica una tècnica de ponderació i agregació no paramètrica, coneguda com *Benefit of the Doubt*, per avaluar l'activitat acadèmica de cada docent respecte a la resta, sobre la base de les tres dimensions principals de la tasca del professorat universitari: docència, investigació i gestió universitària. Aquesta tècnica permet mesurar l'eficàcia relativa de cada docent segons el seu perfil acadèmic.

Les tècniques metodològiques proposades en aquest treball són aplicables a qualsevol àmbit d'estudi, i ofereixen un esquema de construcció d'indicadors compostos robust i transparent.

Abstract

Over the past few years, higher education institutions have been submitted to various changes largely conditioned by the demands of society. The change from a university for the elite to a university for the masses, the increasing demand for the improvement in quality, the globalization of knowledge, the competitiveness at national and international levels, the development of internal improvement procedures, the changes in relationships with administration and society in general; are some of the unavoidable factors of growing importance in the field of higher education in Europe and around the world.

Accountability, to build trust between society and universities, and the demand for transparency, for the design and implementation of plans for improvement and decision-making, are key aspects for meeting society's needs. Nevertheless, it is a fact that universities offer less openness than desired, although in recent years there has been a common effort to change this situation.

Spanish universities have developed several initiatives to assess the quality of university activities. In particular, one of the relevant challenges to ensure quality in the Spanish university system is the assessment of the university teaching staff activity. The importance of the assessment activities is ever increasing. Such activities can help ensure the quality of educational processes affecting students, the university community and society in general. The evaluation results will allow the design and implementation of plans for improvement for the performance of university teaching staff at universities. It is therefore important to have measurement tools to assess the university teaching staff activity in three dimensions: teaching, research and management. However, commonly the assessment instrument used does not meet the need for the transparency which society demands.

The aim of this research is to design methodological strategies for building composite indicators to assess and measure multidimensional concepts related to higher education management. To that end, a thorough review of the statistical techniques that can be applied in the development of composite indicators is performed. Furthermore, the importance of the indicators as a basic tool for suitable quality management and accountability in universities is studied.

After reviewing a summary of the use of indicators in higher education, this study focuses on the assessment of the academic activities of the Universitat Politècnica de València teaching staff. First of all, an exploratory analysis of all indicators proposed by the university, for their teaching staff assessment, is performed. To do this, statistical techniques of dimensionality reduction, and classification techniques are applied. Subsequently, different methodologies for constructing a composite indicator that measures teaching activity are proposed, according to the normalization and weighting techniques applied. Lastly, the analytical characteristics of

each of the methodologies proposed are studied, using uncertainty and global sensitivity analysis techniques to evaluate the quality and robustness of the proposed models, according to the goals set by the university. The uncertainty analysis is performed using a quasi-random simulation process and applying a non-compensatory aggregation technique on the results, known as the Copeland method. A nonparametric approach, known as the State Dependent Parameter, based on the Kalman filter and smoothing techniques, is performed to apply global sensitivity analysis. The uncertainty and sensitivity analysis are considered necessary techniques to provide validity and methodological support for the composite indicators proposed.

Finally, the generalization of the composite indicator referring to teaching activity is studied, and a composite indicator is designed to evaluate the academic activity of each teacher with respect to the rest. This is based on the three main dimensions of university teachers: teaching, research and university management. In order to do this, a non-parametric weighting and aggregation technique called Benefit of the Doubt is applied. This technique measures the relative effectiveness in the development of academic activities of each member of university teaching staff, according to their academic profile.

The methodological techniques proposed in this study are applicable in any field of research, and they provide a robust and transparent construction scheme of composite indicators.

Índice general

1. Introducción	27
2. Objetivos	33
3. Metodología y proceso de construcción de un Indicador Compuesto	37
3.1. Introducción a los Indicadores Compuestos	37
3.2. Consideraciones preliminares	39
3.2.1. Definición y objetivo de un indicador	39
3.2.2. Escalas de medida de un indicador	41
3.2.3. Tipos de indicadores	41
3.2.4. Propiedades que debe cumplir un indicador	42
3.2.5. Propiedades de un sistema de indicadores	43
3.3. Definición y objetivo de un Indicador Compuesto	44
3.4. Construcción de un Indicador Compuesto	48
3.4.1. Desarrollo de un marco conceptual teórico	53
3.4.2. Selección del conjunto de indicadores simples	54
3.4.3. Tratamiento de datos	56
3.4.4. Análisis multivariante	66
3.4.5. Normalización de los datos	80
3.4.6. Ponderación y agregación de los indicadores simples	89
3.4.7. Análisis de incertidumbre y sensibilidad del Indicador Compuesto	105
3.4.8. Conexión con otros indicadores	117
3.4.9. Identificación de los indicadores simples en el Indicador Compuesto	117
3.4.10. Visualización de los resultados	119
4. El papel de los indicadores en el contexto académico de la educación superior	125
4.1. El concepto de ‘calidad universitaria’	125
4.1.1. Diferentes enfoques del concepto ‘calidad’ y ‘calidad universitaria’	125
4.1.2. Hacia una nueva concepción de ‘calidad universitaria’	129
4.2. La evaluación como estrategia de mejora de la calidad educativa	131
4.3. Obtención de información en evaluación: el uso de indicadores	135
4.3.1. Contexto y orígenes de los indicadores en el ámbito de la educación superior	135
4.3.2. Controversia sobre el uso de indicadores en la educación superior	137
4.4. El papel de los indicadores de educación superior en proyectos de investigación internacionales	139

4.4.1. El proyecto “European indicators and ranking methodology for university third mission” (E3M).....	140
4.4.2. El proyecto “Sistema integral de información sobre las instituciones de educación superior de América Latina para el área común de educación con Europa” (INFOACES)	141
4.4.3. El proyecto “The european classification of higher education institutions” (U-MAP)	145
4.4.4. El proyecto “Multi-dimensional global ranking of universities, a feasibility project” (U-MULTIRANK).....	146
4.5. Un panorama de los rankings universitarios.....	148
5. Evaluación de la actividad docente y académica del profesorado universitario.....	155
5.1. Introducción	155
5.2. Modelo de evaluación de la actividad docente del profesorado aplicado al contexto académico universitario español.....	156
5.2.1. Modelo de evaluación de la actividad docente del profesorado de la Universitat Politècnica de València.....	159
5.3. Modelo de evaluación de la actividad académica del profesorado propuesto por la UPV	176
6. Propuesta metodológica para la construcción de un Indicador Compuesto que evalúe la actividad docente en la UPV	179
6.1. Introducción	179
6.2. Técnicas metodológicas para la construcción del Índice de Actividad Docente del profesorado de la UPV.....	180
6.2.1. Desarrollo del marco conceptual teórico.....	180
6.2.2. Selección del conjunto de indicadores simples.....	182
6.2.3. Tratamiento de datos.....	183
6.2.4. Análisis multivariante	186
6.2.5. Normalización de los datos.....	218
6.2.6. Ponderación y agregación de los indicadores simples	220
6.2.7. Escenarios de trabajo	223
6.2.8. Análisis de Sensibilidad Global	225
6.2.9. Análisis de Incertidumbre	247
6.3. Resumen y conclusiones.....	259
7. La evaluación de la actividad académica del profesorado universitario	263
7.1. Introducción	263
7.2. Índice de Actividad Académica del profesorado de la UPV.....	265

7.2.1. Desarrollo del marco conceptual teórico y definición de los indicadores del estudio	265
7.2.2. Tratamiento de datos	266
7.2.3. Análisis del Índice de Actividad Académica propuesto por la UPV	267
7.2.4. Ponderación y agregación mediante la técnica DEA-BOD	278
7.2.5. Escenarios de trabajo propuestos mediante la técnica BOD	289
7.2.6. Visualización de los resultados de la metodología BOD	323
7.3. Resumen y conclusiones.....	327
8. Conclusiones y líneas futuras de investigación.....	329
9. Referencias	335
10. Apéndices.....	353
Anexo A. Análisis Envolvente de Datos. Modelo DEA-CCR	353
Anexo B. Resultados del análisis de sensibilidad global	356
Anexo C. Resultados del análisis de incertidumbre	362
Anexo D. Secuencias de baja discrepancia	367
Anexo E. Escenarios de trabajo propuestos para la construcción del IAA	369

Índice de tablas

3.1. Ventajas y limitaciones del uso de ICs.....	45
3.2. Etapas de la construcción de un IC	52
3.3. Criterios de selección del número de componentes principales.....	71
3.4. Supuestos para la aplicación del ACP/AF.....	73
3.5. Distancias más empleadas para las técnicas de conglomerados con datos cuantitativos	76
3.6. Ventajas e inconvenientes de la aplicación de las técnicas de análisis multivariante	79
3.7. Características más importantes de los métodos de normalización.....	88
3.8. FA aplicado al IC TAI mediante el método de PCA con una rotación VARIMAX	91
3.9. Diferencias entre la agregación lineal y geométrica	103
3.10. Compatibilidad entre métodos de ponderación y métodos de agregación	105
4.1. Rankings nacionales e internacionales según el país donde se elaboran.....	151
5.1. Evaluación de la labor docente en las universidades públicas españolas.....	158
5.2. Relación de los indicadores simples con las dimensiones y criterios del modelo de evaluación	171
6.1. Abreviaturas de los indicadores simples	183
6.2. Estadísticos descriptivos de los indicadores.....	184
6.3. Parámetros λ empleados en las transformaciones Box-Cox.....	190
6.4. Matriz de correlación de los indicadores simples	191
6.5. Matriz de correlaciones anti-imagen de los indicadores simples	193
6.6. Valores propios de la matriz de correlaciones y porcentaje de varianza explicada por cada uno de los factores.....	196
6.7. Método de extracción: PCA. Método de rotación: Normalización Varimax con Kaiser ...	198
6.8. Coeficiente alpha de Cronbach para cada constructo del análisis factorial.....	200
6.9. Coeficiente alpha de Cronbach tras la eliminación de cada uno de los indicadores	200
6.10. Criterio bayesiano para la selección del número de conglomerados con los indicadores simples y la capacidad docente	204
6.11. Distribución de las observaciones en los seis conglomerados	205
6.12. Distribución de las observaciones con respecto al indicador binario en los seis conglomerados	206
6.13. de centroides.....	207
6.14. Correlación del indicador CAPDOC con el resto de indicadores simples	208
6.15. Porcentaje de profesores según distintas categorías asociados cada uno de los seis clusters	209
6.16. Criterio bayesiano para la selección del número de conglomerados con todos los indicadores simples	213

6.17. Distribución de las observaciones en los cuatro conglomerados	213
6.18. Distribución de las observaciones con respecto al indicador binario en los cuatro conglomerados	214
6.19. Porcentaje de profesores según distintas categorías en cada uno de los cuatro clusters ..	215
6.20. Escenarios de trabajo definidos.....	223
6.21. Métodos GSA basados en el cálculo de varianzas	230
6.22. Análisis de sensibilidad global aplicado al escenario EUPV	237
6.23. Análisis de sensibilidad global aplicado al escenario E1	240
6.24. Análisis de sensibilidad global aplicado al escenario E11	242
6.25. Análisis de sensibilidad aplicado al escenario E7	243
6.26. Análisis de sensibilidad aplicado al escenario E15	245
6.27. Función de densidad de probabilidad asignada a cada fuente de incertidumbre en los escenarios E1, E11, EUPV	247
6.28. Función de densidad de probabilidad asignada a cada fuente de incertidumbre en los escenarios E15, E7	248
6.29. Restricciones en los pesos de cada escenario de trabajo	249
7.1. Estadísticos descriptivos de las dimensiones del IAA	266
7.2. Coeficientes de correlación entre las dimensiones de estudio.....	267
7.3. Efectos principales de las dimensiones del IAA	276
7.4. Resultados del modelo BOD para un conjunto de profesores.....	291
7.5. Resultados de las puntuaciones de los profesores X, Y en el modelo del escenario E2	296
7.6. Profesores X, Y con un IAA_UPV mayor que IAA_E3	299
7.7. Asociación de categorías laborales a los perfiles definidos en el escenario E3	303
7.8. Asociación de categorías laborales a los perfiles definidos en el escenario E4	315
10.1. Análisis de sensibilidad global aplicado al escenario E2	356
10.2. Análisis de sensibilidad global aplicado al escenario E3	356
10.3. Análisis de sensibilidad global aplicado al escenario E4	357
10.4. Análisis de sensibilidad global aplicado al escenario E5	357
10.5. Análisis de sensibilidad global aplicado al escenario E6	358
10.6. Análisis de sensibilidad global aplicado al escenario E8	358
10.7. Análisis de sensibilidad global aplicado al escenario E9	359
10.8. Análisis de sensibilidad global aplicado al escenario E10	359
10.9. Análisis de sensibilidad global aplicado al escenario E12	360
10.10. Análisis de sensibilidad global aplicado al escenario E13	360
10.11. Análisis de sensibilidad global aplicado al escenario E14	361
10.12. Escenarios de trabajo propuestos para la construcción del IAA	371

Índice de figuras

3.1. Evolución del número de ICs construidos.....	37
3.2. Proceso de elaboración de índices.....	44
3.3. Efecto de la imputación con la media no condicionada	61
3.4. Esquema del algoritmo de Imputación Múltiple	65
3.5. Modelo DEA con un <i>input</i> y dos <i>outputs</i>	93
3.6. Modelo de construcción de un IC	107
3.7. Análisis de incertidumbre y sensibilidad de un IC.....	107
3.8. Esquema del procedimiento GSA basado en el cálculo de varianzas	109
3.9. TAI vs. GDP (PPP)	117
3.10. TAI por dimensiones	118
3.11. TAI (Finlandia)	118
3.12. TAI (Finlandia) vs. TAI (EE.UU.) vs. TAI (Valor medio)	119
3.13. Presentación del Índice de Desarrollo Humano en tablas	120
3.14. Presentación del Índice de Consumo mundial de cereales per cápita en gráfico de barras	121
3.15. Presentación del Índice de Mercado Interno en gráfico de líneas	122
3.16. Presentación del Índice de Sostenibilidad desarrollado por el gobierno británico en señales de tráfico.....	122
3.17. Leyenda de las señales de tráfico	122
3.18. Presentación del Índice de Competitividad en forma de ranking.....	123
3.19. Presentación del Índice de Sostenibilidad del grupo SPEar en gráfico de rada	124
4.1. Enfoques de la calidad en la universidad	128
4.2. Nuevas estrategias para la mejora de la ‘calidad universitaria’	131
4.3. Marco conceptual teórico para la definición del sistema básico de indicadores.....	143
4.4. Fases de desarrollo del proyecto INFOACES	144
4.5. El “Profile viewer” para una institución.....	146
5. 1. Dimensiones del modelo de evaluación DOCENTIA.....	163
6.1. Diagrama de flujo del marco conceptual teórico del IAD.....	182
6.2. Representación de un valor atípico a nivel multivariante	187
6.3. Gráfico de sedimentación.....	197
6.4. Porcentaje de varianza acumulado	197
6.5. Gráfico de sectores con la distribución de las observaciones en los seis conglomerados..	205
6.6. Contribución de los indicadores K_2 y $IAD_{INFORMA}$ a la formación de los conglomerados	206
6.7 Contribución de CAPDOC a los clusters	210

6.8. Contribución de IAD_{DOC} a los clusters.....	210
6.9. Contribución de IAD_{EVA} a los clusters.....	210
6.10. Contribución de $IAD_{INFORMA}$ a los clusters.....	210
6.11. Contribución de K_4 a los clusters.....	210
6.12. Contribución de K_1 a los clusters.....	210
6.13. Contribución de IAD_{FORMA} a los clusters.....	210
6.14. Contribución de IAD_{GUIA} a los clusters.....	211
6.15. Contribución de IAD_{MAT} a los clusters.....	211
6.16. Contribución de IAD_{OTRAS} a los clusters.....	211
6.17. Contribución de IAD_{TUT} a los clusters.....	211
6.18. Contribución de K_2 a los clusters.....	211
6.19. Gráfico de sectores con la distribución de las observaciones en los cuatro conglomerados.....	214
6.20. Contribución de IAD_{DOC} a los clusters.....	216
6.21. Contribución de IAD_{EVA} a los clusters.....	216
6.22. Contribución de $IAD_{INFORMA}$ a los clusters.....	216
6.23. Contribución de K_4 a los clusters.....	216
6.24. Contribución de IAD_{FORMA} a los clusters.....	216
6.25. Contribución de K_1 a los clusters.....	216
6.26. Contribución de IAD_{GUIA} a los clusters.....	217
6.27. Contribución de IAD_{MAT} a los clusters.....	217
6.28. Contribución de IAD_{TUT} a los clusters.....	217
6.29. Contribución de IAD_{OTRAS} a los clusters.....	217
6.30. Contribución de K_2 a los clusters.....	217
6.31. Ponderación y agregación aplicada al conjunto de indicadores simples para la obtención del IC.....	220
6.32. Ponderación y agregación aplicada a distintos niveles para la obtención del IC.....	220
6.33. Fuentes de variación que afectan a la construcción del IC.....	226
6.34. Estimación no lineal de IAD_{EVA} frente a E15.....	231
6.35. Efectos principales y pesos objetivo de los indicadores simples del IC_{UPV}	239
6.36. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC_{E1}	241
6.37. Efectos principales y pesos objetivo de los indicadores simples del IC_{E11}	242
6.38. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC_{E7}	244
6.39. Relación entre $f_{IAD_{DOC}}$ (IAD_{DOC}) y IAD_{DOC}	244

6.40. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC _{E15}	246
6.41. Resultados del AI donde se muestra el E15 referencial (cruces azules), la media (marcas rojas) y el correspondiente P ₅ y P ₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores	250
6.42. Resultados del AI donde se muestra el EUPV referencial (cruces azules), la media (marcas rojas) y el correspondiente P ₅ y P ₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores	251
6.43. Diagrama de Hasse aplicado a los resultados del A.I. del escenario E15	254
6.44. Comparación de las posiciones del escenario referencial E15 y las posiciones aplicando Copeland	256
6.45. Histograma de las diferencias en posiciones del escenario referencial E15 y las posiciones aplicando Copeland	257
6.46. Comparación de las posiciones del escenario referencial EUPV y las posiciones aplicando Copeland	257
6.47. Histograma de las diferencias en posiciones del escenario referencial EUPV y las posiciones aplicando Copeland	258
7.1. Histograma del escenario IAA_UPV_N	268
7.2. Distribución del escenario IAA_UPV_N por categoría laboral	269
7.3. Contribución de las dimensiones al IAA por categoría laboral para el escenario de la UPV	269
7.4. Contribuciones de las dimensiones al IAA_UPV en coordenadas baricéntricas	271
7.5. Contribuciones de las dimensiones para profesores con IAA_UPV>0,8 en coordenadas baricéntricas	271
7.6. IAA_UPV por categoría laboral para el profesorado con IAA_UPV>0,8	272
7.7. Contribución de las dimensiones para profesores con IAA_UPV>0,8	272
7.8. Contribuciones de las dimensiones para profesores con IAA_UPV<0,2 en coordenadas baricéntricas	273
7.9. IAA_UPV por categoría laboral para el profesorado con IAA_UPV<0,2	273
7.10. Contribución de las dimensiones por categoría laboral para profesores con IAA_UPV<0,2	274
7.11. Histogramas de las dimensiones IAD, VAIP y AG	274
7.12. Pesos nominales y efectos principales de las dimensiones	276
7.13. Efecto principal de la dimensión IAD a la formación del IAA_UPV	277
7.14. Efecto principal de la dimensión VAIP a la formación del IAA_UPV	277
7.15. Efecto principal de la dimensión AG a la formación del IAA_UPV	277
7.16. Comparación del IAA_E1 y IAA_UPV	291
7.17. Histograma del escenario IAA_E1	292
7.18. Contribución de cada dimensión al IAA_E2 según el experto 1	294

7.19. Contribución de cada dimensión según la opinión del grupo de expertos	294
7.20. Histogramas de los escenarios E1 y E3	298
7.21. Comparación del IAA_E3 y IAA_UPV	298
7.22. Comparación del IAA_E3 y IAA_UPV según la contribución de la docencia.....	299
7.23. Comparación del IAA_E3 y IAA_UPV según la contribución de la investigación.....	300
7.24. Comparación en posiciones de los escenario E3 y UPV	300
7.25. IAA_E3 por categoría laboral	301
7.26. Contribución de las dimensiones al IAA_E3 por categoría laboral	301
7.27. Contribuciones de las dimensiones al IAA_E3 en coordenadas baricéntricas.....	302
7.28. IAA_E3 por perfiles de profesorado	304
7.29. IAA_E3 por categoría laboral para el perfil Gestor	305
7.30. Puntuaciones de las dimensiones para el perfil Gestor en el escenario E3	305
7.31. IAA_E3 por categoría laboral para el perfil Docente-Investigador	306
7.32. Puntuaciones de las dimensiones para el perfil Docente-Investigador en el escenario E3	306
7.33. IAA_E3 por categoría laboral para el perfil Docente.....	307
7.34. Puntuaciones de las dimensiones para el perfil Docente en el escenario E3.....	307
7.35. IAA_E3 por categoría laboral para el perfil Investigador	308
7.36. Puntuaciones de las dimensiones para el perfil Investigador en el escenario E3	308
7.37. Contribuciones de las dimensiones para profesores con $IAA_{E3} < 0,2$ en coordenadas baricéntricas	309
7.38. $IAA_{E3} < 0,2$ por perfil de profesorado	309
7.39. Contribuciones de las dimensiones para profesores con $IAA_{E3} > 0,8$ en coordenadas baricéntricas	310
7.40. $IAA_{E3} > 0,8$ por perfil de profesorado	310
7.41. Comparación del IAA_E4 y IAA_UPV	312
7.42. Comparación en posiciones de los escenarios E4 y UPV	312
7.43. Histograma del escenario E4.....	313
7.44. IAA_E4 por categoría laboral	313
7.45. Contribución a las dimensiones por categoría laboral para el escenario E4	314
7.46. Contribuciones del IAA_E4 en coordenadas baricéntricas	315
7.47. IAA_E4 por perfiles de profesorado	316
7.48. IAA_E4 por categoría laboral para el perfil Gestor	317
7.49. Puntuaciones de las dimensiones para el perfil Gestor en el escenario E4	317
7.50. IAA_E4 por categoría laboral para el perfil Docente-Investigador	318
7.51. Puntuaciones de las dimensiones para el perfil Docente-Investigador en el escenario E4	318
7.52. IAA_E4 por categoría laboral para el perfil Docente.....	319

7.53. Puntuaciones de las dimensiones para el perfil Docente en el escenario E4.....	319
7.54. IAA_E4 por categoría laboral para el perfil Investigador	320
7.55. Puntuaciones de las dimensiones para el perfil Investigador en el escenario E4	320
7.56. Contribuciones de las dimensiones para los profesores con IAA_E4<0,3 en coordenadas baricéntricas	321
7.57. IAA_E4<0,3 por perfil de profesorado	321
7.58. Contribuciones de las dimensiones para profesores con IAA_E4>0,8 en coordenadas baricéntricas	322
7.59. IAA_E4>0,8 por perfil de profesorado	322
7.60. Visualización de los resultados del IAA_E4 para ID 2659, ID 359, ID 1947.....	325
7.61 Visualización de los resultados del IAA_E4 para ID 1743, ID 160, ID1	326
10.1. Resultados del AI donde se muestra el E1 referencial (cruces azules), la media (marcas rojas) y el correspondiente P ₅ y P ₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores	362
10.2. Comparación de las posiciones del escenario referencial E1 y las posiciones aplicando Copeland	362
10.3. Histograma de las diferencias en posiciones del escenario referencial E1 y las posiciones aplicando Copeland.....	363
10.4. Resultados del AI donde se muestra el E7 referencial (cruces azules), la media (marcas rojas) y el correspondiente P ₅ y P ₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores	363
10.5. Comparación de las posiciones del escenario referencial E7 y las posiciones aplicando Copeland	364
10.6. Histograma de las diferencias en posiciones del escenario referencial E7 y las posiciones aplicando Copeland.....	364
10.7. Resultados del AI donde se muestra el E11 referencial (cruces azules), la media (marcas rojas) y el correspondiente P ₅ y P ₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores	365
10.8. Comparación de las posiciones del escenario referencial E11 y las posiciones aplicando Copeland	365
10.9. Histograma de las diferencias en posiciones del escenario referencial E11 y las posiciones aplicando Copeland.....	366

Lista de abreviaturas

AG	Actividades de Gestión
AHP	Analytic Hierarchy Process
ANECA	Agencia Nacional de Evaluación de la Calidad y Acreditación
AY	Ayudante
AYD	Ayudante Doctor
BA	Budget Allocation
BOD	Benefit of the Doubt
CA	Conjoint Analysis
CED	Comisión para la Evaluación de la Docencia
CEU	Catedrático de Escuela Universitaria
COD	Colaborador Doctor
COL	Colaborador
CRUE	Conferencia de Rectores de Universidades Españolas
CU	Catedrático de Universidad
DEA	Data Envelopment Analysis
EEES	Espacio Europeo de Educación Superior
EM	Expectación-Maximización
EUROSTAT	Statistical Office of the European Communities
EW	Equal Weighting
FA	Factorial Analysis
GRW	Generalized Random Walk
GSA	Global Sensitivity Analysis
HDMR	High Dimensional Model Representation
IAA	Índice de Actividad Académica
IAD	Índice de Actividad Docente
IAI	Índice de Actividad Investigadora
IC	Indicador Compuesto
ICs	Indicadores Compuestos
IE	Instituto de Evaluación
IES	Instituciones de Educación Superior
INE	Instituto Nacional de Estadística
IRW	Integrated Random Walk
JRC	Joint Research Centre
MAR	Missing at Random
MCAR	Missing Completely at Random
NMAR	Not Missing at Random

NOD	Normativa de Organización Docente
OCDE	Organización para la Cooperación y del Desarrollo Económico
PCA	Principal Component Analysis
PDI	Personal Docente Investigador
PDF	Probability Density Function
SDP	State Dependent Parameter
SDR	State Dependent Regression
SEE	Sistema Estadístico Europeo
TAI	Technology Achievement Index
TEU	Titular de Escuela Universitaria
TU	Titular de Universidad
UCM	Unobserved Components Model
UE	Unión Europea
UIS	UNESCO Institute for Statistics
UPV	Universitat Politècnica de València
VAIP	Valoración de la Actividad Investigadora Personalizada

Capítulo 1

Introducción

En las últimas décadas se ha discutido abiertamente sobre el uso de indicadores con el fin de orientar y justificar la toma de decisiones en la enseñanza de educación superior. Hay quienes están absolutamente en contra del uso de los indicadores como herramienta de medición de aspectos relacionados con la educación superior y, por el contrario, quienes están a favor de su uso para la toma de decisiones educativas. Pero lo más común es encontrar, en estudios relacionados con indicadores, reflexiones en las que se hacen advertencias sobre el uso que se les debe dar y sobre las precauciones que se deben tener (Higher Education Quality Council, 1996).

Algunas de las razones de esta controversia son la discrepancia en la definición de unos objetivos claros sobre el uso que se le va a dar al indicador, la dificultad en la definición de indicadores diseñados para la posterior comparación entre instituciones o la preocupación de convertirlos en una meta por sí mismos cuando simplemente son una manifestación parcial de una realidad compleja (Vroeijenstijn, 1995). Otro de los argumentos que se suele utilizar en contra del uso de indicadores, que es una de las razones por las que se ha avanzado tan poco en el desarrollo de información sobre las universidades, es el deber de proteger la solidez de unas instituciones centenarias de las modas o rápidas fluctuaciones del mundo exterior (Mora, 1999a).

A pesar de los muchos inconvenientes que presentan por encontrarse todavía en fase de perfeccionamiento e incluso de creación, lo cierto es que facilitan la toma de decisiones, aportan información relevante acerca de fenómenos educativos y permiten analizar su evolución en el tiempo, siempre y cuando se definan claramente los objetivos de su uso y vayan acompañados de juicios y estimaciones cualitativas bien argumentados y centrados en la singularidad de la institución, programa o persona evaluados.

Según Mora (1999a), el uso de indicadores como herramienta de medición puede ayudar considerablemente a establecer un cierto orden y a hacer previsiones en un sistema tan complejo como es en la actualidad cualquier institución universitaria. No cabe duda que cada vez es más común su uso en el ámbito de la educación superior, tanto para el funcionamiento a nivel interno de las universidades como para la realización de comparaciones entre instituciones o países (ARWU, 2003; QS, 2005; SCImago, 2007; Aguillo, Ortega y Fernández, 2008; Torres-Salinas et al., 2011; Instituto de Evaluación, 2011a; Instituto de Evaluación, 2011b). Como

consecuencia, se han aumentado los recursos para su elaboración y se ha mejorado la metodología empleada para su construcción. Así pues, son muchas las organizaciones dedicadas a la definición de indicadores de educación y a la recopilación de datos, como por ejemplo la OCDE, EUROSTAT, el Instituto de Estadística de la UNESCO (UIS), el Sistema Estadístico Europeo (SEE) que incluye el Instituto Nacional de Estadística (INE) de todos los estados miembros, el Instituto de Evaluación (IE) del Ministerio de Educación, Cultura y Deporte o la CRUE.

Una de las conclusiones a las que se ha llegado sobre el uso de indicadores es que por sí solos son una herramienta limitada, pero utilizados dentro de los actuales procesos de calidad en las universidades pueden convertirse en una herramienta auxiliar y valiosa para la mejora, además de servir como medio de información para las administraciones públicas y la sociedad sobre el funcionamiento interno de las instituciones de educación superior (rendición de cuentas) (Mora, 1999a). Desde el mismo punto de vista, Olivera (1997) afirma que no se debe magnificar su utilidad ni condenarlos, sino utilizarlos como instrumentos que faciliten el análisis para la medición de complejos fenómenos educativos y su planificación.

Sin embargo, muy pocas veces un indicador proporciona por sí solo información suficiente para la comprensión de esos fenómenos educativos tan complejos. Incluso muchas veces resulta complicado interpretar una tendencia común en un sistema de indicadores simples para la formulación y el análisis de políticas públicas. Es por eso que tanto los tomadores de decisiones como los grupos de interés y los medios de comunicación han manifestado, durante los últimos años, un interés creciente por diseñar mecanismos o herramientas útiles que, aplicados sobre un sistema de indicadores, permitan evaluar conceptos complejos en diferentes áreas de estudio con el objetivo de hacer comparaciones simples entre las distintas unidades de análisis medidas. Dichas herramientas de medición desarrolladas se conocen como indicadores compuestos o índices, y son la combinación o agregación matemática de un conjunto de indicadores simples con el objetivo de resumir un concepto multidimensional en un índice simple o unidimensional en base a un modelo conceptual subyacente (Nardo et al., 2008).

La mayoría de estudios realizados hasta ahora sobre indicadores compuestos miden conceptos multidimensionales sobre un grupo de países para una posterior comparación entre ellos. De esta forma, parece más fácil para el público en general interpretar los indicadores compuestos que identificar tendencias comunes en diferentes indicadores simples. El número existente de indicadores compuestos crece año tras año y se aplica en diferentes ámbitos de interés. La proliferación de estas herramientas de medición es un claro síntoma de su importancia política y de su relevancia operativa en la toma de decisiones (Granger, 2001). Para hacerse una idea, la búsqueda en “Google Scholar” de la palabra “Composite Indicators” en septiembre de 2013 dio un resultado de 9.270 entradas, mientras que en octubre de 2005 fueron 992. La mayoría de ellos se han desarrollado en los últimos años, siendo la UE y la OCDE las organizaciones pioneras en el desarrollo de iniciativas relacionadas con estos conceptos generando una colección de documentación que sirve de punto de partida para su estudio (Nardo et al., 2008). Así mismo, organizaciones como Naciones Unidas y la Comisión Europea han

desarrollado indicadores compuestos de gran interés para la sociedad en general (European Commission-DG ECFIN, 2000; United Nations Development Programme, 2001; European Commission-DG MARKT, 2001; European Commission-DG ENTR, 2001a; European Commission-DG ENTR, 2001b; European Commission-DG ECFIN, 2004; European Commission-DG ENTR, 2004; Annoni y Kozovska, 2010).

La construcción de indicadores compuestos suele realizarse en múltiples áreas de la gestión pública tales como la economía y sus diversos sectores (industria, agricultura, servicios, etc.), el desarrollo social y el análisis integrado del medio ambiente y su interacción con el desarrollo económico y social. También suelen aplicarse en la formulación de políticas de promoción de la innovación y la investigación científica (Griliches, 1990; Cox et al., 1992; Färe et al., 1994; Lovell, Pastor y Turner, 1995; Cribari-Neto, Jensen y Novo, 1999; Forni et al., 2001; Saisana y Tarantola, 2002; Wilson y Jones, 2002; Huggins, 2003; Grupp y Mogege, 2004; Munda, 2005; Nardo et al., 2008).

En el contexto de la educación superior los indicadores compuestos más conocidos son los famosos *rankings* de universidades, cuyo objetivo principal es realizar comparaciones entre instituciones o países (ARWU, 2003; QS, 2005; SCImago, 2007; Saisana, 2008; Aguillo, Ortega y Fernández, 2008; Torres-Salinas et al., 2011). Estos indicadores compuestos surgen como guías para la ayuda en la toma de decisiones por parte de los potenciales usuarios y los estudiantes. Además, también pueden ser utilizados como herramientas de *benchmarking* por parte de las propias instituciones de educación superior. Adicionalmente a los *rankings*, también se han construido indicadores compuestos a nivel interno de funcionamiento universitario (Pulido y Pérez, 2003; Universitat Politècnica de València, 2008; Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV, 2010; Ramón, Ruiz y Sirvent, 2010; Bana e Costa y Oliveira, 2011).

La construcción cada vez mayor de indicadores compuestos en el contexto de educación superior puede ser debida a los cambios de estrategias que exige el paso de una universidad tradicional a una universidad al servicio del conocimiento y, por tanto, a los cambios en las relaciones con la administración, en el sistema de rendición de cuentas, en las relaciones con la sociedad y en la elaboración de procesos internos de mejora (Mora, 1999b).

El principal objetivo del proceso de convergencia de la enseñanza de educación superior es convertir a los sistemas universitarios europeos en un referente internacional por la calidad de la enseñanza impartida (Medina, 2005). Son numerosas las iniciativas ya aplicadas en nuestro país que inculcan la “filosofía de la calidad” de las actividades universitarias medidas a través de la evaluación (Calderón y Escalera, 2008). Ejemplos de esas iniciativas son el Programa Experimental de Evaluación, el proyecto Piloto Europeo de la Evaluación de la Enseñanza, los procesos de evaluación interna y externa que se llevaron a cabo en el marco del Plan Nacional de Evaluación de la Calidad de las Universidades y en el II Plan de la Calidad de las Universidades, y las actividades de evaluación llevadas a cabo por ANECA (el programa de

evaluación institucional, el programa de certificación, el programa de Convergencia Europea y el programa de evaluación del profesorado).

La universidad española se plantea alcanzar muchos retos relevantes con el fin de conseguir una mayor calidad en el sistema universitario español y competitividad en relación con el resto de universidades europeas. Este trabajo de investigación se centra en uno de los retos en particular, la evaluación del profesorado universitario. La transformación surgida como consecuencia de la implantación de las directrices del Espacio Europeo de Educación Superior tiene una influencia directa en la actividad docente del profesorado, observándose la importancia cada vez mayor de las actividades de evaluación que aseguren la calidad de los procesos formativos que afecten a los estudiantes, a la comunidad universitaria y a la sociedad en general (European Commission Education and Training, 2009). Los resultados de la evaluación permitirán a las universidades el diseño y la ejecución de planes de mejora de la docencia y la investigación. Por tanto, es importante disponer de herramientas de medición que permitan evaluar la actividad del profesorado universitario en sus tres dimensiones: docencia, investigación y gestión.

Esta continua demanda del uso de indicadores en la educación superior se ha visto incrementada por el desarrollo de las nuevas tecnologías que han generado nuevas posibilidades en el manejo de datos con gran cantidad de información cuantitativa sobre la que es importante organizar y sintetizar la información de forma rápida y eficiente (Vidal, 1999). Para concluir con decisiones finales ricas en información sobre los indicadores compuestos definidos es necesario combinar la inevitable opinión de los expertos como fuente principal de información para los tomadores de decisiones y los procedimientos estadísticos que justifican la correcta construcción.

Sin embargo, el descuido metodológico durante el diseño del índice, la falta de información y el uso de datos inconsistentes pueden dar lugar a conclusiones demasiado simplificadas. La construcción de un indicador compuesto requiere de etapas subjetivas en las que se emiten juicios de valor para la elección de los diferentes supuestos. Ejemplos de estas etapas son: la elección de los indicadores simples, la imputación de datos faltantes, la ponderación de cada indicador y dimensión, la elección del método de agregación, etc. Dichas elecciones subjetivas forman la base de su construcción, por lo que es imprescindible evaluar la calidad y transparencia del índice mediante técnicas de análisis de sensibilidad y de incertidumbre.

En conclusión, la Tesis Doctoral tiene una relación directa con dos de los objetivos estratégicos a los que la universidad española debería prestar una especial atención, puesto que hoy por hoy constituyen uno de los puntos más débiles del sistema universitario. Estos objetivos son la mejora de la calidad de los procesos de enseñanza-aprendizaje y la transparencia informativa sobre lo que las universidades ofrecen. Este último aspecto es sin duda una herramienta esencial para la rendición de cuentas de las universidades tanto públicas como privadas (Mora y Vilalta, 2011). Por ese motivo, las universidades necesitan disponer de herramientas de medición que les permitan medir y evaluar los diversos fenómenos complejos

de la enseñanza de educación superior. De esta forma, pueden disponer de información útil y fiable que ayude en la toma de decisiones y que permita innovar y mejorar en la calidad de los procesos educativos. En este contexto, el trabajo de investigación aporta una mejora en el diseño de estrategias metodológicas para la construcción de indicadores compuestos como herramientas de medición en la gestión universitaria y, en particular, en la evaluación del profesorado universitario, basadas en técnicas estadísticas y analizadas posteriormente mediante técnicas de análisis de sensibilidad y técnicas de incertidumbre para darle sustento técnico, fiabilidad y validez a la construcción del indicador compuesto.

Capítulo 2

Objetivos

El objetivo principal de la Tesis Doctoral es diseñar estrategias metodológicas para la construcción de indicadores compuestos que permitan medir conceptos multidimensionales relacionados con la gestión de la educación superior. La finalidad es proporcionar herramientas que contribuyan a aumentar la transparencia y calidad de los procesos de evaluación de las actividades del profesorado universitario.

El objetivo general del trabajo se concreta en los siguientes objetivos específicos:

1. Adaptar las directrices recogidas en los principios teóricos generales desarrollados en el campo de la construcción de indicadores compuestos para la evaluación de conceptos multidimensionales en el campo de la educación superior.
2. Estudiar la aplicación de técnicas estadísticas en cada una de las etapas de construcción de un indicador compuesto que mida la evaluación docente del profesorado, siguiendo los objetivos estratégicos de la Universitat Politècnica de València (UPV).
3. Validar la calidad y robustez de las metodologías propuestas aplicando un análisis de sensibilidad y de incertidumbre, y proponer diferentes técnicas de visualización de los resultados de este último.
4. Diseñar un modelo de evaluación de la actividad académica del profesorado universitario de la UPV que permita medir la eficacia relativa de cada profesor con respecto al resto teniendo en cuenta su perfil docente, investigador y gestor.

ESTRUCTURA DEL TRABAJO

La Tesis Doctoral se ha estructurado en nueve capítulos y cinco anexos como se describe a continuación.

En el capítulo 3 se profundiza sobre el concepto de indicador compuesto ya que es la herramienta de medición empleada en este estudio. Desde una perspectiva metodológica, se realiza una revisión exhaustiva de las posibles técnicas estadísticas que se pueden aplicar en cada una de las etapas de la construcción de un indicador compuesto. En estas etapas se incluyen el tipo de normalización de los datos, el tipo de ponderación y agregación de los indicadores simples y, finalmente, el tipo de análisis de sensibilidad y de incertidumbre que se puede aplicar para evaluar la calidad y robustez del indicador compuesto construido.

En el capítulo 4 se estudia el papel de los indicadores en el contexto de la educación superior. Asimismo, se describen proyectos de investigación relevantes en los que se emplean los indicadores como herramientas de medición en las instituciones de educación superior. Estos son el proyecto “European indicators and ranking methodology for university third mission”, “Sistema integral de información sobre las instituciones de educación superior de América Latina para el área común de educación con Europa”, “The european classification of higher education institutions” y el proyecto “Multi-dimensional global ranking of universities, a feasibility project”.

En el capítulo 5 se presentan dos modelos de evaluación del profesorado propuestos por la Universitat Politècnica de València. En primer lugar, se describe el modelo de construcción del Índice de Actividad Docente que evalúa, como su nombre indica, las actividades docentes realizadas por el profesorado. Seguidamente, se expone la metodología de construcción del Índice de Actividad Académica que recoge las tres actividades principales del profesorado universitario: docencia, investigación y gestión académica.

El objetivo de estos tres primeros capítulos es revisar las distintas metodologías existentes para la construcción de indicadores compuestos y mostrar la importancia y utilidad de los indicadores como herramienta de evaluación académica y de rendición de cuentas.

En el capítulo 6 se realiza un estudio metodológico para la construcción de un indicador compuesto que evalúe la actividad docente del profesorado de la UPV. El procedimiento de construcción se divide en etapas en las que se aplican las técnicas de análisis más adecuadas en base a los datos disponibles para el estudio. Por último, se validan las metodologías empleadas mediante la aplicación de técnicas de análisis de sensibilidad y de incertidumbre. El objetivo de este capítulo es mostrar la importancia que tiene la aplicación de técnicas estadísticas en el procedimiento de construcción de un indicador compuesto empleado para la medición de actividades realizadas por el profesorado.

En el capítulo 7 se propone una metodología alternativa a la propuesta por la UPV para la medición de la actividad académica del profesorado. Esta metodología se conoce como *Benefit of the Doubt (BOD)* y deriva de la técnica *Data Envelopment Analysis (DEA)*. Primeramente se describe, de forma teórica, la metodología empleada así como las ventajas y limitaciones de su aplicación, y se resume el estado del arte de la aplicación de esta técnica en el ámbito académico universitario. Seguidamente, se proponen escenarios de trabajo alternativos en los que se emplea la metodología BOD y, finalmente, se valida y justifica la metodología empleada para la evaluación de la actividad académica del profesorado de la UPV. El objetivo de este capítulo es diseñar modelos de indicadores compuestos para la evaluación de la actividad académica del profesorado que tengan en cuenta los distintos perfiles docentes.

En el capítulo 8 se exponen las conclusiones obtenidas como resultado del trabajo de investigación desarrollado en la presente Tesis Doctoral, así como las futuras líneas de investigación a desarrollar después de este estudio.

Finalmente en el capítulo 9 se muestra la bibliografía consultada para la redacción de la memoria y en los anexos del capítulo 10 se adjunta información adicional sobre técnicas aplicadas en el estudio y sobre resultados obtenidos de forma más detallada.

Capítulo 3

Metodología y proceso de construcción de un Indicador Compuesto

3.1. Introducción a los Indicadores Compuestos

Los Indicadores Compuestos (ICs), también conocidos como índices, resumen la información contenida en los sistemas de indicadores. Además, se consideran herramientas eficaces que contribuyen a la formulación y al análisis de políticas públicas, así como a su evaluación y comunicación. Por su capacidad de síntesis, permiten atraer la atención del público dando lugar a la creación de argumentos convincentes y contribuyendo en los debates de políticas integradas que promueven el tema de interés.

La mayoría de estudios sobre ICs realizados hasta el momento miden conceptos multidimensionales sobre un grupo de países con el objetivo de realizar una posterior comparación entre ellos. De esta forma, parece más fácil para el público en general interpretar los ICs que identificar tendencias comunes en diferentes indicadores simples.

El número existente de ICs crece año tras año y se aplica a diferentes campos de estudio (revisión reciente que cita más de 160 ICs (Bandura y Martin, 2006)).

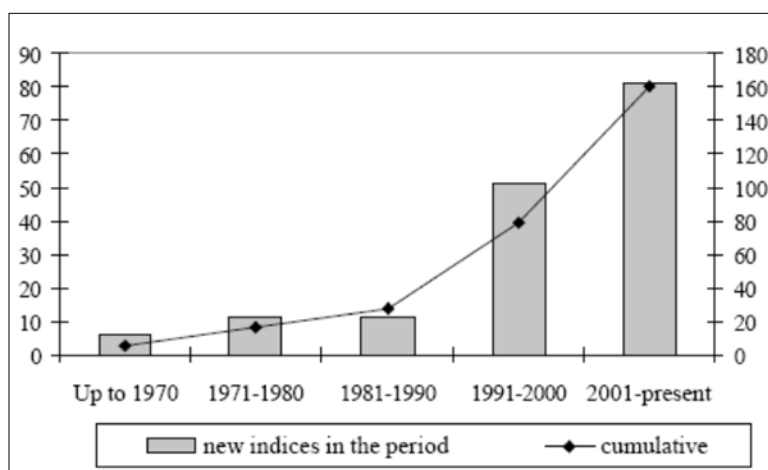


Figura 3.1. Evolución del número de ICs construidos

La proliferación de los ICs es un claro síntoma de su importancia política y de su relevancia operativa en la toma de decisiones (Granger, 2001). Para hacerse una idea del crecimiento de ICs desarrollados la búsqueda en “Google Scholar” de la palabra “Composite Indicators” dio un resultado de 9.270 entradas en septiembre de 2013, mientras que en octubre de 2005 fueron 992. La mayoría de ellos se han desarrollado en los últimos cuatro años (Joint Research Centre, 2010).

La Unión Europea y la Organización para la Cooperación y el Desarrollo Económico (OCDE) han sido organizaciones pioneras en el desarrollo de iniciativas relacionadas con estos conceptos en diferentes ámbitos de estudio: innovación/tecnología, sociedad, globalización, medio ambiente, economía,...; generando una colección de documentación que sirve de punto de partida para su estudio (Nardo et al., 2008). Asimismo, organizaciones como Naciones Unidas y la Comisión Europea han desarrollado ICs de gran interés por parte del público en general (European Commission-DG ECFIN, 2000; United Nations Development Programme, 2001; European Commission-DG MARKT, 2001; European Commission-DG ENTR, 2001a; European Commission-DG ENTR, 2001b; European Commission-DG ECFIN, 2004; European Commission-DG ENTR, 2004).

En el contexto de análisis de políticas, tanto a nivel nacional como internacional, los ICs suelen utilizarse para establecer ejercicios de *benchmarking* acerca del desempeño de las unidades de análisis con respecto a una unidad de análisis referencial (Huggins, 2003). De esta forma, el uso de esta herramienta de medición facilita la interpretación de los escenarios que deben considerar los tomadores de decisiones, permite identificar tendencias generales y establecer una serie de prioridades políticas (Freudenberg, 2003; Saltelli, 2007). Por otra parte, algunos de los ICs diseñados hasta el momento tienen como objetivo general la construcción de un *ranking* de unidades de análisis y su evaluación comparativa de acuerdo con algunas de las dimensiones globales (Cherchye, 2001; Kleinknecht, Van Montfort y Brouwer, 2002).

En el proceso de definición de un IC existen tres niveles de agrupación de indicadores (Freudenberg, 2003):

1. El **conjunto de indicadores simples** o indicadores individuales. Suele ser visto como un primer paso en la acumulación de información cuantitativa ya existente
2. Los **indicadores temáticos** o también llamados dimensiones. Son indicadores simples o individuales agrupados según un tema o área específica. Este procedimiento requiere identificar conjuntos de indicadores conectados o relacionados conceptualmente, o según propiedades estadísticas similares
3. Los **ICs**. Se generan al combinar los indicadores temáticos y presentarlos como una medida compuesta única

Sin embargo, el descuido metodológico durante su diseño, la falta de información y el uso de datos inconsistentes puede dar lugar a conclusiones, si no erróneas, demasiado simplificadas. Su “imagen general” resultante invita a los usuarios, en especial a los tomadores de decisiones, a obtener conclusiones muy simplistas. No obstante, los ICs deben ser vistos como un medio para iniciar el debate y fomentar el interés público, y no como una única herramienta a partir de la cual se obtienen resultados definitivos e indiscutibles.

Las principales críticas de la construcción de un IC se centran en aquellas etapas en las que se deben hacer juicios de valor: la selección de los indicadores simples, el tratamiento de los datos, la elección del modelo de agregación, la asignación de pesos a los indicadores, etc. Las diferentes elecciones metodológicas pueden dar lugar a una manipulación en los resultados. Como consecuencia, la mejora en la construcción y utilización de los ICs parece ser un tema actual de investigación muy importante, tanto desde el punto de vista teórico como operativo (European Commission, 2001; Paruolo, Saltelli y Saisana, 2013).

3.2. Consideraciones preliminares

3.2.1. Definición y objetivo de un indicador

Uno de los principios básicos de la gestión de la calidad de un proceso, institución, producto o servicio es el enfoque objetivo hacia la toma de decisiones. Esto significa que las decisiones deben basarse en el análisis de datos y en la información disponible. Para obtener información relevante de manera eficiente es necesario disponer de un conjunto de datos medido de forma sistemática, periódica y precisa, y organizado de manera que facilite la lectura y el análisis a los distintos miembros de una organización. A este conjunto de datos es lo que se conoce como un sistema de indicadores.

Existen numerosas definiciones del concepto indicador. Según Chevalier, Choiniere y Bernier (1992) un indicador es una “*medida de la parte observable de un fenómeno que permite valorar otra porción no observable de dicho fenómeno*”. Por otra parte, según Ott (1978), un indicador resulta ser la “*forma más simple de reducción de una gran cantidad de datos, manteniendo la información esencial para las cuestiones planteadas a los datos*”.

A menudo son un compromiso entre la exactitud científica y la información disponible a un precio razonable. La aproximación de Gallopín (1996) está inclinada hacia la definición de indicador desde la visión de la Teoría de Sistemas. Este autor define los indicadores como variables (y no valores), es decir, “*representaciones operativas de un atributo, las cuales son definidas en términos de un procedimiento de medida u observación determinado. Cada variable puede asociarse a una serie de valores o estados a través de los cuales se manifiesta*”.

Según Sizer, Spee y Bormans (1992), un indicador es una *“afirmación cuantificada sobre los recursos utilizados o los resultados obtenidos en áreas relacionadas con los objetivos específicos en una determinada empresa”*.

Una de las definiciones más utilizadas en la literatura es la que proporciona Bauer (1996). Este autor señala que los indicadores son *“estadísticas, series estadísticas o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto”*.

La norma UNE 66175 (AENOR, 2003) define un indicador como un *“dato o conjunto de datos que ayudan a medir objetivamente la evolución de un proceso o de una actividad”*. También lo define como *“un parámetro que permite evaluar de forma cuantitativa la eficacia y/o eficiencia de los procesos”*.

De forma matemática se define un indicador simple como una función de una o más variables que conjuntamente miden una característica o atributo de los individuos de estudio (Schuschny y Soto, 2009). Por otra parte, Saltelli y Tarantola (2002) definen los indicadores como piezas de información que resumen las características de un sistema o subrayan lo que está pasando en el sistema.

El uso de indicadores permite llevar a cabo un seguimiento de los objetivos propuestos midiendo su grado de cumplimiento y su calidad. Además, permite resumir extensos conjuntos de datos y, enfocándolos en una dimensión temporal, permite observar tendencias de la situación de un país, región, empresa, investigación, producto o servicio en lo referente a su economía, calidad, desarrollo, etc. Se trata de herramientas muy útiles que ayudan a entender la situación de los problemas, a analizar la evolución en el tiempo y a hacer previsiones futuras. Además, una de sus funciones más importantes es la comparabilidad entre las observaciones, bien sea de la calidad de servicios o productos, del desarrollo económico y social de países, etc.

La confección de un indicador no es una tarea fácil y en su proceso intervienen muchas etapas a desarrollar. Así, cuando se define un indicador se le debe asignar un nombre y establecer el objetivo para el que se crea. Además, dicho indicador estará enclavado dentro de una tipología concreta, ya que los indicadores se pueden clasificar según multitud de variables.

En el proceso de definición de un indicador debe especificarse de forma precisa la unidad de medida, los elementos que lo componen, el origen de los datos y el significado que se le debe dar al resultado numérico final. La periodicidad de medida prevista y la forma de cálculo serán también importantes peculiaridades de cada indicador. Por último, será necesario que quede especificada cual va a ser la unidad responsable de la medición y evaluación del indicador.

3.2.2. Escalas de medida de un indicador

El análisis de datos implica la separación, identificación y medida de la variación en un conjunto de indicadores. La medida es importante para representar con precisión el concepto de interés y es crucial en la selección de posteriores métodos estadísticos (Hair et al., 2007).

Cabe destacar que existen múltiples clasificaciones de indicadores a menudo confluyentes. Una clasificación muy utilizada es la que distingue entre indicadores cuantitativos y cualitativos:

- **Indicador de carácter cuantitativo o métrico.** Los indicadores de datos métricos están constituidos de tal forma que las unidades de análisis pueden ser identificadas por diferencias entre grados o cantidades. Las medidas métricas son las más apropiadas para casos que involucran cantidades o magnitudes. Pueden ser:
 - **Numérico absoluto:** cuando las unidades del indicador derivan directamente de su significado
 - **Numérico relativo:** cuando se consideran ratios o porcentajes de diferentes variables
- **Indicador de carácter cualitativo o no métrico.** Los datos no métricos son atributos, características o propiedades categóricas que identifican o describen a un sujeto. Describen diferencias en tipo o clase indicando la presencia o ausencia de una característica o propiedad. Muchas propiedades son discretas porque tienen una característica peculiar que excluye todas las demás características. Las medidas no métricas pueden tener:
 - **Escalas nominales.** La medida con una escala nominal asigna números que se usan para etiquetar o identificar sujetos u objetos. Las escalas nominales, también conocidas como escalas de categoría, proporcionan el número de ocurrencias en cada clase o categoría de la variable que se está estudiando. Por tanto, los números o símbolos asignados a los objetos no tienen más significado cuantitativo que indicar la presencia o ausencia del atributo o característica bajo investigación
 - **Escalas ordinales.** Las escalas ordinales representan un nivel superior de precisión de medida. Las variables pueden ser ordenadas o clasificadas con escalas ordinales en relación a la cantidad del atributo poseído. Cada subclase puede ser comparada con otra en términos de una relación de “mayor que” o “menor que”

3.2.3. Tipos de indicadores

Existe una amplia literatura sobre tipología, características y propiedades de los diferentes indicadores según el área de estudio y el objetivo perseguido. A continuación se mencionan cuatro tipos de indicadores:

- **Indicadores de Actividad:** indicadores formulados en términos absolutos que se concentran en Qué actividades se desarrollan y en Qué cantidad
- **Indicadores de Rendimiento/Eficiencia:** indicadores que miden el nivel de ejecución del proceso, se concentran en el Cómo se hicieron las cosas y miden el rendimiento de los recursos utilizados por un proceso. En este grupo se consideran solamente indicadores objetivos
- **Indicadores de Percepción:** indicadores basados en juicios de valor que reflejan percepciones y opiniones de grupos de interés sobre una institución o su entorno
- **Indicadores de Impacto:** indicadores que miden las consecuencias derivadas de las actividades que reflejan

3.2.4. Propiedades que debe cumplir un indicador

Para que los indicadores sean utilizados en plena confianza y resulten de gran utilidad se deben establecer unos criterios que deben reunir todos ellos.

Son muchos los autores que han propuesto criterios de selección de indicadores. Sin embargo, parece haber un acuerdo general en el cumplimiento de al menos las propiedades siguientes (Carot et al., 2012):

- **Pertinencia:** un indicador debe reflejar los rasgos distintivos que identifican y caracterizan el fenómeno que pretende describir
- **Validez:** de forma resumida, se dice que hay validez cuando el indicador mide lo que dice medir
- **Fiabilidad:** los indicadores deben producir medidas estables y replicables
- **Comparabilidad:** entre mediciones realizadas a objetos distintos o en el mismo objeto en distintos momentos
- **Comunicabilidad:** los indicadores deben ser fácilmente entendibles por los distintos grupos de interés implicados
- **Resistencia a la manipulación:** los indicadores deben ser resistentes a la manipulación, tanto en el diseño como en la interpretación final de éstos. Si un indicador puede ser fácilmente manipulado por el individuo u organismo evaluado, su validez se reduce

- **Factibilidad:** los indicadores deben poderse calcular en distintos momentos y para distintos objetos
- **Economía:** tanto en la recogida de datos o generación como en el procesamiento
- **Vinculación con los objetivos:** cada tipo de objetivo debe tener asociado claramente sus indicadores y éstos deben ser los adecuados a su naturaleza

3.2.5. Propiedades de un sistema de indicadores

Habitualmente los indicadores no se manejan de forma individual, sino formando parte de sistemas estructurados de indicadores que permitan medir el grado de cumplimiento de los objetivos planteados por cualquier organización. Es deseable que estos sistemas satisfagan, a su vez, determinadas propiedades adicionales que debe cumplir cada uno de los indicadores.

Las características deseables son múltiples y muy variadas. A continuación se describen algunas de ellas:

- **Completitud:** el sistema de indicadores en su conjunto debe reflejar la totalidad del objeto de la medición, en el que cada indicador representa parcialmente un aspecto de ese objeto
- **Pertinencia:** del mismo modo que cada indicador debe ser pertinente en el aspecto concreto que pretende medir, el sistema en su totalidad debe serlo también
- **Simplicidad:** además de ser completo y pertinente, el sistema también debe ser lo más reducido posible en número de indicadores y éstos a su vez lo más simples posible
- **Precisión:** el conjunto de indicadores debe estar definido sin dar lugar a ambigüedad, de modo que pueda ser medido o interpretado por cualquier usuario
- **Utilidad:** el sistema de indicadores debe responder a los intereses de los distintos grupos implicados proporcionando una herramienta útil de apoyo a la toma de decisiones

Además, los sistemas de indicadores deben ser aceptados por los agentes implicados, por lo que es muy importante la participación de todos ellos en el diseño.

3.3. Definición y objetivo de un Indicador Compuesto

Normalmente se distingue entre indicadores simples e ICs (o índices). En términos técnicos los primeros, como se ha visto en la sección anterior, se definen como una función de una o más variables que conjuntamente miden una característica o atributo de las unidades de análisis del estudio. Para efectos del presente documento, se denotará como IC al que se construye como función de dos o más variables, en cuyo caso se están midiendo características multidimensionales (p. ej. calidad ambiental, salud humana, sostenibilidad, calidad educativa, etc.). No obstante, esta jerarquía entre indicadores no puede tomarse como una regla general, pues en muchos casos indicadores simples son utilizados como índices para la toma de decisiones (Gallopín, 1997).

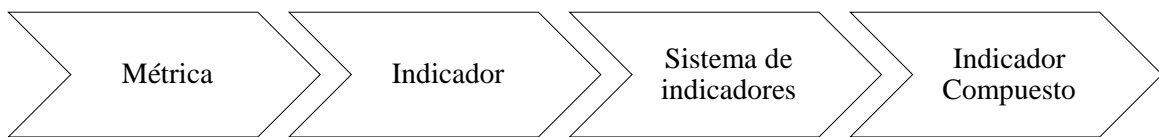


Figura 3.2. Proceso de elaboración de índices

De acuerdo con Nardo et al. (2008), un IC es la combinación (agregación) matemática de un conjunto de indicadores simples con el objetivo de resumir un concepto multidimensional en un índice simple (unidimensional) en base a un modelo conceptual subyacente. Puede ser de carácter cuantitativo o cualitativo, según los requerimientos del analista.

De igual forma, Saisana y Tarantola (2002) definen los ICs como combinaciones matemáticas de indicadores simples que no tienen una unidad de medida común.

Otra definición muy conocida en la literatura señala que un IC es algo que proporciona una pista sobre un asunto de mayor importancia o hace perceptible una tendencia o fenómeno que no es detectable a simple vista (Hammond et al., 1995).

El principal objetivo de un IC es cuantificar y simplificar la información del conjunto de indicadores simples, de forma que se transmita la comprensión del tema que se está analizando tanto a las partes interesadas como al público en general.

La divergencia de opiniones sobre el uso de indicadores aumenta al mismo ritmo que la evidencia del papel de éstos en la vida pública. De entre todos los indicadores, los ICs han experimentado un aumento en su popularidad en los últimos años debido, quizás, a su capacidad de explicar conceptos complejos (Saisana y Tarantola, 2002). Sin embargo, existen ciertas limitaciones de los ICs que se deben conocer para mejorar su diseño y evitar posibles críticas a su construcción (Saisana y Tarantola, 2002; Schuschny y Soto, 2009).

VENTAJAS DEL USO DE ICs	LIMITACIONES EN EL USO DE ICs
<ul style="list-style-type: none"> - Sirven para resumir temas complejos y multidimensionales. Integran y sintetizan diferentes dimensiones del concepto que miden. Facilitan la evaluación de la eficacia de las políticas y la rendición de cuentas - Permiten disponer de una ‘imagen de contexto’ (o ‘imagen general’) que facilita la interpretación de temas complejos. Por el contrario, encontrar una tendencia en cada uno de los indicadores por separado es mucho más costoso y complicado. Facilitan, por tanto, la tarea de construcción de rankings de unidades de análisis en temas multidimensionales - Atraen el interés público por su capacidad de proporcionar una comparación y evolución entre las diferentes unidades de análisis - Ayudan a reducir el tamaño de la lista de indicadores simples 	<ul style="list-style-type: none"> - Pueden transmitir un mensaje confuso si están mal contruidos o son mal interpretados - La ‘imagen general’ que transmiten puede llevar a conclusiones simplistas en la opinión de los políticos. Los ICs se deben utilizar en combinación con los indicadores simples para extraer conclusiones políticas adecuadas - Su construcción requiere de fases en las que se deben hacer juicios de valor. Para una correcta construcción, los juicios de valor deben ser transparentes y deben basarse en principios estadísticos siempre que sea posible. Además, deben ir acompañados de un análisis de sensibilidad y de incertidumbre que evalúe la calidad y robustez de los resultados - Su diseño debe realizarse a partir de un conjunto de información medible. Ello requiere que los datos estén ampliamente disponibles, sus frecuencias de muestreo sean razonables en relación a los objetivos que se plantean y las unidades de análisis hayan consensuado un tolerable nivel de armonización sobre las estadísticas e indicadores a utilizar

Tabla 3.1. Ventajas y limitaciones del uso de ICs

A partir de la tabla 3.1 se concluye que es más fácil interpretar el resultado de un IC que encontrar una tendencia común en los indicadores individuales para cada unidad de análisis. Por otra parte, los ICs son considerados de gran utilidad para los ejercicios de comparación en los rankings de países. Sin embargo, pueden transmitir mensajes políticos erróneos y poco robustos si son pobremente contruidos o mal interpretados.

La construcción de un IC requiere de etapas en las que se realizan elecciones subjetivas. Ejemplos de estas etapas son: la selección de los indicadores simples, el tratamiento de datos faltantes, la asignación de pesos a los indicadores simples, la elección del modelo de agregación, etc. Es por eso que se puede correr el riesgo de que las elecciones subjetivas puedan ser utilizadas con el fin de manipular los resultados según convenga. Por ejemplo, un país que

no esté contento con su situación en un ranking podría luchar para que se considerasen dimensiones adicionales en la metodología que le hicieran mejorar su posición.

Para evitar este problema, la metodología aplicada en el diseño del IC debe mostrarse de forma transparente, clara y justificada. Además, un IC debe basarse fundamentalmente en los tres aspectos siguientes:

- Un marco conceptual sólido (estructura del fenómeno a medir)
- Un proceso sólido de construcción (selección de indicadores, normalización y agregación)
- Buena calidad de los datos subyacentes

Una buena base técnica en su construcción puede lograr una mayor robustez del IC y una mayor resistencia al paso del tiempo, de forma que facilite negociaciones en lugar de potenciar el desacuerdo. Sin embargo, una buena base técnica con datos subyacentes pobres producirá resultados pobres. Es por eso que los tres aspectos descritos representan tres condiciones necesarias que se deben cumplir para que la construcción de los ICs sea sólida y fundamentada. En resumen, siguiendo a Nardo et al. (2008), la transparencia y coherencia debe estar presente en todo el proceso de construcción, desde la formulación del problema hasta la interpretación de los resultados.

Son muchas las críticas que se realizan sobre el uso de ICs y los problemas que conlleva su cálculo. Algunos ejemplos son: la generalización excesiva, la selección de los indicadores simples, la comparabilidad del IC en situaciones diversas, el contrapeso que se produce entre indicadores de naturaleza muy diversa, etc. El economista Koopmans (1947) critica abiertamente la construcción de ICs que no tienen una adecuada y justificada base teórica a partir de la cual iniciar el análisis. Entre sus argumentos destaca que si no se realiza una selección adecuada y correcta de los indicadores simples que formarán parte del IC o si ésta no abarca las principales dimensiones del objeto de estudio, difícilmente el índice desarrollado mostrará algo representativo acerca del concepto que se desea estudiar. De forma general, las principales críticas que reciben los ICs se centran en el incumplimiento de alguna de estas hipótesis de partida. En particular, una de las críticas que genera más controversia es la elección de las técnicas de agregación utilizadas para sintetizar los indicadores simples en un único valor (Hendrik, Howard y Maximilian, 2008). En este sentido, según Sharpe (2004):

“The aggregators believe there are two major reasons that there is value in combining indicators in some manner to produce a bottom line. They believe that such a summary statistic can indeed capture reality and is meaningful, and that stressing the bottom line is extremely useful in garnering media interest and hence the attention of policy makers. The second school, the non-aggregators, believe one should stop once an appropriate set of indicators has been created and not to the further step of producing a composite index. Their key objection to aggregation is what they see as the arbitrary nature of the weighting process by which the variables are combined”.

De forma similar, Saisana, Saltelli y Tarantola (2005) realizan la siguiente afirmación:

“[...] it is hard to imagine that debate on the use of composite indicators will ever be settled [...] official statisticians may tend to resent composite indicators, whereby a lot of work in data collection and editing is “wasted” or “hidden” behind a single number of dubious significance. On the other hand, the temptation of stakeholders and practitioners to summarise complex and sometime elusive processes (e.g. sustainability, single market policy, etc.) into a single figure to benchmark country performance for policy consumption seems likewise irresistible”.

Stiglitz, Sen y Fitoussi (2010) afirman que el uso de indicadores estadísticos se ha incrementado en las últimas dos décadas, sin embargo, existe una crítica general debido al carácter arbitrario de los procedimientos de ponderación y a la falta de una justificación clara de las implicaciones de la elección de estos procedimientos:

“[...] the role of statistical indicators has increased over the two decades.”

“[...] a general criticism that is frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components.”

“[...] the problem is not that these weights procedures are hidden, non-transparent or non-replicable – they are often very explicitly presented by the authors of the indices, and this is one of the strengths of this literature. The problem is rather that their normative implication are seldom made explicit or justified”.

Al igual que otros autores, Cox et al. (1992) exponen que las dificultades comunes que se plantean en la construcción de un IC se centran en la asignación de pesos. Es por ello que las discusiones sobre la elección del método de ponderación más apropiado no son fácilmente resolubles.

A pesar de las críticas recibidas, los ICs son útiles para proporcionar a los expertos, a las partes interesadas y a los tomadores de decisiones los siguientes aspectos:

- La dirección de los acontecimientos
- La comparación entre unidades de análisis
- La evaluación de estado y tendencia en relación con los objetivos y metas propuestos
- La rápida alerta o advertencia sobre posibles situaciones extremas
- La identificación de planes de acción y de mejora
- La anticipación de condiciones futuras y tendencias
- El canal de comunicación para el público en general y para los tomadores de decisiones

Toda iniciativa que busque diseñar un IC deberá estar orientada por una demanda potencial puesta de manifiesto por los actores sociales que pudieran estar involucrados en el área bajo análisis. Por tanto, es esencial la interacción con pares y expertos durante el proceso de

construcción de un IC para conseguir la mutua aceptación del índice, el cual debe tratarse como una herramienta útil de investigación y comparación sobre la base del consenso. Sin embargo, a pesar de basarse en valoraciones subjetivas en algunas de las etapas de su construcción, el diseño del IC requiere de la aplicación minuciosa de principios estadísticamente fundamentados y procedimientos cuantitativos transparentes (Paruolo, Saltelli y Saisana, 2013).

Aunque la ciencia no pueda proporcionar un método objetivo para el desarrollo de un único y verdadero IC que simplifique un sistema complejo, puede contribuir de manera significativa a asegurar que los procesos de agregación sean tan sólidos y transparentes como sea posible. Esta conclusión se resume en el siguiente punto del Comunicado de la Comisión de Indicadores Estructurales (European Commission, 2001):

“41. The Commission is also considering an approach based on composite indicators. Such indicators are already used in some of the more detailed sectoral processes such as the Innovation Scoreboard. Composite indicators are calculated by weighting together a set of well chosen constituent indicators to provide a summary of each Member State's progress in a particular policy domain. Composite indicators would have the advantage of providing a broader coverage of information than can be included in the current list of structural indicators and they would also allow for a reduction in the number of indicators presented in the list. However, because composite indicators invite strong policy messages to be concluded they need to be robust and based on a sound methodology. More work is therefore necessary to develop such indicators, to examine how they could be integrated into the list of structural indicators and to assess the consistency of the policy messages they send.”

Una revisión más profunda de los diferentes puntos de vista de la construcción de ICs se puede encontrar en el artículo publicado por Saltelli (2007) y titulado “Composite Indicators between analysis and advocacy”.

3.4. Construcción de un Indicador Compuesto¹

La construcción de un IC requiere de dos condiciones básicas a tratar (Schuschny y Soto, 2009):

- La definición clara del atributo que se desea medir
- La existencia de información confiable para poder realizar la medición

Estas dos condiciones son indispensables para poder plantearse la posibilidad de construir un IC. El cumplimiento de la primera condición proporcionará al IC un sustento conceptual,

¹ Esta sección está basada en el libro titulado “Handbook on Constructing Composite Indicators: Methodology and user guide” desarrollado conjuntamente por la OCDE y el centro de investigación Joint Research Centre de la Comisión Europea (Nardo et al., 2008).

mientras que el cumplimiento de la segunda le otorgará validez. Ambas condiciones deben validarse antes de considerar los aspectos metodológicos de la construcción del índice.

Un requerimiento adicional a los expuestos anteriormente es la definición de un objetivo claro por el cual se está creando el IC. En la mayoría de los casos, los ICs se construyen con el objetivo de medir el desempeño de unidades de análisis, bien sean países, regiones, organizaciones, etc. en un área o tema determinado.

METODOLOGÍA DE CONSTRUCCIÓN DE UN IC

La construcción de un IC debe basarse en una metodología sólida y transparente que pueda ser fácilmente interpretada. Esta metodología se divide en diferentes etapas (Saisana y Tarantola, 2002):

- **Decisión sobre el atributo que se quiere medir** y si la utilización de un IC es beneficioso para el estudio del atributo
- **Selección de los indicadores simples.** Es necesario tener una idea clara para justificar que la elección de los indicadores simples es relevante y que éstos son apropiados para el estudio del concepto multidimensional que se quiere medir. No existe un método objetivo para la elección de los indicadores relevantes
- **Evaluación de la calidad de los datos.** Es necesario disponer de datos de buena calidad para todos los indicadores simples. De lo contrario, el analista deberá decidir si eliminar parte de los datos o aplicar la técnica adecuada de imputación para el cálculo de los datos faltantes
- **Evaluación de la relación entre indicadores simples.** La aplicación de técnicas estadísticas, como el análisis de componentes principales, informa de la relación existente entre los indicadores simples. Este tipo de análisis preliminar sobre los indicadores es un paso previo necesario a la construcción del IC
- **Selección del método de normalización, ponderación y agregación de los indicadores simples.** Muchos de los métodos de normalización, ponderación y agregación de los indicadores simples están descritos en la literatura, pero la elección del método apropiado en cada caso depende de los datos, del analista y del objetivo del estudio
- **Evaluación de la calidad del IC.** Inevitablemente, cambios en la elección del método de agregación y normalización, y las diferentes formas de selección de los indicadores simples son algunos de los factores que afectan al resultado final del IC. Por consiguiente, es

necesario realizar un análisis de sensibilidad y de incertidumbre con el fin de evitar críticas sobre la robustez del IC ante pequeños cambios en su estructura

ETAPAS DE CONSTRUCCIÓN DE UN IC

Algunos autores sugieren diez etapas para el diseño de un IC que engloban todo el proceso de construcción, desde el desarrollo de un marco conceptual teórico hasta la presentación y disseminación del IC. Según Nardo et al. (2008), cada etapa de la construcción es muy importante, pero la coherencia en todo el proceso es crucial. Las decisiones tomadas en cada una de las etapas pueden repercutir en las decisiones de las posteriores etapas. Como consecuencia, la construcción del IC no sólo debe tener en cuenta la elección de la metodología más apropiada en cada paso, sino que se debe seleccionar aquella que en su conjunto encaje bien con todo el proceso de construcción.

En la tabla 3.2 se describen los objetivos generales de cada una de las etapas de construcción de un IC que se enumeran a continuación (Nardo et al., 2008; Saisana, 2008):

1. Desarrollo de un marco conceptual teórico
2. Selección del conjunto de indicadores simples
3. Tratamiento de datos
4. Análisis multivariante
5. Normalización de los datos
6. Ponderación y agregación de los indicadores simples
7. Análisis de incertidumbre y sensibilidad del IC
8. Conexión con otros indicadores
9. Identificación de los indicadores simples en el IC
10. Visualización de los resultados

Etapa	Objetivos a alcanzar al final de la etapa
<p>1. Marco conceptual teórico Proporciona la base para la selección y combinación de los indicadores simples utilizados en la construcción del IC bajo unos principios de adecuación a los fines. En esta etapa se requiere la opinión de expertos.</p>	<ul style="list-style-type: none"> - Tener una clara comprensión y definición del concepto multidimensional que se pretende medir - Definir los diferentes subgrupos o dimensiones del concepto, si fuera necesario - Crear una lista de criterios de selección para los indicadores simples
<p>2. Selección del conjunto de indicadores simples La selección debe basarse en la mensurabilidad, la cobertura de las unidades de análisis, la pertinencia de los indicadores que se utilizan para medir el fenómeno y la relación entre ellos. El uso de variables proxy debe considerarse cuando los datos son escasos. En esta etapa se requiere de la opinión de expertos.</p>	<ul style="list-style-type: none"> - Comprobar la calidad de los indicadores simples disponibles - Estudiar las fortalezas y debilidades de cada indicador seleccionado - Crear una tabla resumen donde aparezcan las características de los indicadores como, por ejemplo, disponibilidad periódica de los datos, fuente de datos, interpretación, etc.
<p>3. Tratamiento de datos El tratamiento de datos consiste en:</p> <ul style="list-style-type: none"> - Imputar datos faltantes - Examinar la presencia de valores atípicos (ya que pueden convertirse en puntos de referencia no deseados) - Transformar datos sesgados 	<ul style="list-style-type: none"> - Obtener un conjunto de datos completo - Obtener una medida de fiabilidad de cada valor imputado con el fin de estudiar el impacto de la técnica de imputación en la puntuación final del IC - Tratar la presencia de valores atípicos en el conjunto de datos - Aplicar ajuste de escalas, si fuera necesario - Transformar los indicadores, si fuera necesario
<p>4. Análisis Multivariante Se aplica para estudiar la estructura subyacente de la base de datos, evaluar la idoneidad de los indicadores simples seleccionados y guiar las elecciones metodológicas posteriores (asignación de pesos, agregación,...).</p>	<ul style="list-style-type: none"> - Comprobar la estructura subyacente del conjunto de datos a través de dos dimensiones principales: los indicadores simples y las unidades de análisis (observaciones) (aplicación de métodos multivariantes apropiados: análisis de componentes principales, análisis de conglomerados,...) - Identificar los subgrupos de indicadores individuales o grupos de unidades de análisis que son estadísticamente ‘similares’ e interpretar los resultados obtenidos - Analizar la estructura estadística de la base de datos y comparar los resultados con el marco conceptual teórico con el fin de tratar las posibles diferencias o similitudes
<p>5. Normalización de los datos Debe llevarse a cabo para hacer comparables los indicadores simples</p>	<ul style="list-style-type: none"> - Seleccionar un procedimiento de normalización adecuado con el marco conceptual teórico y con las propiedades de los datos

Etapa	Objetivos a alcanzar al final de la etapa
<p>6. Ponderación y agregación de los indicadores simples Debe realizarse siguiendo las líneas del marco conceptual teórico definido.</p>	<ul style="list-style-type: none"> - Seleccionar el procedimiento de ponderación y agregación apropiado con los objetivos del estudio - Discutir si se va a considerar o no la compensación entre los indicadores o entre las dimensiones
<p>7. Análisis de incertidumbre y sensibilidad Se realiza para evaluar la calidad y robustez del IC teniendo en cuenta las diferentes elecciones metodológicas realizadas durante su construcción (argumentos para decidir la inclusión o exclusión de un indicador, el método de normalización, la imputación de datos faltantes, la elección de los pesos, el método de agregación, etc.).</p>	<ul style="list-style-type: none"> - Considerar procedimientos metodológicos alternativos para la construcción del IC y, si fuese posible, escenarios conceptuales alternativos - Identificar las fuentes de incertidumbre en el desarrollo del IC y proporcionar las puntuaciones de este según dichas fuentes de incertidumbre - Llevar a cabo un análisis de sensibilidad y de incertidumbre de cada uno de los escenarios definidos
<p>8. Conexión con otros indicadores Debe calcularse la correlación entre el IC (o sus dimensiones) e indicadores existentes en el mismo ámbito de estudio (simples o compuestos) para determinar los vínculos a través de regresiones.</p>	<ul style="list-style-type: none"> - Calcular la correlación del IC con otros fenómenos medibles relevantes teniendo en cuenta las variaciones del IC determinadas según el análisis de sensibilidad
<p>9. Identificación de los indicadores simples en el IC Deben identificarse los indicadores simples que afectan positiva o negativamente al valor final del IC para cada unidad de análisis. La transparencia, tanto en el proceso como en los resultados, es primordial para un buen análisis y una buena formulación de políticas.</p>	<ul style="list-style-type: none"> - Estudiar el comportamiento de las unidades de análisis a través de los indicadores simples para revelar cuáles son los más influyentes en la puntuación final del IC y así, poder aplicar políticas de mejora individualizadas para cada unidad de análisis
<p>10. Visualización de los resultados Esta etapa debe recibir una especial atención dado que la visualización del IC puede influir (o contribuir para mejorar) en su interpretación.</p>	<ul style="list-style-type: none"> - Identificar un conjunto coherente de herramientas de presentación para el público objetivo - Seleccionar aquellas técnicas de visualización que proporcionen mayor información - Visualizar los resultados del IC de forma clara, sencilla y precisa

Tabla 3.2. Etapas de la construcción de un IC

3.4.1. Desarrollo de un marco conceptual teórico

Lo que está mal definido será erróneamente medido (Nardo et al., 2008).

La primera etapa de la construcción de un IC corresponde a la definición de un marco conceptual teórico desde el que se seleccionan los indicadores simples y se establece la manera en que estos se combinan entre sí a través de dimensiones. El marco conceptual debe ser lo más inteligible y detallado posible para disponer de una mejor definición del IC y de las relaciones entre los indicadores simples que lo componen. Para ello, es necesario tener categorizado, de forma amplia, el contexto de análisis y tener comprensión del fenómeno a medir.

El marco conceptual teórico debe estar basado en lo que se desea medir y no en lo que está disponible para medir. Es conveniente que en esta etapa participen expertos en el área de estudio y grupos de interés, *stakeholders*, para tener en cuenta múltiples puntos de vista y aumentar la solidez del marco conceptual y del conjunto de indicadores. De lo contrario, el IC puede proporcionar unos resultados pobres en información. En este sentido, el analista Freudenberg (2003) afirma: “*Composite Indicators risk becoming exercises in measurement without a theoretical underpinning*”.

Sin embargo, no todos los conceptos multidimensionales poseen fundamentos teóricos sólidos y empíricos a partir de los cuales se deriva un marco conceptual teórico bien definido. Los ICs construidos en el ámbito político y de actualidad, como la competitividad, el desarrollo sostenible, etc., podrían ser muy discutidos, puesto que la investigación en dichos campos todavía está en desarrollo. Por lo tanto, la transparencia en la construcción de ICs es esencial. Para ello, se deben seguir los siguientes pasos en el desarrollo del marco conceptual teórico (Nardo et al., 2008):

- **Definición del concepto multidimensional.** La definición del concepto multidimensional a medir debe proporcionar una idea clara y concisa de lo que se quiere evaluar mediante el IC. Sin embargo, algunos conceptos complejos son muy difíciles de definir y medir con precisión, y pueden ser objeto de controversia entre las partes interesadas. Finalmente, los usuarios de los ICs deben evaluar su calidad y relevancia
- **Clasificación en subgrupos o dimensiones.** Es recomendable dividir el concepto multidimensional en subgrupos o dimensiones. No es necesario que las dimensiones sean estadísticamente independientes entre sí, pero los vínculos existentes deben describirse teórica y empíricamente cuando la consistencia lógica en relación con los fenómenos considerados favorezca a la organización conceptual y no a la puramente estadística. Esta división en dimensiones facilita la asignación de pesos en los diferentes factores. Se debe hacer referencia al marco conceptual conectando los diferentes subgrupos o dimensiones con los indicadores simples

- **Identificación de los criterios de selección para el conjunto de indicadores subyacentes.** Los criterios de selección se utilizan como una guía para la decisión de la inclusión o exclusión de un indicador simple en el IC global. Deben ser lo más claros y concisos posible para facilitar la elección del conjunto de indicadores relevantes que va a ser objeto de estudio en posteriores análisis

3.4.2. Selección del conjunto de indicadores simples

La calidad de un IC es consecuencia directa de la calidad de los indicadores simples que lo definen (Nardo et al., 2008). Por ello, los indicadores simples deben seleccionarse en base a su calidad, su relevancia, su disponibilidad y la frecuencia con la que se muestrean.

Tal y como se expone en el apartado 3.2.5, el sistema de indicadores debe cumplir ciertas propiedades que permitan medir el grado de cumplimiento de los objetivos planteados en el estudio.

Además, existen diferentes criterios para la selección de los indicadores simples que aseguran la calidad de un sistema de indicadores para la construcción de un IC. Un ejemplo, basado en los cuatro criterios siguientes, se propuso para la construcción del Índice de Innovación de las empresas e industrias desarrollado por la Comisión Europea (European Commission-DG MARKT, 2001):

- **Relevancia política:** se deben seleccionar los indicadores simples que resulten relevantes en la toma de decisiones políticas (p. ej. Objetivos de Lisboa)
- **Redundancia:** si dos indicadores aportan la misma información se recomienda seleccionar sólo uno de ellos
- **Correlación:** si dos indicadores están muy correlacionados pero ambos transmiten mensajes políticos fuertes y distintos se pueden incluir en la lista final de indicadores relevantes
- **Disponibilidad:** se recomienda utilizar indicadores que estén disponibles para un gran número de unidades de análisis y que se puedan obtener con regularidad de una base de datos actualizada

Por otra parte, el proyecto europeo “E3M-European indicators and ranking methodology for university third mission” (Carot et al., 2010) presenta otra propuesta metodológica para la definición de un conjunto de indicadores que tiene como objetivo medir la tercera misión de las universidades. A continuación se expone la propuesta metodológica desarrollada en este proyecto.

En primer lugar, se definen las dimensiones que miden la tercera misión. Posteriormente, se identifican grandes procesos dentro de cada una de las dimensiones. Dentro de cada proceso, a su vez, se obtiene una serie de indicadores asociados que son considerados en un principio como relevantes. El proceso de selección de esta primera batería de indicadores se realiza en base a técnicas de consenso. Finalmente, el conjunto de indicadores relevantes se simplifica en una serie de indicadores factibles, de forma que su medición resulte viable.

En este contexto, se definen los indicadores como parámetros que permiten evaluar de forma cuantitativa la eficacia y/o eficiencia de los procesos, entendiendo como proceso el conjunto de pasos que se realizan de forma secuencial para conseguir elaborar productos o servicios a partir de determinados factores de entrada.

Mientras que la elección de los indicadores debe estar guiada por el marco conceptual teórico, el proceso de selección de los datos puede ser un tanto subjetivo. En esta fase de la construcción del IC deben actuar expertos y partes interesadas. Por tanto, puede ser que no haya un conjunto definitivo y único de indicadores.

Un aspecto importante a tener en cuenta durante la etapa de selección de los indicadores simples es el objetivo o uso que se le quiere dar posteriormente al IC. La selección de los indicadores debe estar enfocada de una manera particular dependiendo del tipo de objetivo planteado (p. ej. objetivo de diagnóstico, de evaluación de desempeño o de pronóstico sobre escenarios futuros) (Schuschny y Soto, 2009). Además, el proceso de selección debe estar documentado mediante la construcción de metadatos donde se especifiquen las características de los indicadores, su disponibilidad, las fuentes responsables de su cálculo, el tipo de indicador, las unidades de medida con las que se expresa, etc.

Para determinados sectores, la escasez de datos cuantitativos comparables a nivel internacional podría limitar la capacidad de construcción de un buen IC. Para solucionar este problema se incluyen datos cualitativos de encuestas o estudios de políticas. Las variables *proxy* también se pueden utilizar cuando los datos deseados no están disponibles o cuando la comparación de los distintos países es limitada. Esto ocurre, por ejemplo, cuando se quiere medir el número de trabajadores que utilizan ordenadores y este dato no está disponible. En su lugar, el número de trabajadores que tienen acceso a ordenadores puede usarse como una variable *proxy*. Sin embargo, en el caso de datos cualitativos el uso de variables *proxy* no es muy recomendable. En la medida en que los datos lo permitan, la exactitud de las variables *proxy* se debe comprobar a través de análisis de correlación.

La actual tendencia en la construcción de ICs que miden el rendimiento de países, definidos en un marco de políticas, puede proporcionar un mayor ímpetu en la colección de datos y de información, así como una mejora en la identificación de nuevas fuentes de datos y comparación internacional de estadísticas. No obstante, es importante mencionar la mayor limitación que presenta el diseño de un IC y que consiste en la ausencia de información

estadística en la que se basa el indicador. Para ello, existen métodos de imputación de datos faltantes.

Con todo lo anterior y para finalizar esta etapa de la construcción de un IC, se concluye que si no se realiza una selección correcta de los indicadores simples o si esta no abarca las principales dimensiones del objeto de estudio, difícilmente el índice desarrollado mostrará algo representativo acerca del concepto que se desea estudiar.

3.4.3. Tratamiento de datos

Detección de valores atípicos y de datos faltantes

Los valores atípicos son observaciones con unas características muy diferentes, en todos o en algunos de los aspectos analizados, al resto de las observaciones. Su presencia puede tener un gran impacto en posteriores análisis estadísticos como el análisis de correlación, la imputación, el análisis multivariante, etc. y pueden producir sesgos indeseables si se trata de casos atípicos problemáticos (valores mal recogidos en la base de datos, errores de transcripción,...). Por otra parte, existen casos atípicos que no son problemáticos a pesar de ser diferentes a la mayor parte de la muestra o la población. Estos son, por ejemplo, datos reales con comportamientos anómalos.

Se pueden detectar valores atípicos desde una perspectiva univariante o multivariante (Hair et al., 2007):

- **Nivel univariante.** En este caso se examina la distribución de las observaciones detectando como valores atípicos aquellos que caen fuera de los rangos de la distribución. Se recomienda estandarizar los datos (con media cero y desviación típica uno) para poder realizar comparaciones entre los indicadores simples. Para muestras pequeñas (de 80 observaciones o menos) se suelen considerar como valores atípicos aquellos que tienen un valor, en valor absoluto, de 2,5 o superior. Para muestras grandes el umbral se sitúa entre 3 o 4.

También se pueden detectar valores atípicos analizando la asimetría y la curtosis de las distribuciones de los indicadores simples. Si la asimetría, en valor absoluto, es mayor que 1 o la curtosis mayor que 3,5 se considera que existe algún caso atípico en el indicador simple evaluado (Saisana, 2010). Al realizar un histograma del indicador simple también se puede detectar el caso atípico. En este caso, el analista debe decidir si mantenerlo o excluirlo del análisis, o simplemente ajustar el valor para obtener una asimetría y curtosis dentro de los rangos establecidos.

Otra forma de detección de casos atípicos se realiza estimando el rango intercuartílico definido como la distancia que hay entre el primer y el tercer cuartil. Como regla empírica, se considera dato atípico un valor inferior al primer cuartil menos 1,5 veces el rango intercuartílico o superior al tercer cuartil más 1,5 veces el mismo rango.

- **Nivel multivariante:** se pueden identificar los valores atípicos multivariantes con la medida D^2 de Mahalanobis o mediante la T^2 de Hotelling. Sin embargo, en el ámbito de los ICs se suelen detectar posibles valores atípicos sólo a nivel univariante (Nardo et al., 2008; Saisana, 2010; Annoni, 2010). No obstante, en este estudio se ha considerado interesante aplicar también las técnicas multivariantes de detección de valores atípicos (véase capítulo 6).

Una vez identificados los valores atípicos, el analista debe decidir si retener o excluir cada uno de ellos juzgando no sólo las características del caso atípico, sino también los objetivos del análisis.

Seguidamente al análisis de los valores atípicos, se recomienda realizar un análisis exploratorio para evaluar si, efectivamente, los datos con la información seleccionada están en concordancia con las ideas que dieron lugar a su elección. Este es un primer proceso de validación de la utilidad de los indicadores seleccionados en el cual pueden manifestarse problemas de ausencia parcial de información. Ello puede conducir a problemas en posteriores etapas, dado que puede generar errores en los análisis que deriven en conclusiones finales equívocas. Por consiguiente, es necesario recurrir a las técnicas de imputación de datos perdidos o faltantes.

Se define la imputación como el “*proceso de estimación de los datos faltantes basado en datos disponibles de otras variables y/o casos de la muestra*” (Hair et al., 2007). El objetivo de la imputación es emplear relaciones conocidas que puedan identificarse en los datos disponibles para estimar los datos faltantes. Sin embargo, la imputación tiene un gran impacto sobre análisis posteriores de los datos, por tanto, el analista debe considerar cuidadosamente el uso de esta técnica en cada momento: “*La idea de imputación es tan seductora como peligrosa. Es seductora porque puede llevar al usuario a creer que los datos están completos después de todo, y es peligrosa porque une situaciones donde el problema es suficientemente menor con situaciones donde los estimadores estándar aplicados a los datos reales e imputados tienen sesgos sustanciales*” (Dempster y Rubin, 1983).

La ausencia de datos en algunas unidades de análisis está presente en casi todos los casos de estudio de ICs. A menudo, la falta de datos obstaculiza el desarrollo de un IC robusto. Muchas de las situaciones en las que se detectan valores ausentes se deben a factores de procedimientos como errores en la recogida de datos, fallos al completar cuestionarios, etc. Y otras están relacionadas directamente con el encuestado, como por ejemplo la renuncia a responder ciertas preguntas de una encuesta. Para evitar estos casos de datos ausentes el analista debe cerciorarse

de la buena calidad de la recogida de datos y de la respuesta en las encuestas realizadas durante el diseño de la investigación.

Para poder aplicar una solución a la ausencia de datos se debe averiguar el grado de aleatoriedad en los datos faltantes. Para ello, supóngase que se observan dos indicadores I_1 e I_2 , siendo I_1 el indicador sin datos ausentes e I_2 el indicador con datos ausentes. A continuación se plantean tres patrones de comportamiento a los que pueden obedecer los datos ausentes según el grado de aleatoriedad (Hair et al., 2007):

- **Pérdida de datos completamente ocasional (*Missing completely at random* – MCAR).** Cuando hay un mayor nivel de aleatoriedad el proceso es completamente aleatorio. En este caso, los valores observados de I_2 son un muestra aleatoria de los valores de I_2 sin un proceso subyacente que tiende a sesgar los datos observados. Si los datos ausentes siguen este patrón, cualquier solución se podría aplicar sin tener en cuenta el impacto de cualquier otro indicador o proceso de datos ausentes
- **Pérdida de datos ocasional (*Missing at random* – MAR).** Los datos ausentes siguen un patrón MAR si los valores ausentes de I_2 dependen de I_1 , pero no en I_2 . Es decir, que los valores observados para I_2 representan una muestra de los valores reales de I_2 para cada valor de I_1 , pero los datos observados para I_2 no representan necesariamente una muestra verdaderamente aleatoria para todos los valores de I_2 . Aunque el proceso de datos ausentes es aleatorio en la muestra, sus valores no son generalizables para la población
- **Pérdida de datos sistemática (*Not missing at random* – NMAR).** Si se encuentra un proceso de datos ausentes entre I_1 e I_2 , donde existen diferencias significativas para casos de I_2 con datos válidos y datos ausentes en función de los valores de I_1 , entonces los datos ausentes no son aleatorios. En la práctica es común que se presenten situaciones en las que los datos faltantes no siguen un patrón completamente aleatorio (MCAR) y tampoco aleatorio (MAR)

Desafortunadamente, no hay ningún método estadístico para tratar la pérdida de datos sistemática y, a menudo, no se puede diferenciar si los datos faltantes se han producido por pérdida ocasional o sistemática. La mayoría de los métodos que se utilizan para la imputación de datos requieren un mecanismo de pérdida de datos completamente aleatoria MCAR o al menos aleatoria MAR.

Un método utilizado para probar si los datos faltantes siguen un patrón MCAR consiste en aplicar la prueba de hipótesis propuesta por Little (1988) que se incluye en la opción MVA (*Missing Value Analysis*) del paquete SPSS cuando se aplica el algoritmo EM (Expectación-Maximización). El test de hipótesis que se plantea es el siguiente:

H_0 : Los datos ausentes siguen un patrón MAR

H_1 : Los datos ausentes siguen un patrón MCAR

El estadístico del test sigue una distribución χ^2 con f grados de libertad. Conforme a la regla de decisión, se debe rechazar la hipótesis nula si el valor del estadístico es mayor al valor en tablas según un nivel de significación predeterminado α .

Técnicas de Imputación

La literatura existente sobre el tratamiento de los datos faltantes es muy extensa y está en actual desarrollo. A lo largo de esta sección se van a describir los principales métodos de imputación de datos. Para un estudio profundo de los diferentes métodos de imputación e identificación de los patrones que siguen los datos ausentes véase (Little y Rubin, 1987; Little, 1988; Hair et al., 2007; Medina y Galván, 2007).

IMPUTACIÓN POR ELIMINACIÓN

En este procedimiento se omiten los valores ausentes para posteriores análisis. En la construcción del IC significa omitir un indicador para todas las unidades de análisis u omitir una unidad de análisis completa, lo cual implica descartar datos que pueden haber sido costosos de obtener.

Este tipo de imputación no es muy utilizada, puesto que el objetivo del IC es proporcionar una puntuación para todas las unidades de análisis manteniendo el mayor número posible de indicadores subyacentes.

Por otra parte, los errores estándar serán más grandes en muestras reducidas dado que se utiliza menos información. Como regla general, si un indicador tiene más de un 5% de datos faltantes los casos no se eliminan (Little y Rubin, 1987).

IMPUTACIÓN SIMPLE

La imputación simple consiste en el método efectivo de sustitución de los valores ausentes por datos estimados sobre la base de información disponible en la muestra. Esta medida puede realizarse de muchas maneras, que van desde una sustitución directa de valores hasta procesos de imputación basados en relaciones entre indicadores. A continuación se van a nombrar los métodos más ampliamente utilizados, aunque existen otras muchas formas de imputación simple (Little y Rubin, 1987; Hair et al., 2007):

Modelización Implícita

En esta técnica de imputación la atención se centra en un algoritmo con supuestos implícitos que deben evaluarse. Además de la necesidad de verificar cuidadosamente si las suposiciones implícitas son razonables y se ajustan a la cuestión tratada, el riesgo de este tipo de imputación de datos faltantes es considerar al conjunto de datos resultante como completo, olvidándose que se realizó una imputación. La modelización implícita incluye:

- **Imputación Hot deck:** en este método de imputación se llenan los ‘vacíos’ de información a partir de unidades con comportamiento similar. Por ejemplo, en una encuesta se podría agregar la información que respondieron ciertos encuestados a aquellos, de similares características, que no las respondieron
- **Sustitución de caso:** en este método las observaciones con datos ausentes se sustituyen por otras observaciones no muestrales. Un ejemplo común es reemplazar un encuestado que está en la muestra pero con el que no se ha podido contactar o que tiene gran cantidad de datos ausentes por otro encuestado que no está en la muestra, preferiblemente muy similar al de la observación original. Este método es el que más se utiliza para sustituir las observaciones con datos ausentes completos, aunque también puede emplearse para reemplazar observaciones con menor cantidad de datos ausentes
- **Imputación Cold deck:** en este método el analista sustituye los datos ausentes por un valor constante derivado de fuentes externas o investigación previa

Modelización Explícita

La modelización explícita se realiza considerando un modelo estadístico que parte de supuestos concretos y explícitos. La modelización explícita incluye:

- **Imputación por el método de la Media/Mediana/Moda no condicionada:** este método consiste en sustituir los valores ausentes por una indicador cuyo valor medio se calcula sobre todas las respuestas disponibles. Así, las respuestas de la muestra disponible se usan para calcular el valor de sustitución (media/mediana/moda). Este tipo de imputación es uno de los métodos más empleados.

La sustitución de los valores ausentes por la media es extensamente utilizada, sin embargo presenta tres inconvenientes. En primer lugar, invalida las estimaciones de la varianza derivadas de las fórmulas estándar de la varianza para conocer la verdadera varianza de los datos. Por la manera en que se realiza la sustitución de los datos, la suma de cuadrados de las desviaciones de las observaciones respecto de la media permanece inalterada pero se incrementa el tamaño de la muestra, lo cual origina que la varianza del indicador disminuya

y se generen, de forma artificial, intervalos de confianza más estrechos. En segundo lugar, la distribución real de los valores se encuentra distorsionada por la sustitución de los datos ausentes por la media. Finalmente, este método modifica la correlación observada puesto que todos los datos ausentes tendrán un valor único constante. Sin embargo, tiene la ventaja de poderse llevar a cabo fácilmente y proporcionar una información completa para todos los casos.

En la figura 3.3 se ilustran las consecuencias de utilizar el método de la media no condicionada (Medina y Galván, 2007). En el primer gráfico se observa la distribución de frecuencias para las observaciones completas, en donde se asume que la desviación estándar del indicador que se está analizando es $\sigma = 310$ unidades. En el segundo gráfico se considera que el 30% de los datos no tiene información, por lo que la desviación estándar se incrementa a $\sigma = 350$ unidades, y en el último gráfico se observa que cuando los datos faltantes son sustituidos por la media, la desviación estándar se reduce a $\sigma = 249$ unidades.

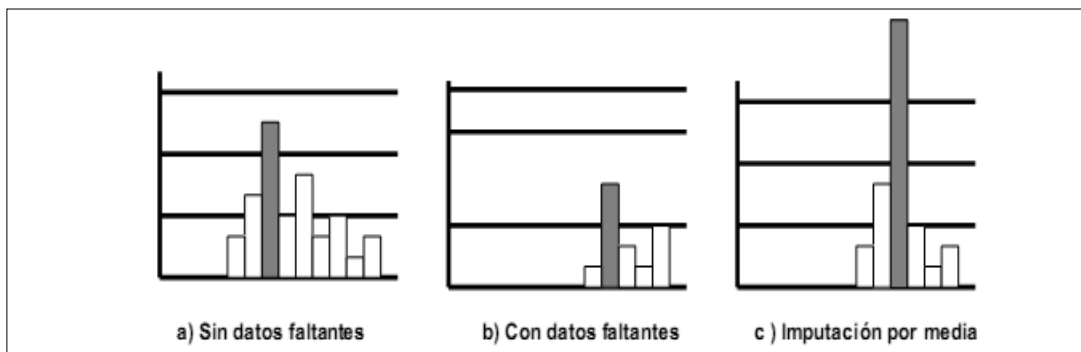


Figura 3.3. Efecto de la imputación con la media no condicionada

Bajo este procedimiento de imputación el valor medio del indicador se mantiene, pero otros estadísticos que definen la forma de la distribución como la varianza, covarianza, cuartiles, curtosis, etc. pueden verse afectados.

A continuación se va a formalizar en términos matemáticos el procedimiento de imputación por el método de medias no condicionadas. Sea X_q la variable aleatoria asociada al indicador simple q con $q=1, \dots, Q$ y x_{qc} el valor observado de X_q para la unidad de análisis c , con $c=1, \dots, M$. Para algunas unidades de análisis c , sea m_q el número de valores registrados en X_q , y $M - m_q$ el número de valores ausentes. La media no condicional se calcula de la siguiente forma (Nardo et al., 2008):

$$\bar{x}_q = \frac{1}{m_q} \sum_{\text{registrados}} x_{qc}$$

De forma similar, se calcula la mediana (el valor que divide en dos partes iguales la distribución de la variable aleatoria) y la moda (el valor con mayor frecuencia) de la

distribución sobre la muestra de datos disponible con el fin de sustituir los valores faltantes por estos valores.

- **Imputación por regresión:** en este método se usa el análisis de regresión para predecir los valores ausentes de un indicador basándose en su relación con otros indicadores del conjunto de datos. La ventaja de este procedimiento es el uso que se le da a las relaciones ya existentes en la muestra como base de predicción. Sin embargo, la aplicación de este método i) refuerza las relaciones ya existentes en los datos, por lo que los datos resultantes finalmente son más característicos de la muestra y menos generalizables, ii) se subestima la varianza de la distribución (menos cuando se añaden valores estocásticos a los valores estimados), iii) se supone que el indicador con datos ausentes tiene correlaciones sustanciales con otros indicadores, en caso contrario este método no es preferible como método de imputación y iv) los valores estimados puede que no pertenezcan a los rangos válidos de los indicadores, requiriendo por tanto alguna forma de ajuste adicional.

A pesar de los inconvenientes que presenta este método, se utiliza mucho cuando se presentan niveles moderados de dispersión de los datos ausentes y cuando las relaciones entre los indicadores son suficientemente significativas.

Para describir el método, supóngase que se tiene un conjunto de $h - 1 < Q$ indicadores con datos completos (x_1, \dots, x_{h-1}) y un indicador x_h observado en r unidades de análisis, pero con datos perdidos en $M-r$ de ellas (Nardo et al., 2008). Este método realiza una regresión de x_h con (x_1, \dots, x_{h-1}) usando las r observaciones completas de tal manera que la imputación se haga a partir de la predicción:

$$\hat{x}_{ih} = \hat{\beta}_0 + \sum_{j=1}^{h-1} \hat{\beta}_j x_{ij}, \quad i = 1, \dots, M - r$$

Por lo general, la estrategia para definir la “mejor” regresión se basa en un procedimiento que presenta dos fases. En la primera, todos los subconjuntos diferentes de predictores se tienen en cuenta en la regresión. En la segunda fase, el mejor subconjunto se determina usando uno de los siguientes criterios:

- El valor de R^2
- El valor del Cuadrado Medio Residual (RMS)
- El valor de la C_k de Mallows
- Regresión por pasos (regresión “stepwise”)

Una variante a este procedimiento es la imputación por “regresión estocástica”, en la cual los datos faltantes se obtienen con un modelo de regresión más un valor aleatorio asociado al término de error.

$$\hat{x}_{ih} = \hat{\beta}_0 + \sum_{j=1}^{h-1} \hat{\beta}_j x_{ij} + \epsilon_i, \quad i = 1, \dots, M - r; \epsilon_i \sim N(0, \sigma^2)$$

donde σ^2 es la varianza residual de la regresión de x_h con (x_1, \dots, x_{h-1}) basada en los r casos completos.

Este procedimiento garantiza variabilidad en los valores imputados y contribuye a reducir el sesgo en la varianza y en el coeficiente de determinación del modelo.

- **Algoritmo Expectación-Maximización (EM):** es un algoritmo propuesto por Dempster, Laird y Rubin (1977) que presenta una técnica iterativa general para realizar una estimación de máxima verosimilitud de parámetros de problemas con datos ausentes. En el algoritmo intervienen dos etapas (las etapas “E” y “M”). La etapa de Expectación, “E”, realiza las mejores estimaciones posibles de los datos ausentes mientras que la etapa de Maximización, “M”, realiza estimaciones de los parámetros (medias, desviaciones típicas o correlaciones) con la suposición de reemplazamiento de todos los datos ausentes. El proceso continúa realizando las dos etapas hasta obtener una diferencia despreciable en los valores estimados con respecto a etapas anteriores y reemplazar todos los datos ausentes.

Sea X el conjunto de datos. Para describir el algoritmo EM supóngase que los datos se generan por un modelo descrito por la función de distribución $f(X|\theta)$, donde $\theta \in \Omega_\theta$ es el vector de parámetros desconocidos del espacio de parámetros. La función $f(X|\theta)$ captura la relación entre el conjunto de datos y los parámetros del modelo de datos. Como se desconocen los parámetros pero se conocen los datos tiene sentido estimar la probabilidad de observar cierto conjunto de parámetros dados los datos, es decir la función de verosimilitud.

El algoritmo EM alterna etapas de expectación en las que se calcula la verosimilitud esperada mediante la inclusión de variables latentes como si fueran observables y una etapa de maximización, donde se calculan estimadores de máxima verosimilitud de los parámetros mediante la maximización de la verosimilitud esperada del paso anterior. Los parámetros que se encuentran en la etapa de maximización se usan para comenzar la nueva etapa de expectación y así, el proceso se repite recursivamente.

Dado X , sea la función de verosimilitud $L(\theta|X)$ proporcional a $f(X|\theta)$:

$$L(\theta|X) = k(X)f(X|\theta) \text{ con } k(X) > 0$$

Para M observaciones (x_1, \dots, x_M) consideradas independientes e idénticamente distribuidas conforme una distribución normal $N(\mu, \sigma^2)$ se tiene que la función de densidad conjunta es:

$$f(X|\mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{-M/2} \exp\left(-\frac{1}{2} \sum_{i=1}^M \frac{(x_i - \mu)^2}{\sigma^2}\right)$$

Por tanto, el logaritmo de la función de verosimilitud será:

$$l(\mu, \sigma^2|X) = \ln[L(\mu, \sigma^2|X)] = \ln[K(X)] - \frac{M}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^M \frac{(x_i - \mu)^2}{\sigma^2}$$

La condición de primer orden para la maximización de esta función es:

$$\frac{\partial \ln L(\theta|X_{obs})}{\partial \theta} = 0$$

Es decir:

$$\frac{\partial \ln L(\mu, \sigma^2|X_{obs})}{\partial \mu} = 0, \quad \frac{\partial \ln L(\mu, \sigma^2|X_{obs})}{\partial \sigma^2} = 0$$

De esta forma se buscan aquellos valores de $\theta \in \Omega_\theta$ que más se acomodan a la muestra de datos X . Dado que los datos perdidos forman parte de X , el algoritmo debe estimar tanto μ y σ^2 como los valores perdidos. Para llegar a esta solución se suele proceder iterativamente. En la etapa de maximización se estiman los parámetros como si no hubiera datos perdidos (estos son reemplazados por estimaciones) y en la etapa de expectación se estiman los datos perdidos a partir de los conocidos y los previamente estimados. Luego se establece un ciclo repitiendo estas etapas hasta alcanzar un cierto criterio de convergencia preestablecido, como por ejemplo la ausencia de cambios significativos de los valores estimados. El resultado final resulta de la obtención de un máximo local de la función de verosimilitud (Nardo et al., 2008).

El algoritmo EM puede aplicarse en muchas situaciones en las que se desea estimar un conjunto de parámetros que describen una distribución de probabilidad subyacente con sólo una parte observada de los datos completos producidos por la distribución. Además, es muy simple de construir tanto en el aspecto conceptual como en el práctico, cada etapa tiene una interpretación estadística y siempre converge. El único inconveniente que presenta este método de imputación es que la convergencia puede resultar de larga duración cuando se parte de un conjunto de datos con muchos valores ausentes.

IMPUTACIÓN MÚLTIPLE

La imputación múltiple utiliza métodos de simulación Monte Carlo vía cadenas de Markov y sustituye los datos faltantes a partir de un número de simulaciones (normalmente ese número se ubica entre 3 y 10) (Rubin, 1987). La metodología consta de varias fases y en cada una de las simulaciones realizadas se aplican métodos estadísticos convencionales para analizar la matriz de datos completa. Posteriormente, se combinan los resultados para generar estimadores robustos, su error estándar e intervalos de confianza (Medina y Galván, 2007). En la figura 3.4 se muestran los pasos anteriormente descritos.

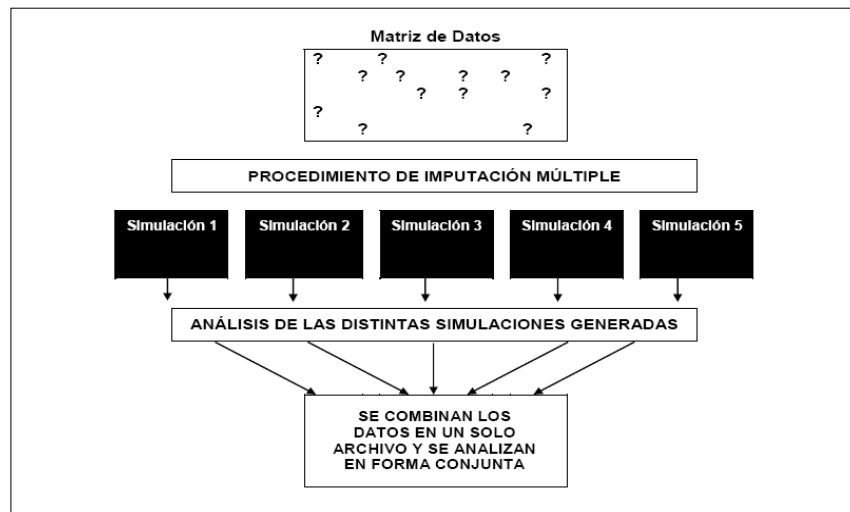


Figura 3.4. Esquema del algoritmo de Imputación Múltiple

A pesar de minimizar los problemas que pueden surgir en cualquier método de imputación simple, no se recomienda que se apliquen los métodos de imputación múltiple como la mejor opción estadística para la sustitución de datos faltantes (Little, 1988; Robins, Rotnitzky y Zhao, 1994). Cada situación es diferente y dependiendo del indicador que se analice, del porcentaje de respuesta y de su patrón de comportamiento es probable que se presenten situaciones en las que se obtengan mejores resultados con los métodos de imputación simple que con el que se acaba de nombrar. Por tanto, no se va a describir con más detalle este procedimiento de imputación. Un análisis detallado sobre el uso y manejo de esta técnica de imputación se puede encontrar en las referencias citadas en esta sección.

Durante la etapa de tratamiento de datos se debe tener presente que el uso de técnicas de imputación no puede sustituir por completo la información perdida, por lo que en primer lugar se debe tratar de recuperar los datos ausentes desde las fuentes originales de información. Además, el abuso de los métodos de imputación puede llevar a conclusiones que no reflejen la realidad de lo que se pretende medir.

Si se utilizan las técnicas de imputación, la aplicación del análisis de sensibilidad e incertidumbre es primordial para tener en cuenta los efectos de la imputación en las conclusiones. Finalmente hay que resaltar que es muy importante elegir el método de imputación adecuado para cada análisis, ya que todas las decisiones que se asuman en el tratamiento de datos se verán reflejadas en el IC construido y, por tanto, en las conclusiones finales de la investigación.

3.4.4. Análisis multivariante

Analizar la estructura subyacente de los datos sigue siendo un ‘arte’ (Nardo et al., 2008).

Durante las últimas décadas ha habido un crecimiento en la construcción de ICs diseñados por diversos organismos nacionales e internacionales. Desafortunadamente, muchas veces se seleccionan los indicadores simples de una forma aleatoria, prestando poca atención a las posibles relaciones entre ellos. Esto puede conducir a la construcción de ICs con un mensaje confuso y equívoco para las partes interesadas y el público en general. Algunos analistas bautizan esta situación como ‘indicadores ricos con escasa información’. Es por eso que la naturaleza de los datos subyacentes debe analizarse con mucho cuidado previamente a la construcción del IC. El uso de técnicas multivariantes permite evaluar la idoneidad del conjunto de datos y facilitar la comprensión de las elecciones metodológicas tomadas durante el proceso de construcción del índice.

La información debe agruparse y analizarse como mínimo en función de dos dimensiones del conjunto de datos: los indicadores individuales que son los que definen el concepto multidimensional que se pretende medir y las unidades de análisis que corresponden a cada una de las observaciones sobre las cuales se miden los indicadores simples definidos (Nardo et al., 2008). En las siguientes secciones se describen las técnicas más frecuentes en el contexto de ICs.

Información agrupada con respecto a los indicadores individuales

En primer lugar, el analista debe decidir si la estructura conceptual del IC está bien definida (Etapa 1) y si el conjunto de indicadores individuales disponibles es apropiado para describir el fenómeno a medir (Etapa 2). Esta decisión puede tomarse con la ayuda de la opinión de expertos y de la estructura estadística del conjunto de datos. Diferentes procedimientos estadísticos se pueden utilizar para explorar si las dimensiones del fenómeno están, desde un punto de vista estadístico, bien equilibradas en el IC. En caso de no ser así sería necesaria una revisión de los indicadores individuales o del marco conceptual teórico (esto no sería necesario si existe una justificación teórica de dicha agrupación de indicadores en dimensiones).

ANÁLISIS FACTORIAL (Factorial Analysis - FA)

El Análisis Factorial es una técnica estadística de modelación de datos cuya idea principal es explicar la variabilidad de Q indicadores observados en términos de un número menor m de variables no observadas llamadas factores, cuya influencia queda matizada por unos pesos o cargas incluyendo un término de error.

Los indicadores observados se modelan como combinaciones lineales de factores más expresiones de error (Uriel, 1995):

$$\begin{aligned} I_1 &= \alpha_{11}F_1 + \alpha_{12}F_2 + \cdots + \alpha_{1m}F_m + e_1 \\ I_2 &= \alpha_{21}F_1 + \alpha_{22}F_2 + \cdots + \alpha_{2m}F_m + e_2 \\ &\quad \dots \\ I_Q &= \alpha_{Q1}F_1 + \alpha_{Q2}F_2 + \cdots + \alpha_{Qm}F_m + e_Q \end{aligned}$$

donde:

I_i son los indicadores observados que se consideran tipificados o estandarizados, $i = 1, \dots, Q$

$\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im}$ son las cargas factoriales o saturaciones del indicador I_i en los factores F_1, F_2, \dots, F_m , $i = 1, \dots, Q$

F_1, F_2, \dots, F_m son los factores no correlacionados, cada uno de ellos con media cero y varianza la unidad

e_i son los errores independientes e idénticamente distribuidos con media cero, con $i = 1, \dots, Q$

Existen muchos procedimientos para resolver el modelo expuesto: factores de máxima verosimilitud, mínimos cuadrados no ponderados, mínimos cuadrados generalizables, etc. Pero el procedimiento más usado para extraer los m factores es el análisis de componentes principales, PCA, puesto que permite la construcción de pesos que representan la información contenida en los indicadores simples. Este método de extracción es el que más se aplica en la construcción de ICs y, por tanto, el que se va a describir a continuación. Nótese que diferentes procedimientos para la resolución del modelo implican diferentes valores de los factores extraídos y diferentes pesos para los indicadores simples influyendo, por tanto, en la puntuación del IC final.

ANÁLISIS DE COMPONENTES PRINCIPALES (Principal Component Analysis - PCA)

La técnica de Análisis de Componentes Principales fue descrita por Karl Pearson en 1901. Una descripción de su metodología de cálculo fue introducida más tarde por Hotelling en 1933. El objetivo de esta técnica es explicar la mayor parte de la variabilidad total observada en un conjunto de variables con el menor número de componentes posible (Uriel, 1995). Esto es posible transformando las variables correlacionadas en un nuevo conjunto de variables no correlacionadas, denominadas factores o componentes principales, relacionadas con las

variables originales mediante una transformación lineal y ordenadas de forma decreciente según el porcentaje de variabilidad que explican. Dicho de otro modo, la técnica utilizada en el análisis de componentes principales consiste en proyectar la nube de observaciones sobre un subespacio afín de dimensión menor, determinado de tal manera que la nube proyectada se deforme lo menos posible.

El análisis de componentes principales está relacionado con el análisis factorial, pero existen ciertas diferencias; i) los componentes principales se construyen para explicar las varianzas, mientras que los factores se construyen para explicar las covarianzas o correlaciones entre las variables, ii) el análisis de componentes principales es una técnica descriptiva, mientras que el análisis factorial presupone un modelo estadístico formal de generación de datos como se ha descrito anteriormente (Peña, 2002).

Para describir la técnica PCA supóngase que se tienen Q indicadores en el análisis $I_i, i = 1, \dots, Q$ medidos sobre n unidades de análisis (Uriel, 1995; Nardo et al., 2005). Sea X la forma matricial que representa los datos del estudio:

$$X = \begin{pmatrix} I_{11} & \dots & I_{Q1} \\ \dots & \dots & \dots \\ I_{1n} & \dots & I_{Qn} \end{pmatrix} \in R^{n \times Q}$$

La matriz de covarianza muestral CM de los datos originales es:

$$CM = E[(X - E[X])(X - E[X])'] = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \dots & \dots & \dots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{pmatrix} \in R^{n \times n}$$

Para evitar que algún indicador tenga una influencia indebida en las componentes principales se suele estandarizar la matriz de variables originales. En este caso, la matriz de varianzas-covarianzas se convierte en la matriz de correlaciones:

$$R = \begin{pmatrix} 1 & \dots & r_{1Q} \\ \dots & \dots & \dots \\ r_{Q1} & \dots & 1 \end{pmatrix} \in R^{Q \times Q} \text{ con } r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}, 1 \leq i, j \leq Q$$

Las componentes principales pueden estimarse a partir de cualquiera de estas dos matrices que son las que proporcionan información acerca de la relación en la variabilidad observada en las variables cuando son tomadas de dos en dos.

Las componentes principales son un conjunto de variables $Z_j, j = 1, \dots, Q$, ortogonales entre sí que surgen de una combinación lineal de las variables originales con la propiedad de contener en conjunto la misma varianza total que el conjunto original.

$$\begin{aligned}
Z_1 &= a_{11}I_1 + a_{12}I_2 + \dots + a_{1Q}I_Q \\
Z_2 &= a_{21}I_1 + a_{22}I_2 + \dots + a_{2Q}I_Q \\
&\dots \\
Z_Q &= a_{Q1}I_1 + a_{Q2}I_2 + \dots + a_{QQ}I_Q
\end{aligned}$$

La primera componente principal retendrá la máxima porción de la varianza del conjunto de las variables originales, la segunda retendrá el máximo de la varianza restante y así sucesivamente hasta la última componente principal que contendrá el resto de varianza no incluida en las componentes principales anteriores.

La primera componente se expresará como la combinación lineal siguiente:

$$\begin{pmatrix} Z_{11} \\ \dots \\ Z_{1n} \end{pmatrix} = \begin{pmatrix} I_{11} & \dots & I_{Q1} \\ \dots & \dots & \dots \\ I_{1n} & \dots & I_{Qn} \end{pmatrix} \begin{pmatrix} a_{11} \\ \dots \\ a_{1Q} \end{pmatrix}; Z_1 = X * a_1$$

El vector a_1 se obtiene maximizando la varianza de Z_1 :

$$\begin{aligned}
\max Var(Z_1) \text{ con } Var(Z_1) &= \frac{\sum_{i=1}^n Z_{1i}^2}{n} = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} a_1' X' X a_1 = a_1' \left[\frac{1}{n} X' X \right] a_1 \\
s. a \sum_{j=1}^Q a_{1j}^2 &= 1
\end{aligned}$$

- Si las variables están normalizadas, $\left[\frac{1}{n} X' X \right] = R$
- Si las variables están expresadas como desviaciones típicas alrededor de la media, $\left[\frac{1}{n} X' X \right] = CM$

Sin pérdida de generalidad, supóngase la segunda situación. Por tanto, para maximizar la varianza de Z_1 , se construye el lagrangiano:

$$\mathcal{L} = a_1' * CM * a_1 - \lambda(a_1' a_1 - 1)$$

cuya condición de primer orden es:

$$\frac{\partial \mathcal{L}}{\partial a_1} = 2CM * a_1 - 2\lambda a_1 = 0, \quad (CM - \lambda I) * a_1 = 0$$

y dado que a_1 es un vector no nulo, se tiene que λ es el autovalor de la matriz de covarianzas y a_1 su autovector.

El resto de las componentes se obtienen aplicando el mismo procedimiento pero añadiendo una nueva restricción de ortogonalidad respecto de las componentes anteriores ya calculadas.

En resumen, el PCA trata de encontrar los autovalores λ_j de la matriz de covarianza CM de los datos originales que son las varianzas de las componentes principales. Además se cumple:

$$\lambda_1 + \lambda_2 + \dots + \lambda_Q = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_Q^2$$

Con esto se obtienen Q componentes principales, tantas componentes como variables del análisis. El siguiente paso es seleccionar $P < Q$ componentes que conserven la mayor cantidad de varianza acumulada de los datos originales.

Los coeficientes de correlación entre las componentes principales Z_j y los indicadores I_i se llaman **cargas factoriales** o **puntos-variable**. En el caso en que los indicadores no estén correlacionados, las cargas factoriales coinciden con los pesos a_{ij} . Cuanto mayor sea la carga factorial de un indicador con respecto a una componente significa que la relación entre ambos es alta.

Análogamente al coeficiente de correlación de Pearson, el cuadrado de la carga factorial del indicador I_i , a_{ij}^2 , se denomina **Comunalidad** e indica la proporción de variabilidad del indicador i que queda explicada por la componente principal j . De esta forma, la **comunalidad del indicador I_i** , h_i^2 , se define como la suma de todas las comunalidades de cada factor respecto al indicador (Cuadras, 2006):

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{iQ}^2$$

Las puntuaciones de cada una de las observaciones en cada componente principal se llaman **puntos-individuo**. Los puntos-individuo para una observación particular respecto a un componente principal se calculan estandarizando el valor para cada indicador, multiplicándolo por la correspondiente carga factorial de ese componente principal y sumando los productos.

La falta de correlación en las componentes principales nos indica que cada una de ellas mide “dimensiones estadísticas” diferentes en los datos. No siempre la aplicación de PCA reduce el número de indicadores originales en un número de variables latentes menor. Esto último ocurre cuando los indicadores originales no están correlacionados. Se conseguirá reducir notablemente el número de componentes principales cuando los indicadores originales estén altamente correlacionados, tanto si es positiva como negativamente.

Existen diversos criterios para decidir el número de componentes principales que se deben extraer de tal forma que se conserve la mayor cantidad de varianza acumulada de los datos originales y posibilite su posterior interpretación. En la tabla 3.3 se describen algunos de ellos aplicados a la rama de ciencias sociales (Dunteman, 1989; Hair et al., 2007).

Criterios de selección del número de componentes principales	
1. Criterio de Kaiser o de la raíz latente	Es el criterio que se utiliza más frecuentemente. Se basa en retener aquellos factores que tienen raíces latentes o autovalores que exceden de la media del conjunto de todos los autovalores. Como el autovalor asociado a un factor es su varianza, este criterio selecciona aquellos autovalores que explican un mayor porcentaje de la varianza total. En el caso de variables tipificadas, se seleccionarían aquellas componentes cuyos autovalores excedan de 1. Los factores con autovalores menores que 1 explican menos varianza que la contenida en una variable y, por tanto, no son significativos
2. Criterio de Joliffe	Este criterio se basa en el mismo razonamiento que el criterio de Kaiser. La única diferencia es que en el criterio de Joliffe se seleccionan aquellos componentes cuyos autovalores exceden de 0,7
3. Criterio de contraste de caída o test del codo de Castell	Se basa en identificar el número óptimo de factores que se pueden extraer antes de que la cantidad de la varianza única, aquella asociada con una variable específica, empiece a dominar la estructura de la varianza común, aquella que una variable comparte con el resto de variables. Para ello se trazan los autovalores de manera decreciente en función del número de factores (gráfico de sedimentación) escogiéndose aquellas componentes hasta el punto en que la curva decreciente converge hacia una línea horizontal, lo cual indica que a partir de ese punto la varianza explicada no aporta mucho más
4. Criterio del porcentaje de la varianza explicada	Este criterio se basa en retener aquellos factores hasta obtener un porcentaje acumulado específico de la varianza total extraída, asegurando que explican una cantidad específica de la varianza. No existe un umbral predefinido para todas las aplicaciones. Mientras que en el campo de las ciencias naturales se seleccionan aquellos factores que expliquen el 95% de la varianza, en ciencias sociales el umbral se reduce al 60% de la varianza
5. Criterio de comprensibilidad	Se aplica cuando la construcción de las componentes da lugar a algún tipo de interpretación que sea entendible en términos de las clases que se constituyen

Tabla 3.3. Criterios de selección del número de componentes principales

Una herramienta que se suele emplear después de seleccionar el número de componentes y que mejora la interpretación de los resultados es la rotación de factores. Esta herramienta se basa en girar los ejes de referencia hasta alcanzar una determinada posición. La rotación de factores hace que se redistribuya la varianza de los primeros factores a los restantes. De esta forma, se consigue un patrón de factores más simple y más fácilmente interpretable (Hair et al., 2007). Existen varios métodos de rotación pero, según la literatura, los métodos más comunes son la **rotación ortogonal “Varimax”** y la **rotación oblicua “Oblimin”** (Nardo et al., 2005;

Hair et al., 2007). No existen reglas concretas que indiquen la selección de una técnica de rotación. La rotación afecta a las cargas factoriales de las variables y puede ocurrir que los grupos obtenidos con PCA sin aplicar rotación no sean los mismos que al aplicar rotación.

En resumen, los pasos a seguir para aplicar un análisis PCA/FA como un análisis exploratorio son:

1. Cálculo de la matriz de correlaciones: si las correlaciones entre los indicadores simples son bajas es muy probable que estos no compartan factores comunes
2. Identificar el número de factores necesarios para representar el conjunto de datos y el método para calcularlos
3. Aplicar, en el caso que sea necesario, una rotación sobre los factores para facilitar la interpretación de los resultados

Para poder aplicar el análisis PCA/FA se deben cumplir una serie de supuestos que se van a enumerar a continuación y que a su vez pueden encontrarse en cualquier libro de texto (Hair et al., 2007; Nardo et al., 2008). Sin embargo, estos supuestos se descuidan, con frecuencia, en el desarrollo de ICs.

Supuestos básicos para la aplicación del PCA/FA
<ul style="list-style-type: none"> - Poseer un número de observaciones suficientemente grande. La cuestión sobre cuántas observaciones mínimas se deben tener en cuenta para poder aplicar un análisis PCA/FA está científicamente por resolver y existen muchas opiniones que difieren. A continuación se nombran algunas reglas empíricas propuestas por diferentes autores: <ul style="list-style-type: none"> - Regla del 10: disponer de al menos 10 observaciones por cada variable - Ratio 3:1: el número de observaciones debe ser, como mínimo, el triple que el de variables - Ratio 5:1: el número de observaciones debe ser, como mínimo, el quintuple que el de variables - Regla del 100: el número de observaciones debe ser cinco veces el número de variables y superiores a 100 - Regla del 150: el número de observaciones debe ser mayor a 150 cuando hay muy pocas variables correlacionadas - Regla del 200: disponer de más de 200 observaciones sin tener en cuenta las variables - Regla de la significación: tener 51 casos más que el número de variables con el fin de poder realizar la prueba chi-cuadrado - Que no haya sesgo de selección de las variables. La exclusión de variables relevantes y la inclusión de variables irrelevantes afectará a la matriz de covarianzas y, por lo tanto, a la representatividad del resultado obtenido - Que no haya datos atípicos. La presencia de datos atípicos puede afectar la interpretación del resultado de un análisis PCA/FA - Linealidad. El PCA/FA es una técnica basada en el álgebra lineal y, por tanto, la relación entre las variables debe ser lineal - Normalidad multivariante. Si se supone que las variables están distribuidas a partir de funciones de distribución diferentes, será más complicado realizar pruebas de contraste estadístico, pues generalmente las herramientas de software solo contemplan las variables basadas en el supuesto de normalidad - Correlación fuerte entre las variables. Al aplicar un PCA/FA a una matriz de covarianzas con correlaciones bajas se obtienen tantos factores como variables originales se tengan y, por tanto, no se podrá reducir la dimensionalidad del conjunto inicial de datos. Pruebas para comprobar la correlación entre las variables: medida de Kaiser-Meyer-Olkin, test de esfericidad de Bartlett, correlación observada y reproducida, matriz de correlaciones, matriz de correlaciones anti-imagen, etc.

Tabla 3.4. Supuestos para la aplicación del ACP/AF

COEFICIENTE ALPHA DE CRONBACH

Otra forma de investigar el grado de correlación entre los indicadores es utilizando el coeficiente alpha de Cronbach (c -alpha) que es el coeficiente más utilizado para analizar la consistencia interna o fiabilidad de un conjunto de indicadores de un modelo o una encuesta. Esta consistencia se refiere a que los indicadores miden las mismas cosas y, por tanto, están

altamente correlacionados. Indica cuán bien representada está la información de múltiples indicadores en un sólo factor (Cronbach, 1951).

Es una medida de fiabilidad muy sencilla de calcular y presenta la gran ventaja de estar disponible en la mayoría de programas estadísticos. El coeficiente alpha de Cronbach se puede calcular como:

$$\alpha = \left(\frac{Q}{Q-1} \right) \frac{\sum_{i \neq j} cov(I_i, I_j)}{var(I_o)} = \left(\frac{Q}{Q-1} \right) \left(1 - \frac{\sum_j var(I_j)}{var(I_o)} \right) \quad i, j = 1, \dots, Q$$

donde Q es el número de indicadores simples disponibles y $I_o = \sum_{j=1}^Q I_j$ es la suma de todas las puntuaciones de los indicadores simples.

El coeficiente c-alpha mide la fracción de variabilidad total de la muestra de indicadores debido a su correlación. Toma valores entre 0 y 1, y sirve para comprobar si un factor recopila información de los indicadores.

El coeficiente c-alpha no es un test estadístico, sino una medida de fiabilidad basada en las correlaciones de los indicadores simples. Un coeficiente elevado significa que los indicadores considerados miden correctamente el fenómeno latente. Pero el coeficiente c-alpha no es una medida unidimensional. Un conjunto de indicadores puede tener un c-alpha elevado y seguir siendo multidimensional. Esto ocurre cuando existen diferentes conglomerados o dimensiones con elevada correlación dentro de ellos aunque las dimensiones entre sí no estén altamente correlacionadas. Respecto al umbral predefinido para considerar un coeficiente aceptable existen diversas opiniones. Nunnally (1978) sugirió un coeficiente de 0,7 como umbral de fiabilidad aceptable, mientras que otros autores consideran 0,75 o 0,80 como valor de corte. El umbral predefinido depende de las disciplinas en las que se trabajan.

Si las varianzas de los indicadores simples difieren ampliamente entre ellas, previamente al cálculo del coeficiente c-alpha se suelen estandarizar los indicadores para conseguir una desviación típica común e igual a la unidad.

Un ejercicio interesante es observar cómo varía el coeficiente c-alpha con la exclusión de alguno de los indicadores. Esto ayuda a detectar la existencia de dimensiones en los indicadores simples. Si su valor se incrementa con la exclusión se puede afirmar que el indicador eliminado no está muy correlacionado con el resto.

Información agrupada con respecto a las unidades de análisis

ANÁLISIS DE CONGLOMERADOS

El análisis de conglomerados o análisis *cluster* es la denominación de un grupo de técnicas multivariantes cuyo objetivo es agrupar un serie de objetos en conglomerados (grupos o *clusters*) a partir de la información que proporcionan los propios individuos (ya sean entidades, países, universidades, profesores,...) (Hair et al., 2007). Los conglomerados obtenidos deben ser heterogéneos entre ellos y homogéneos dentro de ellos, por tanto, tendrán la particularidad de que los objetos que pertenezcan a una misma categoría sean lo más parecido posible entre ellos y a la vez lo más diferente posible a los de otras categorías. Además, los grupos deben ser exhaustivos (deben comprender a todos los objetos) y exclusivos (un elemento no podrá pertenecer a la vez a dos conglomerados).

El análisis de conglomerados también se utiliza en el desarrollo de ICs para agrupar la información en unidades de análisis que presentan características similares en los diferentes indicadores individuales. Además, el análisis de conglomerados se puede tratar como:

- Herramienta de diagnóstico que sirve para explorar el impacto de las elecciones metodológicas realizadas durante la fase de construcción del IC
- Método de disseminación de información del IC que no pierde las dimensiones de los indicadores individuales
- Método de selección de grupos de unidades de análisis para la imputación de datos faltantes

El análisis de conglomerados engloba numerosos procedimientos que se clasifican en jerárquicos y no jerárquicos dependiendo de si los grupos se establecen o no mediante una jerarquía de partes.

Para aplicar un análisis de conglomerados se parte de una matriz de distancias o similitudes entre las unidades de análisis a agrupar. Dos objetos son más parecidos cuanto más similares son o cuanto menos distancia hay entre ellos. Estas distancias se calculan a partir de una matriz de datos cuantitativos o cualitativos.

En el caso de variables o indicadores cuantitativos las distancias que más se utilizan al aplicar las técnicas de análisis de conglomerados se muestran en la tabla 3.5 (Schuschny y Soto, 2009).

Distancia	Fórmula	Ventajas e Inconvenientes
Distancia euclídea	$d(x, y) = \left(\sum_{i=1}^Q (x_i - y_i)^2 \right)^{1/2}$ x, y representan dos unidades de análisis	<i>Ventaja:</i> distancia que no se ve afectada por la inclusión de nuevas unidades de análisis, por tanto, no se ve afectada por <i>outliers</i> <i>Inconveniente:</i> su valor depende de la escala o unidad de medida de las variables
Distancia euclídea al cuadrado	$d(x, y) = \sum_{i=1}^Q (x_i - y_i)^2$	Al no calcularse la raíz cuadrada las variables más alejadas tendrán más peso
Distancia euclídea al cuadrado normalizada	$d(x, y) = \sum_{i=1}^Q \left(\frac{x_i - y_i}{\sigma_i} \right)^2$	Esta medida soluciona el problema de la medida anterior y, por tanto, le resta peso a las variables más alejadas
Distancia de Chebychev	$d(x, y) = \max_i x_i - y_i $	Medida útil cuando se desea definir elementos como “diferentes” si hay una diferencia apreciable en cualquiera de los mismos
Distancia de Manhattan (o city-block o de Hamming)	$d(x, y) = \sum_{i=1}^Q x_i - y_i $	Distancia que atenúa la presencia de valores atípicos
Distancia de Mahalanobis	$D^2 = (x_i - x_j)' V^{-1} (x_i - x_j)$ siendo V la matriz de varianzas-covarianzas de las variables que intervienen	<i>Ventaja:</i> distancia que tiene en cuenta la correlación de las variables <i>Inconveniente:</i> no cumple la desigualdad triangular que es una condición que debe cumplir cualquier distancia definida
Distancia de Minkowski	$d(x, y) = \left(\sum_{i=1}^Q (x_i - y_i)^s \right)^{1/r}$	Distancia ventajosa cuando se desea incrementar o atenuar el valor de una variable. El valor s mide las diferencias de las variables mientras que r controla el peso de la distancia entre unidades de análisis diferentes
Proporción de discrepancias	$d(x, y) = \#veces\ que\ x_i \neq y_i$	Distancia útil cuando se trabaja con información discreta o categorizada

Tabla 3.5. Distancias más empleadas para las técnicas de conglomerados con datos cuantitativos

En el caso de variables cualitativas se pueden calcular medidas de similitud a partir de las variables binarias asociadas. Toda variable categórica se puede expresar mediante variables binarias con valores 0, no-posección del atributo, y 1, posesión del atributo. El coeficiente de similaridad, como su nombre indica, da una medida del grado de semejanza entre dos objetos en relación con los indicadores.

Los métodos jerárquicos aglomerativos o ascendentes comienzan con tantos grupos como individuos, se van formando grupos entre los individuos más parecidos y termina con un solo grupo que integra a todos los elementos. Una vez elegido el tipo de distancia a utilizar se debe seleccionar una medida de la distancia entre *clusters*, $\delta(C_i, C_j)$, siendo i, j dos *clusters* cualesquiera. Las más comunes son (Späth, 1980):

- **Distancia mínima (enlace simple):** la distancia entre dos grupos se determina por la distancia entre los dos vectores (o unidades de análisis) más cercanos pertenecientes a grupos distintos. En términos matemáticos, dados dos clusters C_i, C_j ,

$$\delta(C_i, C_j) = \min\{d(x, y), x \in C_i, y \in C_j\}$$

- **Distancia máxima (enlace completo):** la distancia entre dos grupos se determina por la mayor distancia entre dos unidades de análisis pertenecientes a grupos distintos. Método útil cuando las unidades de análisis forman claramente grupos definidos. Dados dos clusters C_i, C_j ,

$$\delta(C_i, C_j) = \max\{d(x, y), x \in C_i, y \in C_j\}$$

- **Método del centroide:** calcula la proximidad entre dos conglomerados como la distancia entre sus centroides. Si $\bar{x}_i = \frac{1}{n_i} \sum_{x \in C_i} x$; $\bar{y}_j = \frac{1}{n_j} \sum_{y \in C_j} y$, entonces,

$$\delta(C_i, C_j) = d(\bar{x}_i, \bar{y}_j)$$

- **Distancia media entre grupos (enlace promedio ponderado):** la distancia entre grupos se define como el promedio de las distancias entre todos los pares de elementos de ambos grupos usando el tamaño de los grupos como peso,

$$\delta(C_i, C_j) = \frac{1}{n_i n_j} \sum_{x \in C_i; y \in C_j} d(x, y)$$

- **Método de Ward** (Ward, 1963; Pérez, 2005): para este método se considera la distancia euclídea al cuadrado como medida de disimilitud. Sea $d(x_i, x_j)^2 = \|x_i - x_j\|^2$ la distancia entre los puntos x_i y x_j . La varianza total del conjunto de puntos se define como $I = \sum_i m_i \|x_i - G\|^2$, siendo G el centro de gravedad de los puntos dados con masa respectiva m_i . Sea G_q el centro de gravedad del q -ésimo conglomerado al dividirse el conjunto de individuos en q grupos y m_q su respectiva masa. La varianza total o inercia se puede descomponer de la siguiente forma:

$$I = \sum_q m_q \|G_q - G\|^2 + \sum_q \sum_{i \in q} m_i \|x_i - G_q\|^2$$

Sean x_i y x_j dos elementos de masa m_i y m_j respectivamente que se unen en un elemento x cuya masa es $m = m_i + m_j$, con $x = \frac{(m_i x_i + m_j x_j)}{(m_i + m_j)}$. Se descompone la varianza I_{ij} de x_i y x_j con respecto a G por la ecuación:

$$I_{ij} = m_i \|x_i - x\|^2 + m_j \|x_j - x\|^2 + m \|x - G\|^2$$

La reducción de la varianza se calcula reemplazando x por su valor como función de x_i y x_j :

$$\Delta I_{ij} = \left(\frac{m_i m_j}{m_i + m_j} \right) \|x_i - x_j\|^2$$

Por tanto, el objetivo de este método es encontrar los individuos x_i y x_j con la condición de que hagan mínima ΔI_{ij} .

Los métodos no jerárquicos, también conocidos como partitivos o de optimización, tienen por objeto realizar una sola partición de los individuos en k grupos, por lo que el número de clusters se fija a priori. Existe gran variedad de algoritmos de este tipo. Uno de los algoritmos más conocidos es el de k -medias de MacQueen (1967).

Cuando el número de variables es muy grande o cuando se sabe de antemano que algunas de ellas no contribuyen a identificar la estructura de conglomerados del conjunto de datos, se pueden aplicar secuencialmente modelos continuos y discretos. Normalmente, se suele hacer un análisis de componentes principales y luego aplicar un algoritmo de agrupamiento sobre las puntuaciones objetivas de las primeras componentes principales, es lo que se llama “análisis tándem”. Sin embargo, se debe tener precaución al utilizar estos métodos. Si el análisis de componentes principales o el análisis factorial identifican dimensiones que no necesariamente nos ayudan a comprender la estructura de agrupaciones en los datos pueden enmascarar la información taxonómica.

Un método alternativo que combina el análisis de conglomerados k -medias y el análisis factorial fue desarrollado por Vichi y Kiers (2001) con el objetivo de identificar la mejor partición de las unidades de análisis descritas por las mejores combinaciones lineales ortogonales de las variables (factores) de acuerdo al criterio de mínimos cuadrados. Esta metodología ha sido bautizada como análisis factorial k -medias y ha tenido una amplia aplicación en muchos campos de trabajo por su doble objetivo: reducir la dimensionalidad del conjunto de datos y agrupar las unidades de análisis según los objetivos perseguidos.

En la tabla 3.6 se muestran las ventajas e inconvenientes de la aplicación de las técnicas multivariantes descritas, así como algunos ejemplos de ICs donde se usan estas técnicas (Nardo et al., 2005).

Técnica Multivariante	Ventajas	Inconvenientes	ICs en los que se aplica la técnica
PCA/FA	<ul style="list-style-type: none"> - Reduce el conjunto de indicadores simples preservando al mismo tiempo la máxima proporción posible de la variabilidad total del conjunto de datos originales - Las puntuaciones más altas se asignan a los indicadores simples que tienen la mayor variabilidad entre las unidades de análisis, una característica deseable para las comparaciones entre unidades de análisis 	<ul style="list-style-type: none"> - Las correlaciones no representan necesariamente la influencia real de los indicadores simples sobre el fenómeno que se está midiendo - Sensible a las modificaciones en la base de datos, como por ejemplo la incorporación de nuevas unidades de análisis - Sensible a la presencia de valores atípicos, lo que puede causar una variabilidad errónea en los datos - Sensible a los problemas con muestras pequeñas que son particularmente relevantes cuando el estudio se centra en un conjunto limitado de unidades de análisis - Minimiza la contribución de los indicadores simples que no tienen una alta correlación con otros indicadores individuales 	<ul style="list-style-type: none"> - <i>Environmental Sustainability Index</i> (World Economic Forum, 2001) - <i>General Indicator of Science & Technology</i> (National Institute of Science and Technology Policy, 2004) - <i>Internal Market Index</i> (European Commission-DG MARKT, 2001)
Coefficiente alpha-Cronbach	<ul style="list-style-type: none"> - Mide la consistencia interna del conjunto de indicadores individuales, es decir, cómo de bien describen un constructo unidimensional. Por lo que resulta de gran utilidad para clasificar objetos similares 	<ul style="list-style-type: none"> - Las correlaciones no representan necesariamente la influencia real de los indicadores simples sobre el fenómeno que se está midiendo. - Es significativo sólo cuando el IC se calcula como una “escala” (es decir, como la suma de los indicadores individuales) 	<ul style="list-style-type: none"> - <i>Compassion Fatigue</i> (Boscarino, Figley y Adams, 2004) - <i>Secondary trauma</i> (Boscarino, Figley y Adams, 2004) - <i>Job burnout</i> (Boscarino, Figley y Adams, 2004) - <i>Success of software process implementation</i> (El Emam et al., 1998)
Técnicas de clasificación	<ul style="list-style-type: none"> - Ofrece un método distinto para agrupar las unidades de análisis y da una idea de la estructura del conjunto de datos 	<ul style="list-style-type: none"> - Herramienta puramente descriptiva. Puede que no sea transparente si las elecciones metodológicas tomadas durante el análisis no están claramente justificadas - Se debe identificar el número óptimo de grupos a considerar. La alternativa óptima es parte de la decisión del analista 	<ul style="list-style-type: none"> - <i>Technology Achievement Index</i> (United Nations Development Programme, 2001)

Tabla 3.6. Ventajas e inconvenientes de la aplicación de las técnicas de análisis multivariante

3.4.5. Normalización de los datos

La normalización de los datos es un paso previo a cualquier tipo de agregación de los indicadores simples. Los objetivos de las técnicas de normalización son:

- Ajustar para que los datos no tengan diferentes unidades de medida
- Ajustar para que los datos no tengan diferentes rangos de variación
- Ajustar en el caso en que los datos sigan una distribución asimétrica o ante la presencia de datos atípicos

Los dos primeros objetivos se deben cumplir con el fin de no mezclar diferentes tipos de datos en el análisis. El tercer objetivo es necesario que se cumpla sólo en el caso en que se esté usando un modelo que asuma normalidad en los datos. La desviación de la normalidad está provocada, a menudo, por la presencia de valores atípicos que alteran estadísticos descriptivos como la media, la desviación estándar o el coeficiente de correlación causando, posteriormente, malas interpretaciones.

Hay que tener en cuenta que si se aplica cualquier transformación con el fin de evitar la presencia de valores atípicos o que la distribución sea asimétrica, los datos originales se van a ver alterados por dicha transformación. Por eso hay que elegir cuidadosamente la opción de transformar los datos y hacerlo en el caso en que sea inevitable. Si $|\text{coeficiente de asimetría}| > 1$ y $\text{curtosis} > 3,5$ en alguno de los indicadores se recomienda aplicar algún método de transformación, puesto que la media de una distribución asimétrica no es un buen indicador para emplearse en estos casos. Además, al corregir la asimetría de una distribución muestral, valores originales aparentemente atípicos se encontrarán más cerca del grueso de los datos y, por tanto, se facilitará la identificación de los verdaderos datos atípicos.

Existen diferentes tipos de normalización de los datos (Freudenberg, 2003; Jacobs, Smith y Goddard, 2004). La elección de una u otra metodología dependerá de las características de cada indicador y del juicio experto del analista. Para la descripción de los diferentes métodos de normalización se va a considerar la siguiente notación:

x_{qc}^t : puntuación del indicador simple q para la unidad de análisis c en el momento de tiempo t para $q = 1, \dots, Q$ y $c = 1, \dots, M$, donde Q es el número de indicadores simples y M el número de unidades de análisis/observaciones.

Ranking

$$I_{qc}^t = \text{Rank}(x_{qc}^t)$$

- Es la técnica de normalización más simple
- Se puede usar tanto con datos cualitativos (ordinales) como con datos cuantitativos
- Los indicadores tienen las mismas unidades de medida
- Este método no se ve afectado por valores atípicos
- Para estudios dependientes del tiempo, esta normalización se aplica en cada momento del tiempo a analizar y permite comparar el rendimiento de cada una de las unidades de análisis a lo largo del tiempo y en forma de posiciones relativas (Cribari-Neto, Jensen y Novo, 1999)
- El rango de variación es el mismo para todos los indicadores $[1, M]$ siendo M el número de unidades de análisis
- La desventaja de este método de normalización es que se pierde la información de las diferencias entre las unidades de análisis cuando se agrega a nivel absoluto

ICs en los que se aplica el método Ranking para la normalización de los datos:

- *The Development and application of Information and Communication Technology across countries Index* (Fagerberg, 2001)
- *Medicare Study on Healthcare Performance across the United States Index* (Jencks, Huff y Cuerdon, 2003)

Estandarización (método z-score)

$$I_{qc}^t = \frac{x_{qc}^t - x_{qc=\bar{c}}^t}{\sigma_{qc=\bar{c}}^t}$$

donde $x_{qc=\bar{c}}^t$ representa la media del indicador q para todas las unidades de análisis en el momento t y $\sigma_{qc=\bar{c}}^t$ la desviación típica del indicador q para todas las unidades de análisis en el momento t .

- Se puede aplicar únicamente a indicadores cuantitativos
- Transforma los indicadores a una escala adimensional con media de 0 y desviación típica la unidad manteniendo las distancias relativas, puesto que se trata de una transformación lineal
- Los valores de cada unidad de análisis quedan estandarizados respecto a una misma distribución. Dado que la media es cero se evita la presencia de distorsiones debidas a las diferencias entre las medias de los indicadores en la fase de agregación
- No ajusta los datos a un mismo rango de variación. Si los indicadores provienen de una distribución aproximadamente normal, el 95% de los valores transformados estarán dentro del rango $[-2,2]$

- Los indicadores con valores extremos tendrán mayor efecto sobre el IC. Esto puede que sea deseable si la intención es premiar el comportamiento excepcional de los indicadores, es decir, si se considera mejor resultado cuando el valor de un indicador es muy alto respecto a la media de las puntuaciones de todos los indicadores. Este efecto se puede corregir en la fase de la agregación, bien excluyendo la mejor y peor puntuación de los indicadores simples en el IC o asignando ponderaciones diferentes basadas en la conveniencia de las puntuaciones de los indicadores simples
- Cuando se trabaja con variables registradas a lo largo de los años se estima la media y la desviación estándar en referencia a un año base que suele ser el primer año del que se dispone de la información

ICs en los que se aplica el método z-score para la normalización de los datos:

- *The Environmental Sustainability Index* desarrollado en la Universidad de Yale (World Economic Forum, 2001)
- *The Internal Market Index* (European Commission-DG MARKT, 2001)

Un método alternativo a la normalización z-score es la normalización geométrica definida de la siguiente forma:

$$I_{qc}^t = \frac{x_{qc}^t - x_{qc=\bar{c}}^t}{\sqrt{\sum_{i=1}^M (x_{qi}^t - x_{qc=\bar{c}}^t)^2}}$$

Este tipo de normalización tiene el mismo comportamiento y las mismas propiedades que la normalización z-score excepto el rango de variación que en este caso es [-1,1].

Ambos tipos de normalización reducen el valor de los datos atípicos frente a distribuciones asimétricas, pero no los eliminan. Existe una familia de transformaciones de indicadores llamadas transformaciones Box-Cox muy utilizadas para resolver los problemas de falta de normalidad y de igualdad de varianzas en los indicadores.

Considerando que $x > 0 \forall x \in X$, siendo X el indicador que se pretende transformar, la transformación Box-Cox (1964) se define de la siguiente forma:

$$x \rightarrow y = T(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(x) & \text{si } \lambda = 0 \end{cases}$$

Una de las ventajas que presenta esta familia de transformaciones es que es muy flexible, puede hacer más compactos los valores altos si $\lambda < 1$ y puede estrecharlos si $\lambda > 1$, incluyendo la transformación logaritmo como un caso especial.

Para estimar el valor de λ con el objetivo de obtener un conjunto de indicadores distribuidos lo más parecido posible a la distribución normal se calcula la siguiente función de máxima verosimilitud:

$$U(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda \hat{x}^{(\lambda-1)}} & \text{si } \lambda \neq 0 \\ \hat{x} \log(x) & \text{si } \lambda = 0 \end{cases} \text{ con } \hat{x} = \prod_{i=1}^M x_i^{\frac{1}{M}}$$

Se estima el valor de λ que maximiza la función de verosimilitud:

$$L(\lambda) = -\frac{M}{2} \log \left[\sum_{i=1}^M (U_\lambda(x_i) - \bar{U}_\lambda)^2 \right]$$

Con esto se consigue que el indicador transformado se aproxime a una distribución normal. A veces, puede haber indicadores para los que ninguna transformación es satisfactoria.

Manly (1976) propuso una transformación, versión exponencial, similar a la familia Box-Cox que admite valores negativos y es útil para distribuciones asimétricas:

$$x \rightarrow y = \begin{cases} \frac{e^{\lambda x} - 1}{\lambda} & \text{si } \lambda \neq 0 \\ x & \text{si } \lambda = 0 \end{cases}$$

En 1980 otra propuesta fue introducida por John y Draper (1980) llamada “transformación módulo” para distribuciones asimétricas:

$$x \rightarrow y = \begin{cases} \text{sgn}(x) \frac{(|x| + 1)^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \text{sgn}(x) \log(|x| + 1) & \text{si } \lambda = 0 \end{cases} \text{ con } \text{sgn}(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Re-escalamiento (método min-max)

$$I_{qc}^t = \frac{x_{qc}^t - \min_c(x_q^t)}{\max_c(x_q^t) - \min_c(x_q^t)}$$

donde $\max_c(x_q^t)$ y $\min_c(x_q^t)$ es el máximo y mínimo del valor x_q^t obtenido para todas las unidades de análisis en el momento t .

- Se puede aplicar tanto para datos cuantitativos como cualitativos
- Transforma los indicadores a una escala adimensional manteniendo las distancias relativas

- Normaliza los indicadores para obtener un rango de variación [0,1] para todos ellos
- Los valores atípicos pueden distorsionar el indicador transformado. Por otra parte, la normalización min-max puede ampliar el rango de los indicadores que están dentro de un mismo intervalo pequeño aumentando más el efecto sobre el IC que en la transformación z-score

Un método alternativo al anterior que tiene en cuenta varios instantes de tiempo durante el análisis y considera un año base (t_0) es el siguiente (Schuschny y Soto, 2009):

$$I_{qc}^t = \frac{x_{qc}^t - \min_c(x_q^{t_0})}{\max_c(x_q^{t_0}) - \min_c(x_q^{t_0})}$$

Podría suceder que $\exists t/x_{qc}^t > \max_c(x_q^{t_0})$ y, por lo tanto, el valor normalizado del indicador sería mayor que 1.

Como opción de mejora, podría calcularse:

$$I_{qc}^t = \frac{x_{qc}^t - \min_{t \in T} \min_c(x_q^t)}{\max_{t \in T} \max_c(x_q^t) - \min_{t \in T} \min_c(x_q^t)} \in [0,1]$$

siendo T el conjunto de todos los momentos temporales considerados en el análisis.

El problema de esta transformación es su inestabilidad debido a la llegada de nueva información, puesto que modifica los máximos y mínimos.

Distancia a una unidad de análisis referencial

$$I_{qc}^t = \frac{x_{qc}^t}{x_{qr}^{t_0}} \quad (3.1)$$

donde r es una unidad de análisis considerada como de referencia y t_0 el año base de estudio.

- Se usa únicamente con valores cuantitativos
- Transforma los indicadores a una escala adimensional manteniendo las distancias relativas
- No ajusta los diferentes rangos de variación de los indicadores
- Mide la posición relativa de un indicador dado un punto referencial. Este punto referencial podría ser un objetivo a alcanzar en un marco temporal dado o una unidad “promedio”. La unidad de análisis referencial también podría ser el líder del grupo en el que la principal unidad de análisis recibe una puntuación de 1 y a los otros se les asigna una puntuación

porcentual según la distancia a la unidad de análisis líder. Este procedimiento, sin embargo, está basado en valores extremos que podrían ser valores atípicos no fiables

- En algunos casos no se fija el valor del tiempo y la referencia respecto de la cual se calculan los valores puede ir variando según nueva información
- Un procedimiento diferente considera la unidad de referencia como la misma unidad de análisis y la compara con respecto al año inicial:

$$I_{qc}^t = \frac{x_{qc}^t}{x_{qc}^{t_0}} \tag{3.2}$$

Otro tipo de medida que puede emplearse es:

$$I_{qc}^t = \frac{x_{qc}^t - x_{qr}^{t_0}}{x_{qr}^{t_0}}$$

cuya única diferencia con respecto a la medida anterior es centrar los valores resultantes alrededor de cero. De la misma manera, la unidad de análisis referencial podría ser la unidad “promedio”, un grupo líder o una unidad de análisis benchmarking

ICs en los que se aplica la distancia a una unidad de análisis referencial para la normalización de los datos:

- *Environmental Policy Performance Indicator* (Adriaanse, 1993). En su desarrollo se ha aplicado la normalización según la ecuación (3.1)
- *Concern about environmental problems Index* (Parker, 1991). En su desarrollo se ha aplicado la normalización según la ecuación (3.2)

Categorización de escalas

$$\text{Ejemplo: } I_{qc}^t = \begin{cases} 0 & \text{si } x_{qc}^t < P^{15} \\ 20 & \text{si } P^{15} \leq x_{qc}^t < P^{25} \\ 40 & \text{si } P^{25} \leq x_{qc}^t < P^{65} \\ 60 & \text{si } P^{65} \leq x_{qc}^t < P^{85} \\ 80 & \text{si } P^{85} \leq x_{qc}^t < P^{95} \\ 100 & \text{si } P^{95} \leq x_{qc}^t \end{cases}$$

- Se usa tanto para indicadores cualitativos (ordinales) como para indicadores cuantitativos
- Transforma los indicadores a una escala adimensional
- Ajusta el rango de variación de [1,100] para cada uno de los indicadores
- Asigna una puntuación a cada indicador. La asignación puede ser categórica: una, dos o tres estrellas, o cualitativa: “objetivo alcanzado plenamente”, “objetivo alcanzado

parcialmente” o “objetivo no alcanzado”. Normalmente, a cada categoría se le asigna un rango de valores basados en los percentiles de la distribución del indicador a lo largo de las unidades de análisis, P^i (i-ésimo percentil de la distribución del indicador). Por ejemplo, el top 5% de las unidades de análisis recibe una puntuación de 100, las unidades entre los percentiles 85 y 95 reciben 80 puntos, los valores entre los percentiles 65 y 85 reciben 60 puntos y así sucesivamente hasta la asignación de 0 puntos. De esta forma, se premia a las unidades de análisis con mejores resultados y se penaliza a aquellas con peores resultados

Ventaja: si la transformación del percentil es la misma a lo largo de los años, una pequeña modificación en la definición del indicador no afectará a la variable transformada. Sin embargo, no se podrán analizar las diferencias de rendimiento año tras año

Limitación: se omite mucha información sobre las diferencias entre unidades de análisis en los indicadores transformados. Si hay una pequeña variación en los datos originales, los percentiles obligan a una categorización de los datos

Categorización de valores por encima o por debajo de la media

$$I_{qc}^t = \begin{cases} 1 & \text{si } w > (1 + p) \\ 0 & \text{si } (1 - p) \leq w \leq (1 + p) \\ -1 & \text{si } w < (1 - p) \end{cases} \text{ donde } w = \frac{x_{qc}^t}{x_{qc=\bar{c}}^{t_0}}$$

- Este método de normalización es simple y no se ve afectado por valores atípicos
- Sólo se aplica a datos cuantitativos
- Ajusta los datos para obtener una misma escala adimensional. Sin embargo, no mantiene las distancias relativas entre las unidades de análisis
- La transformación asigna a los valores alrededor de la media una puntuación de 0, mientras que a aquellos valores que están por encima o por debajo de un cierto umbral p se les asigna una puntuación de 1 y -1 respectivamente
- La arbitrariedad del umbral p y la omisión de un nivel absoluto de información es criticable. Por ejemplo, si el valor de un indicador dado para una unidad de análisis A está tres veces por encima de la media (300%) y el valor para la unidad de análisis B es 25% mayor que la media, se considera que ambos países están “por encima de la media” con un umbral del 20% alrededor de la media
- Ajusta el rango de variación para cada uno de los indicadores, ya que se les asigna un número discreto del conjunto $\{-1, 0, 1\}$

Método de normalización para indicadores cíclicos

$$I_{qc}^t = \frac{x_{qc}^t - E_t(x_{qc}^t)}{E_t(|x_{qc}^t - E_t(x_{qc}^t)|)}$$

donde $E_t(x_{qc}^t)$ indica la media de los valores de x_{qc}^t a lo largo de un cierto periodo de tiempo t .

- Este método implícitamente le asigna menos peso a las series más irregulares en el movimiento cíclico del IC

Porcentaje de diferencias anuales en años consecutivos

$$I_{qc}^t = \frac{x_{qc}^t - x_{qc}^{t-1}}{x_{qc}^t} * 100$$

- El indicador transformado es adimensional
- Representa el porcentaje de crecimiento con respecto al año anterior. La transformación sólo se usa cuando los indicadores están disponibles para diferentes años, como por ejemplo en el Índice de Mercado Interno (European Commission-DG MARKT, 2001; Tarantola et al., 2002; Tarantola et al., 2004)

La selección del método apropiado no es trivial. El método de normalización debe tener en cuenta las propiedades de los datos, así como el objetivo del IC. Los test de robustez son necesarios para evaluar su impacto en los resultados.

En la tabla 3.7 se resumen las características principales de los métodos de normalización descritos.

Metodología	Fórmula	Tipo de datos a los que se le puede aplicar	Ajuste de los indicadores a la misma escala de medida	Mantiene las distancias relativas	Ajuste de los indicadores al mismo rango de variación	Ajuste datos ante outliers
Ranking	$I_{qc}^t = Rank(x_{qc}^t)$	Datos cualitativos (ordinales) + datos cuantitativos	Sí	No	Sí. Rango: $[1, M]$ con M : # unidades de análisis)	Sí. Fija la media y la desviación típica
Estandarización	$I_{qc}^t = \frac{x_{qc}^t - x_{qc=\bar{c}}^t}{\sigma_{qc=\bar{c}}^t}$	Datos cuantitativos	Sí. Medida adimensional	Sí, puesto que la transformación es lineal	No. Para indicadores aprox, normales el 95% de los valores transformados estará es el intervalo (-2,2)	No
Re-escalamiento	$I_{qc}^t = \frac{x_{qc}^t - \min_c(x_q^{t_0})}{\max_c(x_q^{t_0}) - \min_c(x_q^{t_0})}$	Datos cualitativos + datos cuantitativos	Sí. Medida adimensional	Sí	Sí. Rango: $[0, 1]$ para cada indicador.	No
Distancia a una unidad de análisis referencial	$I_{qc}^t = \frac{x_{qc}^t}{x_{qr}^{t_0}}$ ó $I_{qc}^t = \frac{x_{qc}^t - x_{qr}^{t_0}}{x_{qr}^{t_0}}$	Datos cuantitativos	Sí. Medida adimensional	Sí	No	No
Categorización de escalas	Ejemplo: $I_{qc}^t = \begin{cases} 0 & \text{si } x_{qc}^t < P^{15} \\ 20 & \text{si } P^{15} \leq x_{qc}^t < P^{25} \\ 40 & \text{si } P^{25} \leq x_{qc}^t < P^{65} \\ 60 & \text{si } P^{65} \leq x_{qc}^t < P^{85} \\ 80 & \text{si } P^{85} \leq x_{qc}^t < P^{95} \\ 100 & \text{si } P^{95} \leq x_{qc}^t \end{cases}$	Datos cualitativos (ordinales) + datos cuantitativos	Sí	No	Sí. Rango: $[1, 100]$ para cada indicador	Sí. Fija la media y la desviación típica para un conjunto de percentiles elegidos.
Categorización de valores por encima o debajo de la media	$I_{qc}^t = \begin{cases} 1 & \text{si } w > (1 + p) \\ 0 & \text{si } (1 - p) \leq w \leq (1 + p) \\ -1 & \text{si } w < (1 - p) \end{cases}$ donde $w = \frac{x_{qc}^t}{x_{qc=\bar{c}}^{t_0}}$	Datos cuantitativos	Sí	No	Sí. Asignación de tres valores discretos $\{-1, 0, 1\}$	Sí
Métodos de indicadores cíclicos	$I_{qc}^t = \frac{x_{qc}^t - E_t(x_{qc}^t)}{E_t(x_{qc}^t - E_t(x_{qc}^t))}$	Datos cuantitativos	Sí. Media adimensional	Sí	No. Depende de la serie temporal del indicador	No
Porcentaje de diferencias anuales en años consecutivos	$I_{q,c}^t = \frac{x_{qc}^t - x_{qc}^{t-1}}{x_{qc}^t} * 100$	Datos cuantitativos	Sí. Media adimensional	No	No	Sí

Tabla 3.7. Características más importantes de los métodos de normalización

3.4.6. Ponderación y agregación de los indicadores simples

La ponderación constituye la etapa crucial de la composición de un IC. Consiste en asignar pesos a los indicadores simples para, posteriormente, agregarlos en un único valor que represente el índice global. La asignación de pesos se puede hacer de forma equitativa o bien estableciendo diferentes factores de peso que indiquen la importancia relativa que se le quiere dar a los indicadores simples en el índice global, la significatividad, fiabilidad u otras características de los datos subyacentes.

Los pesos también deben elegirse de tal forma que se refleje la calidad estadística de los datos. Muchas veces se suelen asignar pesos elevados a datos estadísticamente fiables con una amplia cobertura (Freudenberg, 2003; Nardo et al., 2005). Sin embargo, este método podría sesgar los resultados hacia aquellos indicadores que son fácilmente disponibles, penalizando la información que es estadísticamente más problemática de identificar y de medir. En el caso de las clasificaciones de países, por ejemplo, tendrían mejor posición los países más desarrollados con una exhaustiva recopilación de datos.

El método de ponderación elegido tiene un gran impacto sobre el valor final del IC y, por tanto, sobre la ordenación de los casos bajo estudio. Por ello, la etapa de ponderación debe ser lo más explícita, transparente y justificada posible.

En esta sección se presentan distintas técnicas de ponderación divididas en tres grandes bloques según los diferentes métodos utilizados: i) método de ponderación equitativa; ii) métodos de ponderación basados en modelos estadísticos y iii) métodos de ponderación basados en modelos participativos.

Mientras que algunos analistas optan por asignar únicamente pesos basados en métodos estadísticos, otros recompensan (o castigan) a los indicadores considerados más (o menos) importantes dependiendo de la opinión de los expertos, con el fin de reflejar mejor las prioridades políticas. No existe una metodología de ponderación objetiva, común y única para la construcción de un IC. Como consecuencia, los pesos deben seleccionarse de acuerdo al marco conceptual teórico subyacente que define el concepto que se está midiendo.

Métodos de Ponderación

ASIGNACIÓN DE PESOS IGUALES (Equal Weighting - EW)

Un IC es la agregación de un conjunto ponderado de indicadores simples. A cada indicador simple se le asigna un peso que se suele interpretar como una medida de la importancia relativa del indicador en la construcción del índice. De esta forma, los pesos expresan la relevancia de los indicadores en el contexto del fenómeno que se quiere medir (Saltelli et al., 2008).

En la mayoría de los estudios sobre ICs se plantea la opción de asignar el mismo peso a todos los indicadores simples durante su fase de construcción. Se trata del método de ponderación más sencillo que implica que tienen la misma importancia sobre el IC, pero puede ocultar la ausencia de una base estadística o empírica, por ejemplo cuando hay un conocimiento insuficiente de las relaciones causales o una falta de consenso sobre la alternativa. En cualquier caso, la equidad de pesos no significa que no se les asigna ningún peso a los indicadores sino que los pesos son iguales implícitamente. Además, si los indicadores están agrupados por dimensiones para la obtención del IC, la asignación equitativa de pesos en los indicadores simples puede implicar una asignación de pesos no igualitaria en las dimensiones (las dimensiones con mayor número de indicadores tendrán un peso mayor en la formación del IC), dando lugar a una estructura no equilibrada en el IC (Nardo et al., 2008).

Con el método de ponderación equitativa se corre el riesgo de que combinando indicadores con un elevado grado de correlación ciertos aspectos tengan un elevado peso en el IC. Por tanto, se debe comprobar la correlación estadística entre los indicadores y elegir aquellos que presenten un bajo grado de correlación entre ellos o dar menos peso a indicadores correlacionados.

Ejemplos de ICs en los que se asigna la misma ponderación a cada uno de los indicadores simples son el “Environmental Sustainability Index” (World Economic Forum, 2001) y el “Summary Innovation Index” (European Commission-DG ENTR, 2001b).

PONDERACIÓN BASADA EN MODELOS ESTADÍSTICOS

Análisis Factorial (Factorial Analysis - FA)

Las técnicas de reducción de la dimensión, descritas en la cuarta etapa de la construcción de un IC, también se pueden utilizar para la etapa de ponderación de los indicadores simples. Pueden ser útiles cuando los indicadores simples presentan una alta correlación. Sin embargo, se debe tener en cuenta que la agrupación de las variables se basa en propiedades estadísticas y no en el plano interpretativo del concepto que se desea medir. Además, los pesos que se obtienen (los coeficientes de las combinaciones lineales) pueden ser negativos (Schuschny y Soto, 2009).

Los pesos resultantes de esta técnica se basan en las correlaciones de los indicadores que no necesariamente corresponden a las relaciones conceptuales entre los indicadores y el concepto multidimensional a medir.

Una forma de obtener los pesos mediante las técnicas de reducción de la dimensión fue propuesta por Nicoletti, Scarpetta y Boylaud (2000). El enfoque utilizado por estos autores consiste en agrupar los indicadores simples con las mayores saturaciones en ICs intermedios. Los pasos a seguir son los siguientes:

1. Obtención de las variables latentes de acuerdo a las diferentes reglas de extracción
2. Cálculo de la varianza de todos los indicadores explicada por cada factor. Para obtener la varianza explicada se calcula la suma de los cuadrados de las cargas factoriales de los indicadores
3. Normalización de los cuadrados de las cargas factoriales. Para ello se calcula el cuadrado de la carga factorial dividido por la varianza de los indicadores explicada por el factor
4. Obtención de los pesos de cada factor mediante el cociente entre la varianza de los indicadores explicada por cada factor y la varianza total de los factores retenidos
5. Cálculo del peso de cada indicador ponderando la máxima variabilidad de éste explicada por un factor por el peso del factor. Los pesos obtenidos se normalizan dividiendo por la suma de todos ellos de tal forma que la suma sea la unidad

En la tabla 3.8 se presenta un ejemplo en el que se ha aplicado esta técnica de ponderación. Corresponde a la etapa de asignación de pesos del desarrollo del Índice de Rendimiento Tecnológico (TAI, por sus siglas en inglés) (United Nations Development Programme, 2001) formado por 8 indicadores simples:

Indicadores	Saturaciones				Cuadrado de las saturaciones normalizados				Pesos	Pesos normalizados
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4		
Indicador 1	0,07	0,97	0,06	0,06	0,00 ³	0,67	0,00	0,00	0,13 ⁴	0,17
Indicador 2	0,13	0,07	-0,07	0,93	0,01	0,00	0,00	0,49	0,12	0,15
Indicador 3	0,79	-0,21	0,21	0,42	0,24	0,03	0,04	0,10	0,09	0,11
Indicador 4	-0,64	0,56	-0,04	0,36	0,16	0,23	0,00	0,07	0,04	0,07
Indicador 5	0,37	0,17	0,38	0,68	0,05	0,02	0,12	0,26	0,06	0,08
Indicador 6	0,82	-0,04	0,25	0,35	0,25	0,00	0,05	0,07	0,09	0,12
Indicador 7	0,88	0,23	-0,09	0,09	0,29	0,04	0,01	0,00	0,11	0,14
Indicador 8	0,08	0,04	0,96	0,04	0,00	0,00	0,77	0,00	0,13	0,16
Varianza Explicada por el factor	2,64 ¹	1,39	1,19	1,76						
Varianza Total	6,98 ²	6,98	6,98	6,98						
V. E./V.T.	0,38	0,20	0,17	0,25						

Tabla 3.8. FA aplicado al IC TAI mediante el método de PCA con una rotación VARIMAX

$$^1 2,64 = 0,07^2 + 0,13^2 + 0,79^2 + (-0,64)^2 + 0,37^2 + 0,82^2 + 0,88^2 + 0,08^2$$

$$^2 6,98 = 2,64 + 1,39 + 1,19 + 1,76$$

$$^3 0,00 = \frac{0,07^2}{2,64}$$

$$^4 0,13 = 0,67 * 0,20$$

En este ejemplo existen cuatro ICs intermedios que corresponden a los factores obtenidos. El primero incluye el indicador 3 (peso: 0,24), el indicador 6 (peso: 0,25) y el 7 (peso: 0,29), el segundo incluye los indicadores 1 y 4 (pesos: 0,67 y 0,23 respectivamente), el tercer factor incluye el indicador 8 (peso: 0,77). Finalmente el cuarto factor incluye los indicadores 2 y 5 (pesos: 0,49 y 0,26 respectivamente).

Los cuatro ICs intermedios se agregan de tal forma que se le asigna a cada uno de ellos la proporción de varianza explicada en el conjunto de datos, 0,38 para el primero, 0,20 para el segundo, 0,17 para el tercero y 0,25 para el cuarto IC intermedio.

De acuerdo con esta técnica, la ponderación sólo interviene para corregir la superposición de información entre dos o más indicadores relacionados y no es una medida de la importancia teórica del indicador asociado. Además, los pesos obtenidos dependen del método de extracción usado y de los métodos de rotación.

La asignación de pesos mediante el análisis de componentes principales se basa en la correlación existente entre los indicadores simples que puede no corresponder con las verdaderas relaciones teóricas que pudieran existir entre los fenómenos bajo análisis y los indicadores simples con los que se trabaja. Se debe tener en cuenta que *correlación* no necesariamente implica *redundancia*. Si no existe correlación significativa entre los indicadores simples no se sugiere que los pesos se estimen con estas técnicas.

Otros ejemplos en los que se aplica esta técnica de ponderación a los indicadores simples son el “*Internal Market Index*” (European Commission-DG MARKT, 2001) y el “*Business Climate Indicator for the Euro Area*” (European Commission-DG ECFIN, 2000).

Análisis Envoltente de Datos (Data Envelopment Analysis - DEA)

El Análisis Envoltente de Datos es una herramienta no paramétrica de programación lineal surgida del análisis económico que ha sido utilizada para estudiar el rendimiento de unidades productivas, sectores, países, etc., que producen *outputs* similares a partir de una serie de *inputs* comunes de los cuales no se conocen los pesos de cada uno de éstos ni la forma funcional de la función de producción.

La metodología DEA es una extensión del trabajo desarrollado por Farrell (1957) que posteriormente pusieron en práctica Charnes et al. (1978) en la literatura de investigación operativa aplicada a diferentes campos de estudio. En un principio los modelos DEA fueron empleados para evaluar la eficiencia relativa de organizaciones sin ánimo de lucro, pero dada la naturaleza interdisciplinaria de la metodología su uso se extendió rápidamente al análisis de rendimiento de cualquier tipo de organización. Ejemplos de los distintos sectores en los que se ha aplicado esta metodología son: banca, salud, industria, educación secundaria, etc. Una exhaustiva recopilación bibliográfica sobre la metodología DEA se puede encontrar en el capítulo titulado “A DEA Bibliography” del libro “Data Envelopment Analysis: Theory, Methodology and Applications” (Seiford, 1994).

Esta metodología cuenta con las herramientas de programación lineal para construir una frontera eficiente y utilizarla como punto de referencia para medir el desempeño de un

determinado conjunto de unidades de análisis. Se tratan dos aspectos importantes en este análisis: la construcción de una frontera eficiente y la medida de distancia entre unidades de análisis en un marco teórico multi-dimensional (Nardo et al., 2005).

La metodología se puede aplicar para cualquier número de *inputs* y *outputs*, pero para expresar el concepto gráficamente supóngase que se dispone de un *input* x y de dos *outputs* y_1 , y_2 evaluados sobre $n = 17$ unidades de análisis (Rogge, 2011).

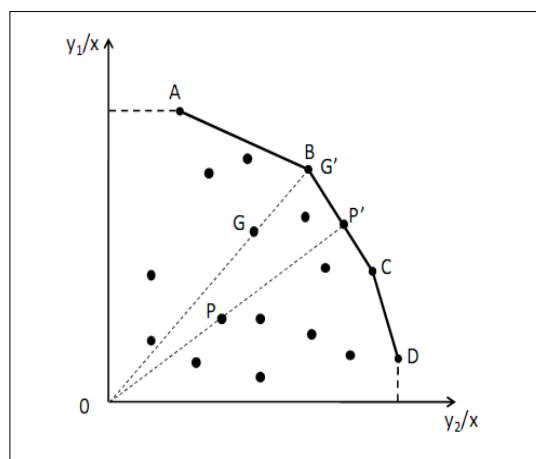


Figura 3.5. Modelo DEA con un *input* y dos *outputs*

En la figura 3.5 se refleja la naturaleza no paramétrica de la metodología DEA en la forma en la que la técnica escoge la frontera eficiente para evaluar la eficiencia de cada una de las unidades de análisis. En lugar de requerir una forma funcional explícita, el modelo DEA identifica la frontera eficiente empírica mediante la poligonal convexa que queda definida por aquellas unidades que poseen el mejor desempeño (aquellas que se ubican más al “nordeste” de los ejes).

De esta manera, las unidades ABCD de la figura son las que forman la frontera eficiente y las que tienen una eficiencia relativa máxima. Por el contrario, las unidades que se sitúan por debajo de la frontera son ineficientes. Para determinar el nivel de eficiencia de dichas observaciones se procede como sigue: i) proyectar la unidad ineficiente en la frontera eficiente y ii) tomar ratios de las distancias radiales.

Supóngase que se quiere evaluar la eficiencia de la unidad P . Es posible considerar una unidad “virtual” de referencia P' con la que P puede compararse. Por tanto, el análisis envolvente de datos no es más que una técnica, basada en el uso de la programación lineal matemática, en la que dada una unidad de análisis se permiten identificar aquellas otras que sirven para construir la unidad “virtual” de referencia que yace en la intersección con la línea que pasa por el origen y la ubicación del punto que definen los datos de la unidad de análisis dada. Debe considerarse que la unidad “virtual” de referencia puede ser que sea real o que sea un ideal con el cual comparar el desempeño de P , construido a partir de una combinación lineal

convexa de las unidades de referencia existentes. En el caso de la unidad P , la unidad virtual P' es una combinación lineal de las unidades B y C .

El desempeño de la unidad de análisis P , (h_p) , es el ratio de la distancia entre el origen y el valor de la unidad de análisis y la distancia entre el origen y el valor de la unidad de análisis referencial que será siempre menor o igual que la unidad:

$$h_p = \frac{OP}{OP'}$$

Cuanto más próxima esté la unidad de análisis evaluada a su unidad de referencia, *benchmark*, mayor será su eficiencia.

La formalización matemática de este método está detallada en el anexo A. Una descripción más exhaustiva de esta metodología se puede encontrar en el Handbook de Análisis Envolvente de Datos desarrollado por Cooper, Seiford y Tone (2004). Un ejemplo de la aplicación de esta técnica para la construcción de un IC puede verse en (Cherchye et al., 2008).

Un enfoque basado en esta técnica y aplicado al campo de los ICs es conocido como “Beneficio de la Duda” (“*Benefit of the Doubt - BOD*”). La descripción de esta técnica se expone en el capítulo 7.

Método de Regresión

Los modelos de regresión lineal pueden proporcionar información valiosa sobre el vínculo existente entre un conjunto de variables independientes $I_{1c}, I_{2c}, \dots, I_{Qc}$ y una variable dependiente \hat{Y}_c . Supóngase que las variables independientes son los indicadores simples que se han definido para la construcción del IC y la variable dependiente representa el objetivo global que debe alcanzar cada unidad de análisis. En muchos estudios sobre ICs la variable dependiente es otro IC muy relacionado con el que se quiere comparar (Saisana, 2008):

$$\hat{Y}_c = \hat{\alpha} + \hat{\beta}_1 I_{1c} + \dots + \hat{\beta}_Q I_{Qc} \text{ con } c = 1, \dots, M$$

donde $\hat{\alpha}$ es la constante estimada, M es el número de observaciones y $\hat{\beta}_1, \dots, \hat{\beta}_Q$ son los coeficientes de regresión asociados a los indicadores simples $I_{1c}, I_{2c}, \dots, I_{Qc}$ que pueden ser considerados, una vez estandarizados, como factores de ponderación.

Se trata de un procedimiento que, aunque es adecuado para un número elevado de variables de diferentes tipos, implica la suposición de que los indicadores simples tienen un comportamiento lineal en relación con el objetivo planteado y que deben ser independientes

entre sí, puesto que si existe multicolinealidad el análisis se torna deficiente. No obstante, el uso de modelos de regresión lineal puede ser útil para cuantificar el efecto relativo de cada objetivo de política, representado por cada variable, y los objetivos globales a ser alcanzados, así como para validar un conjunto de factores de ponderación calculados a partir de otra técnica.

Modelos de componentes no observados (Unobserved Components Models - UCM)

En los modelos de componentes no observados los indicadores simples que componen el IC se suponen dependientes de una variable no observada más un término de error. A través de la estimación de la variable no observada es posible obtener información sobre las relaciones entre el IC y los indicadores individuales. Los pesos obtenidos serán aquellos que minimicen el término de error resultante. El único inconveniente en su aplicación es la complejidad en la estimación y su elevado coste computacional.

Según el enfoque del modelo de componentes no observados descrito por (Nardo et al., 2005) se define como h_c el fenómeno desconocido que se desea medir para cada unidad de análisis, $c = 1, \dots, M$ y q los indicadores simples disponibles con $q = 1, \dots, Q$, cada uno de los cuales mide un aspecto de h_c . El resultado observable de la unidad de análisis c para el indicador q , I_{qc} , puede escribirse como una combinación lineal de la variable no observada más un término de error:

$$I_{qc} = \alpha_q + \beta_q [h_c + \epsilon_{qc}]$$

donde α_q y β_q son parámetros desconocidos que proyectan h_c en I_{qc} .

El término de error capta dos fuentes de incertidumbre: i) la variable no observada puede ser medida incorrectamente en cada unidad de análisis; ii) la relación de proyección puede que no sea perfecta (por ejemplo, I_{qc} puede ser solamente un indicador de ruido del fenómeno desconocido si existen diferencias entre las unidades de análisis para el indicador).

Se asume que el término de error posee media nula y la misma varianza para todas las unidades de análisis con respecto a un indicador simple pero distinta varianza entre indicadores:

$$E(\epsilon_{qc}) = 0; E(\epsilon_{qc}^2) = \sigma_q^2; E(\epsilon_{qc}\epsilon_{ih}) = 0 \text{ para } c \neq i, q \neq h$$

Además, se asume que el término de error es independiente entre indicadores, puesto que cada indicador simple debe medir idealmente un aspecto concreto del fenómeno desconocido independiente de los otros.

Supóngase que h_c es una variable aleatoria con media cero y varianza la unidad y los indicadores simples se normalizan aplicando la técnica min-max. Todos estos supuestos simplifican la estimación de la variable no observada para la unidad de análisis c . Esta estimación se calcula con la media de la distribución condicional de la variable no observada una vez normalizados los valores de los indicadores simples:

$$E[h_c | I_{1c}, \dots, I_{Qc}] = \sum_{q=1}^Q w_{qc} \frac{I_{qc} - \alpha_q}{\beta_q} \quad (3.3)$$

Los pesos asignados son:

$$w_{qc} = \frac{\sigma_q^{-2}}{1 + \sum_{q=1}^Q \sigma_q^{-2}} \quad (3.4)$$

siendo w_{qc} una función decreciente en la varianza del indicador q y creciente en la varianza de los demás indicadores. Por tanto, el peso w_{qc} depende de la varianza del indicador q y de la suma de las varianzas de todos los indicadores (incluyendo q).

Sin embargo, si existen unidades de análisis que carecen de datos en indicadores individuales, el denominador del factor peso puede ser específico de las unidades de análisis que sí tengan datos en todos los indicadores. Esto hace que no se puedan comparar los valores de las unidades de análisis en el IC, como en BOD. Pero si esto no ocurre, los pesos no dependen de unidades de análisis específicas y, por tanto, es posible la comparación.

La varianza de la distribución condicionada viene dada por la expresión:

$$Var[h_c | I_{1c}, \dots, I_{Qc}] = \frac{1}{1 + \sum_{q=1}^Q \sigma_q^{-2}}$$

y puede usarse como una medida de precisión del IC. La varianza decrece con el número de indicadores para cada unidad de análisis y crece en el término de error para cada indicador. Los parámetros desconocidos a ser estimados son α_q, β_q y σ_q^2 para cada indicador que maximizan la función de verosimilitud que se basa en los datos estimados y se sustituyen en (3.3) y (3.4) para obtener el IC y los factores de peso.

PONDERACIÓN BASADA EN MÉTODOS PARTICIPATIVOS

De forma alternativa a las técnicas de ponderación basadas en modelos estadísticos existen otras técnicas de ponderación basadas en métodos participativos en las que se pide opinión a expertos en el tema en cuestión sobre la importancia que debe tener cada indicador simple en el

índice global. La utilización de estas técnicas es adecuada cuando se tiene una base de una cierta política bien definida. Esa base no suele estar disponible para comparaciones internacionales.

Método de Asignación Presupuestaria (*Budget Allocation - BA*)

En esta técnica de ponderación a cada experto se le asigna un determinado “presupuesto” que también se puede definir como una determinada cantidad y se le pide que haga una repartición de esa cantidad entre los diferentes indicadores simples. En este contexto vuelve a aparecer la idea de peso como una medida de la importancia del indicador simple.

Normalmente la cantidad asignada es de 100 puntos a repartir entre los indicadores simples con el fin de tener ya la suma normalizada. Se suelen hacer varias rondas entre los expertos hasta conseguir un consenso en el conjunto de pesos acordado por todos ellos. Si dicho consenso no fuese posible, no sería adecuada la aplicación de esta técnica de ponderación, puesto que el IC no tendría coherencia interna.

Esta técnica de ponderación es óptima para un número máximo de 12 indicadores simples. Si se dispone de más indicadores, este método puede producir confusión y estrés a los expertos (Nardo et al., 2005).

La principal ventaja de este método de ponderación es su transparencia y su simple aplicación. Es importante que el conjunto de expertos seleccionados refleje la diversidad de ramas de conocimiento y las diferentes áreas de experiencia del tema a analizar.

Un ejemplo de IC en el que se ha aplicado la técnica de asignación presupuestaria es “*Composite Indicator on e-Business Readiness*” (European Commission-DG ENTR, 2004).

Opinión pública

Muchas veces los expertos ponderan los indicadores fundamentalmente en base a las urgencias políticas y no tanto a la importancia relativa que tiene cada uno de ellos. Es por eso que otra posible técnica de ponderación se basa en la opinión pública. En este caso se realiza una encuesta para que el público participe en la asignación de los pesos de cada uno de los indicadores simples. No obstante, es más difícil que la opinión pública redistribuya los 100 puntos entre los indicadores que den su opinión sobre el grado de preocupación del tema en cuestión.

Ejemplos en los que se ha aplicado esta técnica son “*Environmental Policy Performance Indicator*” (Adriaanse, 1993) y “*Human Development Index*” (United Nations Development Programme, 2001).

Procesos de Jerarquía Analítica (Analytic Hierarchy Processes - AHP)

Los procesos de jerarquía analítica fueron desarrollados por Thomas Saaty (1980) y se utilizan para la toma de decisiones multi-criterio. Es una técnica participativa de ponderación diseñada para obtener una escala numérica con el fin de dar prioridades a las alternativas de decisión según los juicios subjetivos de un grupo de expertos. Trata de dar prioridades cuando se deben considerar múltiples aspectos, tanto cuantitativos como cualitativos, en una decisión en la que las opiniones se obtienen mediante comparación por pares de los indicadores simples preguntando a los expertos cuál de los dos indicadores es más importante y en qué cantidad. La preferencia se expresa en una escala de Likert de 1 a 9. Una preferencia de 1 significa que para ese experto los dos indicadores son igualmente importantes y una preferencia de 9 indica que uno de los indicadores es 9 veces más importante que el otro.

Esta técnica participativa de ponderación está formada por cuatro etapas (Zahedi, 1986):

Etapa 1: Crear una jerarquía de decisión desagregando el problema a evaluar en una jerarquía de decisiones interrelacionadas entre ellas

Etapa 2: Coleccionar los datos usando comparaciones por pares de los elementos de decisión

Etapa 3: Utilizar el método de los autovalores para estimar los pesos relativos de los elementos de decisión

Etapa 4: Agregar los pesos relativos de los elementos de decisión en un conjunto de clasificaciones sobre las alternativas de decisión

El carácter compensatorio del método implica que los pesos obtenidos indiquen en qué cantidad un grupo de expertos, en promedio, está dispuesto a renunciar a un indicador por otro. Los pesos obtenidos con el método AHP no son medidas de importancia, es decir, no miden el grado de importancia que tiene cada uno de los indicadores sobre el IC. Este aspecto ha generado muchas tergiversaciones en la literatura de los ICs en los que se utilizan pesos AHP como medida de importancia (Nardo et al., 2008).

Los resultados se expresan en una matriz, A , de comparaciones entre indicadores. Por tanto, A es una matriz cuadrada en el número de indicadores, cuyos elementos a_{ij} indican en qué cantidad el indicador de la fila i es más importante que el indicador de la fila j . Para completar la matriz basta considerar que $a_{ii} = 1$ y $a_{ij} = \frac{1}{a_{ji}}$.

Comparando la matriz A con una matriz de razones de pesos W con $w_{ij} = \frac{w_i}{w_j} \in R^{Q \times Q}$ con $w_i > 0$, y siendo $i = 1, \dots, Q$ el número de indicadores disponibles, se cumple que:

$$w_{ji} = \frac{1}{w_{ij}}; w_{ii} = 1; w_{ij} = w_{ik}w_{kj}$$

El rango de la matriz W es la unidad, puesto que cada fila es múltiplo de la primera, por tanto, sólo existirá un autovalor de la matriz no nulo. Sea λ dicho autovalor. Como $w_{ii} = 1 \forall i$ y la suma de todos los autovalores es igual a la traza se tiene que:

$$\sum_{i=1}^q w_{ii} = \lambda$$

donde se cumple que $Wv = \lambda v$ con v el vector columna de pesos.

Por tanto, como A tiene una forma similar en su construcción a W se puede demostrar que:

- λ_{max} es igual o mayor a λ
- el autovector asociado a λ_{max} posee todas las componentes estrictamente positivas

De esta forma, para estimar los pesos a partir de la matriz de comparaciones A bastará con calcular el autovector asociado a su autovalor λ_{max} y normalizarlo de tal manera que la suma sea la unidad. Las componentes de dicho vector son los pesos que se calculan a partir de la matriz de comparaciones realizada por expertos. Los pesos obtenidos por cada experto se pueden promediar para obtener un conjunto de pesos final.

La ventaja de esta técnica participativa de ponderación es que el proceso de definición de los pesos se obtiene de una forma transparente sobre la base de un procedimiento muy utilizado. El principal inconveniente es que las comparaciones por pares de indicadores pueden resultar costosas de obtener. Normalmente se usa esta técnica para un máximo de diez indicadores. También puede ocurrir que las opiniones de los expertos sean inconsistentes. En este caso, se debería medir el grado de inconsistencia para proporcionar unos resultados aceptables.

Análisis Conjunto (Conjoint Analysis - CA)

El análisis conjunto es una técnica estadística utilizada en ciencias sociales aplicadas, particularmente en el marketing, la administración del producto y la investigación operativa, que trata de entender cómo los encuestados desarrollan preferencias acerca de productos o servicios (Hair et al., 2007). Se trata de una metodología de carácter participativo que se basa en la evaluación por parte de los expertos del valor de un servicio, producto o concepto combinando cantidades separadas de valor que proporciona cada atributo.

El análisis conjunto es el único entre los métodos multivariantes en el que el analista debe reformular el problema identificando los atributos a considerar. En el caso de los ICs, los atributos son los indicadores y sus posibles niveles. Un ejemplo sencillo que clarifica el concepto de niveles podría ser un indicador con dos tipos de marcas (niveles), marca X y marca Y. También se podrían elegir diferentes niveles seleccionando un conjunto de valores del

indicador simple según el tipo de indicador que sea, económico, social, institucional o ambiental.

Una vez definidos los diferentes niveles de los indicadores, el siguiente paso es diseñar la encuesta para poder ser evaluada por los expertos. Los tipos de encuesta que se suelen diseñar en la construcción del IC se describen a continuación:

- Método *trade-off* en el que se comparan dos indicadores al mismo tiempo mediante la clasificación de todas las combinaciones de niveles. Este procedimiento puede resultar muy tedioso para los expertos, por eso se recomienda la utilización de esta técnica con menos de diez indicadores simples
- Evaluación directa de cada indicador realizando una evaluación global independiente del resto de indicadores

El análisis conjunto se denomina modelo de descomposición, puesto que el analista necesita conocer solamente una preferencia global del experto encuestado para un objeto creado por el analista a través de la especificación de valores o niveles para cada indicador (Hair et al., 2007).

En el análisis conjunto subyace una relación funcional del siguiente tipo (Nardo et al., 2008):

$$preferencia_c = F(I_{1c}, I_{2c}, \dots, I_{qc})$$

La variable dependiente recoge la utilidad o preferencia que el experto atribuye al concepto que se desea medir y las variables independientes son los indicadores que forman el IC, cada uno de los cuales puede presentar varios niveles. Sólo la variable dependiente recogerá la opinión de los expertos, puesto que las variables independientes son definidas por el analista dependiendo de lo que se quiera someter a evaluación. El análisis conjunto trata de averiguar los niveles de las variables independientes, es decir, los indicadores que optimizan la utilidad de la variable dependiente de tal forma que su variabilidad sea pequeña.

La puntuación de utilidad asociada a un nivel es un valor que indica el grado de preferencia o rechazo del experto por el mismo. A partir de las utilidades de los niveles del indicador se puede calcular la importancia que tiene el indicador para el experto.

Las derivadas con respecto a los indicadores simples de la función de preferencia se pueden usar para determinar los pesos de cada uno de los indicadores. De esta forma:

$$IC_c = \sum_{q=1}^Q \frac{\partial F}{\partial I_{qc}} I_{qc}$$

$\partial F / \partial I_{qc}$ es una medida de trade-off, es decir, indica cómo varían las preferencias al hacer variar el indicador. Esto implica la compensabilidad entre los indicadores.

Uno de los inconvenientes de este método es la necesidad de especificar y estimar la función de utilidad (Nardo et al., 2005).

Técnicas de Agregación

La agregación corresponde a la etapa más polémica en la construcción de un IC. Muchos autores están en contra de agregar toda la información disponible obtenida mediante las etapas anteriores en un único valor que represente a cada unidad de análisis, ya que puede perderse mucha información. Por otra parte, es muy difícil intentar interpretar e identificar diferencias entre las unidades de análisis si no se realiza una agregación de los indicadores simples.

En este estudio se van a considerar diferentes técnicas de agregación de los indicadores simples para la obtención del IC, aunque en la práctica la que más se suele utilizar es la agregación lineal ponderada (véase (Nardo et al., 2008)).

PROCEDIMIENTOS COMPENSATORIOS

Métodos aditivos de agregación lineal

Suma de Rankings

Es el método más simple de agregación consistente en sumar, para cada unidad de análisis c con $c = 1, \dots, M$, el orden o ranking que posee cada uno de los indicadores simples en relación con el resto de unidades de análisis:

$$IC_c = \sum_{q=1}^Q \text{Ranking}_{qc} \text{ con } c = 1, \dots, M$$

Se trata de un método que no se ve afectado por valores atípicos. La principal desventaja al utilizar este método es que se pierde mucha información al calcular el ranking en los indicadores simples.

Conteo de los indicadores que superan o exceden una referencia dada

Este método está basado en el número de indicadores que están por encima o por debajo de una unidad de referencia p :

$$IC_c = \sum_{q=1}^Q \text{sgn}\left[\frac{I_{qc}}{E(I_q)} - (1 + p)\right] \text{ para } c = 1, \dots, M$$

siendo I_{qc} el valor normalizado de la unidad de análisis c respecto al indicador q , para $q = 1, \dots, Q$ y $c = 1, \dots, M$.

El umbral p puede elegirse arbitrariamente por encima o por debajo de la media. Este valor se selecciona después de realizar un análisis exploratorio que permita determinar aproximadamente el rango de valores que toman los indicadores. Al igual que el método anterior, tiene la ventaja de no estar influenciado por valores atípicos y el inconveniente de perder información al agregar a nivel absoluto.

Agregación lineal ponderada

Se trata del método de agregación lineal más utilizado en la construcción de un IC (Bandura y Martin, 2006; Nardo et al., 2008):

$$IC_c = \sum_{q=1}^Q w_q I_{qc}$$

con $\sum_{q=1}^Q w_q = 1$ y $0 \leq w_q \leq 1$, y I_{qc} el valor normalizado de la unidad de análisis c respecto al indicador q , para $q = 1, \dots, Q$ y $c = 1, \dots, M$. La obtención de los pesos w_q debe quedar clara en la etapa anterior.

Agregación Geométrica

Un comportamiento no deseado de las técnicas de agregación lineal descritas anteriormente es la compensación total entre indicadores, de tal forma que si existe un rendimiento bajo en algunos indicadores este comportamiento se compensa por altos valores en el resto de indicadores. Los pesos empleados en las técnicas de agregación compensatoria se tratan como factores de escala o *trade-offs* (Munda, 2008).

La agregación geométrica es una solución intermedia entre la compensación total y la no compensación entre los indicadores proporcionada por las técnicas de agregación multi-criterio no compensatorias.

Este tipo de agregación es similar a la agregación lineal ponderada pero considerando la media geométrica:

$$IC_c = \prod_{q=1}^Q (I_{qc})^{w_q}$$

con $\sum_{q=1}^Q w_q = 1$ y $0 \leq w_q \leq 1$, y I_{qc} el valor normalizado de la unidad de análisis c respecto al indicador q , para $q = 1, \dots, Q$ y $c = 1, \dots, M$. La obtención de los pesos w_q debe quedar clara en la etapa anterior.

A modo de ejemplo, se presenta en la tabla 3.9 las diferencias entre la agregación lineal y geométrica para dos unidades de análisis A y B. Supóngase que se dispone de cuatro indicadores simples para cada unidad:

Unidad de análisis	$IC_c = \sum_{q=1}^4 w_q I_{qc}$	$IC_c = \prod_{q=1}^4 I_{qc}^{w_q}$
A: $I_{qc}=26,2,2,2$ (pesos asignados $w_q=1/4$)	8	3,79
B: $I_{qc}=8,8,8,8$ (pesos asignados $w_q=1/4$)	8	8

Tabla 3.9. Diferencias entre la agregación lineal y geométrica

La agregación geométrica diferencia las dos unidades, mientras que la agregación lineal obtiene el mismo valor para las dos unidades de análisis a pesar de tener puntuaciones muy distintas.

Así mismo, un cambio marginal de los tres últimos indicadores tendrá un mayor efecto sobre el IC si se agrega de forma geométrica. Por tanto, si se emplea esta técnica de agregación habrá un mayor incentivo a aplicar políticas que mejoren la situación que reflejan los indicadores con valores más bajos.

Aun así, en ambas técnicas de agregación subyace un razonamiento compensatorio entre los indicadores, lo que supone la existencia de vínculos entre éstos. Si se desea tratar a los factores de ponderación como medidas de la importancia de cada indicador y no se desea que reflejen la posibilidad de compensación entre éstos se recurre a las técnicas multi-criterio de agregación no compensatoria.

PROCEDIMIENTOS NO COMPENSATORIOS

Análisis multi-criterio no compensatorio

La literatura que apoya a las técnicas de agregación multi-criterio no compensatorio critica abiertamente el uso de los procedimientos de agregación compensatorios (Munda y Nardo, 2005; Munda y Nardo, 2009; Billaut, Bouyssou y Vincke, 2010). El procedimiento multi-criterio no compensatorio trata de resolver los conflictos que surgen a la hora de establecer un ordenamiento entre las unidades de análisis disponibles. Estos conflictos tienen lugar cuando algunos de los indicadores que se utilizan para evaluar el conjunto de observaciones favorecen a ciertas unidades de análisis, mientras que otros indicadores favorecen a otras. Para evitar este problema se usan los procedimientos multi-criterio discretos y no compensatorios.

Estos procedimientos tienen la ventaja de superar algunos inconvenientes que presentan las agregaciones lineales y geométricas, como la posible compensación entre los indicadores y la necesidad de establecer una misma escala de medida para éstos. Además, los pesos empleados con criterios ordinales son la base de los procedimientos de agregación no compensatorio y reflejan la importancia de cada indicador sobre el IC (Munda, 2008). No obstante, los principales inconvenientes son las costosas comparaciones que se deben hacer entre las unidades de análisis y la posible aparición de situaciones cíclicas en las que no se llegue a ninguna solución.

El método de Condorcet y el método de Copeland son dos de las principales técnicas de agregación multi-criterio no compensatorio. En el capítulo 6 se describe con más detalle el método de Copeland, puesto que es el que se utiliza para ilustrar los resultados del análisis de incertidumbre de los escenarios de los ICs propuestos.

COMPATIBILIDAD ENTRE LOS MÉTODOS DE PONDERACIÓN Y LOS MÉTODOS DE AGREGACIÓN

En la tabla 3.10 se presentan las compatibilidades entre los métodos de asignación de pesos y los métodos de agregación que se han presentado anteriormente (Nardo et al., 2005).

Métodos de ponderación	Métodos de agregación		
	Lineal ⁴	Geométrico ⁴	Multi-criterio no compensatorios
EW	Sí	Sí	Sí
FA	Sí	Sí	Sí
BOD	Sí ¹	No ²	No ²
UCM	Sí	No ²	No ²
BA	Sí	Sí	Sí
AHP	Sí	Sí	No ³
CA	Sí	Sí	No ³

Tabla 3.10. Compatibilidad entre métodos de ponderación y métodos de agregación

¹Requiere una normalización con el método Min-Max

²BOD requiere agregación aditiva, lo mismo ocurre con UCM

³Los métodos multi-criterio no compensatorios tratan a los pesos como coeficientes de importancia

⁴En ambas agregaciones, los pesos nominales son factores de escala o de compensación

3.4.7. Análisis de incertidumbre y sensibilidad del Indicador Compuesto

La construcción de un IC requiere de etapas subjetivas en las que se emiten juicios de valor para la elección de los diferentes supuestos. Ejemplos de estas etapas son: la elección de los indicadores simples, la imputación de datos faltantes, la asignación de pesos para cada indicador y dimensión, la elección del método de agregación, etc. Estas elecciones subjetivas forman la base de la construcción del IC, por tanto, para incrementar la transparencia y evaluar la calidad de su diseño es imprescindible aplicar un análisis de sensibilidad y de incertidumbre.

En general, las fuentes de incertidumbre asociadas al diseño de ICs dependen de muchos factores, entre los que destacan (Saltelli et al., 2008):

- La inclusión o exclusión de indicadores simples en la construcción del índice global
- Tratamiento previo de los datos (detección de valores atípicos, detección de distribuciones asimétricas en los indicadores...)
- Elección del método apropiado de imputación de datos faltantes
- Elección del tipo de normalización de los datos
- Elección de los pesos asignados a cada indicador simple y, si se da el caso, a las dimensiones
- Elección del método de agregación de los indicadores simples y, si se da el caso, de las dimensiones

Todas las premisas o supuestos descritos anteriormente pueden influenciar en la puntuación final de cada unidad de análisis del estudio, por lo que se deben tener en cuenta en la construcción de un IC, ya que diferentes metodologías pueden originar diferentes ICs. No

obstante, se debe elegir aquel índice que sea válido y que cumpla que pequeños cambios en su arquitectura den lugar a pequeñas variaciones en la puntuación final.

La construcción de un IC puede identificarse con el desarrollo de un modelo. Supóngase que se asume un modelo representado por una función de la siguiente forma:

$$Y = f(X_1, X_2, \dots, X_n) = f(\mathbf{X})$$

donde Y es la variable *output* del modelo y X_1, X_2, \dots, X_n son los factores *input* del modelo.

El análisis de sensibilidad trata de estudiar cómo la variación en la salida del modelo puede distribuirse cuantitativamente a diferentes fuentes de variación en los supuestos del modelo (Nardo et al., 2008).

El análisis de sensibilidad está ligado al análisis de incertidumbre. Este último trata de cuantificar la incertidumbre del output provocada por diferentes fuentes de incertidumbre de los factores input, mientras que el análisis de sensibilidad trata de identificar de dónde proviene dicha incertidumbre, es decir, cuáles son los factores input que más afectan a la variación del output. Una combinación del análisis de incertidumbre y sensibilidad ayuda a evaluar la robustez y calidad del IC y a aumentar la transparencia de su construcción.

No obstante, se debe tener cuidado y ser objetivo en la decisión sobre el grado de incertidumbre que se debe asignar a los factores input para comprobar la variación del factor output. Si la variación asignada a los factores input es grande, el modelo predictivo tendrá tanta variabilidad que el análisis no será útil. Edward E. Leamer (1983) lo resume como sigue:

“I propose a form of organized sensitivity analysis in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful.”

En la construcción de un IC los factores input suelen identificarse con las posibles metodologías que se pueden aplicar en cada etapa de su construcción, mientras que el factor output se identifica con el objetivo del estudio, es decir, con el IC construido para cada una de las unidades de análisis.

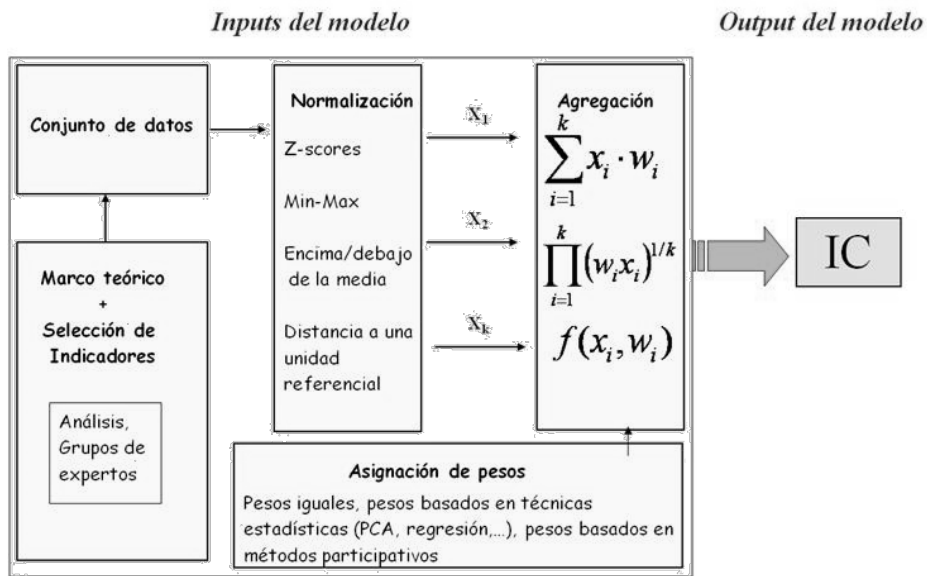


Figura 3.6. Modelo de construcción de un IC

Por tanto, el objetivo de esta etapa es realizar un análisis de incertidumbre con el fin de cuantificar la incertidumbre del IC provocada por las diferentes fuentes de variación del espacio de supuestos e identificar mediante un análisis de sensibilidad los supuestos que más afectan a la incertidumbre del índice global.

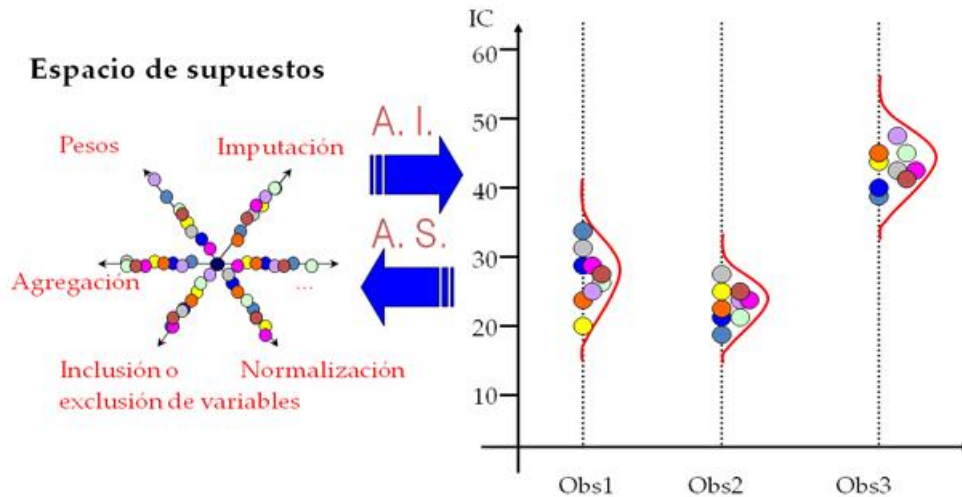


Figura 3.7. Análisis de incertidumbre y sensibilidad de un IC

A continuación se describen los procedimientos de análisis de incertidumbre y sensibilidad que se suelen aplicar a los ICs para el estudio de su robustez y calidad (Saisana, Saltelli y Tarantola, 2005; Saltelli et al., 2008). El ejemplo que se expone para el desarrollo de ambos análisis es ilustrativo, puesto que en la práctica la elección del espacio de supuestos y de las fuentes de incertidumbre depende del objetivo del problema planteado.

Supóngase que en el modelo planteado se consideran tres fuentes de incertidumbre/supuestos:

1. Técnica de normalización
2. Técnica de ponderación
3. Técnica de agregación

Sea IC_j el valor del índice para las unidades de análisis j con $j = 1, \dots, M$,

$$IC_j = f_{rs}(I_{1,j}, I_{2,j}, \dots, I_{Q,j}, w_{s,1}, w_{s,2}, \dots, w_{s,Q})$$

calculado con el método de agregación $r = \{1, 2, 3, \dots\}$ y el método de ponderación $s = \{1, 2, 3, \dots\}$, con un número finito de posibilidades para r y s , y donde los números se identifican con distintas técnicas de agregación y ponderación. Los $I_{Q,j}$ representan los indicadores simples utilizados para la construcción del índice una vez normalizados mediante una técnica de normalización y los $w_{s,Q}$ los pesos de los indicadores simples calculados según el modelo de ponderación s . Puede ocurrir que algunas de las combinaciones de los distintos métodos no se puedan aplicar (véase tabla 3.10).

Un procedimiento utilizado para cuantificar el output del modelo es calcular la posición que ocupa cada unidad de análisis en relación al resto de unidades, $Ranking(IC_j)$, y estimar la discrepancia respecto a una metodología de referencia, $Ranking_{referencial}(IC_j)$:

$$\bar{R}_s = \frac{1}{M} \sum_{j=1}^M |Ranking_{referencial}(IC_j) - Ranking(IC_j)|$$

En este caso, diferentes supuestos en la construcción del IC pueden producir una variación en las variables $Ranking(IC_j)$ y \bar{R}_s . Estas dos variables se identifican como variables output del modelo y constituyen las medidas de interés en el análisis de sensibilidad e incertidumbre. Se pueden utilizar otras medidas de interés según convenga.

Análisis de Sensibilidad Global basado en el cálculo de varianzas

El análisis de sensibilidad de un modelo es un procedimiento relevante, ya que permite identificar los factores input que más afectan a la incertidumbre del output. Existen diversos métodos de análisis de sensibilidad clasificados en dos grandes grupos: análisis de sensibilidad local y análisis de sensibilidad global.

El análisis de sensibilidad local está formado por métodos locales basados en el cálculo de derivadas parciales de los factores output con respecto a los factores input. Se trata de métodos simples de evaluar y con bajo coste computacional. La principal limitación de este tipo de análisis es que proporcionan información únicamente en el punto base donde las derivadas son calculadas, sin explorar todo el intervalo de variación de los factores input. Además, cuando el modelo contiene discontinuidades las derivadas parciales no pueden ser calculadas. En el análisis de sensibilidad global (GSA, por sus siglas en inglés), que supera las limitaciones del análisis de sensibilidad local, se han desarrollado métodos como los coeficientes de regresión estandarizados, la prueba de efectos elementales, el filtrado de Monte Carlo, los métodos basados en el cálculo de varianzas, etc (véase Saltelli, Chan y Scott (2000), Saltelli et al. (2004) y Saltelli et al. (2008)).

Los métodos de análisis de sensibilidad global más usados en la literatura se basan en el cálculo de varianzas, puesto que presentan las siguientes ventajas (Nardo et al., 2008; Saltelli et al., 2008):

- Se pueden aplicar independientemente de la forma funcional del modelo
- Permiten distinguir, mediante el uso de coeficientes de sensibilidad, los principales factores que afectan a la sensibilidad del IC
- Permiten distinguir los efectos principales de primer orden de los efectos de orden superior de los factores input
- Permiten captar la influencia de todo el rango de variación de cada factor input
- Permiten tratar a los factores input de forma agrupada
- Son fáciles de interpretar

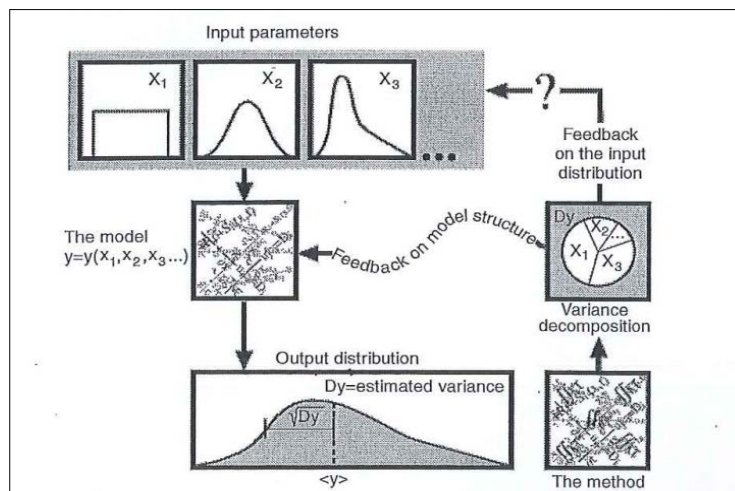


Figura 3.8. Esquema del procedimiento GSA basado en el cálculo de varianzas

A continuación se describe el procedimiento para la obtención de medidas de sensibilidad de los factores input del modelo a partir del cálculo de varianzas (Saltelli et al., 2008).

Supóngase el siguiente modelo genérico:

$$Y = f(X_1, X_2, \dots, X_n)$$

Sean X_i , $i = 1, \dots, n$, los factores input del modelo con un rango de variación o de incertidumbre no nulo e Y el factor output que puede ser *Ranking*(IC_j), \bar{R}_s o cualquier otra medida de interés.

Supóngase que se fija un factor X_i en un punto determinado x_i^* . Sea $V_{\mathbf{X}_{\sim i}}(Y|X_i = x_i^*)$ la varianza resultante de Y sobre $\mathbf{X}_{\sim i}$ (la notación “ $\sim i$ ” significa que en \mathbf{X} se están considerando todos los factores menos el factor o indicador X_i). Este término se llama varianza condicionada, puesto que se calcula al fijar X_i en el valor x_i^* . Cabe esperar que al fijar una fuente de variación X_i , la varianza resultante $V_{\mathbf{X}_{\sim i}}(Y|X_i = x_i^*)$ sea inferior que la varianza total incondicionada $V(Y)$. Por tanto, se puede utilizar $V_{\mathbf{X}_{\sim i}}(Y|X_i = x_i^*)$ como una medida de la importancia relativa de X_i sobre Y , de tal forma que cuanto menor sea $V_{\mathbf{X}_{\sim i}}(Y|X_i = x_i^*)$, mayor será la influencia de X_i sobre Y .

Sin embargo, se han detectado dos limitaciones con este procedimiento. En primer lugar, esta medida de importancia depende de la posición del punto x_i^* para cada factor input. En segundo lugar, se puede definir un modelo tal que para ciertos X_i y puntos fijos x_i^* ocurre que $V_{\mathbf{X}_{\sim i}}(Y|X_i = x_i^*) > V(Y)$. Para resolver estas limitaciones se calcula el promedio de la medida definida sobre todos los posibles valores de x_i^* , de tal forma que la dependencia con x_i^* desaparece. El promedio $E_{X_i}(V_{\mathbf{X}_{\sim i}}(Y|X_i))$ siempre es inferior a $V(Y)$ puesto que:

$$V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i)) + E_{X_i}(V_{\mathbf{X}_{\sim i}}(Y|X_i)) = V(Y)$$

donde al primer término se le llama **efecto principal de la variable X_i sobre el output Y** , y el segundo término es el **residuo**.

Un valor pequeño para el residuo $E_{X_i}(V_{\mathbf{X}_{\sim i}}(Y|X_i))$ o un valor grande para el efecto principal $V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))$ significa que X_i es un factor importante sobre Y . También se cumple por la ecuación anterior que $V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i)) \leq V(Y)$. La varianza condicionada $V_i = V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))$ es un valor comprendido entre cero (cuando X_i no contribuye en la formación de Y) y la varianza no condicional de Y , $V(Y)$ (cuando el resto de factores no influyen en la creación de Y).

A partir de la varianza condicional, se define el coeficiente de sensibilidad S_i conocido como el **coeficiente de sensibilidad de primer orden de X_i sobre Y^2** (Saltelli et al., 2008):

$$S_i = \eta^2 = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)} \quad (3.5)$$

Los coeficientes S_i cuantifican la importancia de un factor de entrada X_i sobre el factor de salida Y .

En la fórmula anterior se evidencia la identidad entre S_i y el coeficiente ratio definido por Karl Pearson a principios de 1905 (Pearson, 1905). Se trata de una medida de la influencia no lineal de un factor input X_i sobre un factor output Y , especialmente para los casos en que mediante la regresión lineal se obtienen unos valores de R^2 relativamente bajos (esto ocurre cuando el grado de linealidad entre el input y el output es bajo).

La esperanza condicional $E_{\mathbf{X}_{\sim i}}(Y|X_i)$ del numerador de la expresión (3.5) puede ser una función lineal o no lineal en X_i . En el caso particular en que $E_{\mathbf{X}_{\sim i}}(Y|X_i)$ es una función lineal en X_i :

$$E_{\mathbf{X}_{\sim i}}(Y|X_i) = a_i + b_i * X_i$$

mediante el procedimiento de mínimos cuadrados se estima $b_i = cov(Y, X_i)/V(X_i)$. Sea $\hat{\beta}_i = b_i * (\sqrt{V(X_i)}/\sqrt{V(Y)})$ el coeficiente de regresión estandarizado de b_i . Por tanto, el coeficiente de sensibilidad de primer orden de X_i sobre Y coincide con el coeficiente de regresión estandarizado al cuadrado de la regresión lineal de Y en X_i :

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)} = \frac{b_i^2 * V(X_i)}{V(Y)} = \frac{cov^2(Y, X_i)}{V(X_i) * V(Y)} = \hat{\beta}_i^2$$

En la práctica, el modelo $E_{\mathbf{X}_{\sim i}}(Y|X_i)$ no suele ser una función lineal en X_i , por tanto, el coeficiente de regresión estandarizado no recoge toda la reducción en varianza del IC al fijar cada uno de los factores input. Es por eso que para el cálculo de los coeficientes S_i resulta conveniente el uso de un modelo de GSA basado en el cálculo de varianzas aplicable a cualquier modelo independientemente de su forma funcional.

Sacks et al. (1989) y Sobol' (1993) definieron los coeficientes de análisis de sensibilidad global S_i a partir de la descomposición de f en un conjunto de funciones de creciente

² En la literatura también se le conoce con el nombre de “efecto principal de X_i en Y ”, “medida de importancia de X_i en Y ” o “peso efectivo de X_i en Y ”

dimensionalidad, conocida como *High Dimensional Model Representation* (HDMR). A continuación se describe el procedimiento analítico.

Sea $Y \equiv f(\mathbf{X}) = f(X_1, X_2, \dots, X_n)$ una función definida en un compacto unitario $K^n \equiv [0,1]^n$ donde X_i son los factores input con dominio de variabilidad U y Y es el factor output.

La función $f(\mathbf{X})$ se puede descomponer mediante la representación HDMR como sigue (Rabitz et al., 1999):

$$f(\mathbf{X}) = f(X_1, X_2, \dots, X_n) = f_0 + \sum_i f_i(X_i) + \sum_{i < j} f_{i,j}(X_i, X_j) + \dots + f_{1,2,\dots,n}(X_1, X_2, \dots, X_n) \quad (3.6)$$

donde cada término de la descomposición es función de los factores input, es decir, $f_i(X_i)$ es función de X_i , $f_{i,j}(X_i, X_j)$ es función de X_i y X_j , y así sucesivamente.

El factor f_0 es una constante que denota el efecto de orden cero, la función $f_i(X_i)$ denota el efecto del factor input X_i sobre f cuando actúa de forma independiente al resto de factores input, la función $f_{i,j}(X_i, X_j)$ describe el efecto de interacción de los inputs X_i, X_j sobre f y el resto de términos de mayor grado describen el efecto conjunto de los factores input que actúan en el término correspondiente sobre Y . Finalmente, el último término $f_{1,2,\dots,n}(X_1, X_2, \dots, X_n)$ indica la dependencia residual de todos los factores input fijados de forma que tienen un efecto conjunto sobre el modelo f .

Si las interacciones entre los inputs no tienen efecto sobre el modelo, la descomposición sólo se define mediante el término de orden cero f_0 y los términos de primer orden $f_i(X_i)$ que son funciones de los factores input X_i . En este caso, el modelo se llama aditivo.

El objetivo, en principio, es buscar descomposiciones que den una buena aproximación a la función $f(\cdot)$ en norma L_2 :

$$J \equiv \int_{K^n} \left[f(X) - f_0 - \sum_i f_i(X_i) - \sum_{i < j} f_{i,j}(X_i, X_j) - \dots - \sum_{i_1 < \dots < i_s} f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s}) \right]^2 dX,$$

con $s \leq I$.

Sin embargo, la forma de las funciones de la descomposición (3.6) no es única.

Cuando los factores input del modelo son independientes y se cumple la siguiente condición de partida:

$$\int_0^1 f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s}) dX_j = 0, \quad j \in \{i_1, i_2, \dots, i_s\}, s \in \{1, \dots, I\}$$

entonces la descomposición HDMR cumple las siguientes propiedades:

1) $\int_{K^n} f(\mathbf{X})dX = E(Y) = f_0 \rightarrow f_0 =$ valor medio de la función $f(\mathbf{X})$

2) Todo par de términos de la descomposición HDMR son ortogonales:

$$\int_{K^n} f_{i_1, \dots, i_s}(X_{i_1}, \dots, X_{i_s}) f_{j_1, \dots, j_p}(X_{j_1}, \dots, X_{j_p}) dX = 0 \text{ con } \{i_1, \dots, i_s\} \neq \{j_1, \dots, j_p\}$$

3) La descomposición HDMR es única

En este caso, la descomposición HDRM de la función $f(\cdot)$ viene dada de la siguiente forma (Rabitz et al., 1999)

$$f_0(X_0) := \int_{K^n} f(\mathbf{X})dX = E(Y)$$

$$f_i(X_i) := \int_{K^{n-1}} f(\mathbf{X}) \prod_{j \neq i} dX_j - f_0 = E_{\mathbf{X}_{-i}}(Y|X_i) - f_0$$

$$f_{i,j}(X_i, X_j) := \int_{K^{n-2}} f(\mathbf{X}) \prod_{k \notin \{i,j\}} dX_k - f_i(X_i) - f_j(X_j) - f_0$$

$$= E_{\mathbf{X}_{-ij}}(Y|X_i, X_j) - E_{\mathbf{X}_{-i}}(Y|X_i) - E_{\mathbf{X}_{-j}}(Y|X_j) - f_0$$

⋮

$$f_{i_1, \dots, i_r}(X_{i_1}, \dots, X_{i_r}) := \int_{K^{n-r}} f(\mathbf{X}) \prod_{k \notin \{i_1, \dots, i_r\}} dX_k - \sum_{j_1 < \dots < j_{r-1} \subset \{i_1, i_2, \dots, i_r\}} f_{j_1, \dots, j_{r-1}}(X_{j_1}, \dots, X_{j_{r-1}})$$

A esta descomposición única se le conoce con el nombre de ANOVA-HDMR (Archer y Saltelli, 1997; Rabitz et al., 1999), puesto que es la definición de la descomposición ANOVA (Análisis de la Varianza) del factor output. Se trata de la descomposición que mejor aproxima en norma L_2 a la función $f(\cdot)$.

La aplicación de la descomposición ANOVA-HDMR es muy útil para medir el efecto de la varianza de cada uno de los factores input de la función $f(\cdot)$ en la varianza total del factor output Y .

Considerando que los factores input son independientes, el esquema de descomposición de la varianza incondicional del output Y , equivalente a la descomposición HDMR, se define como sigue (Sobol', 1993):

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{i,j} + \dots + V_{1,2,\dots,n} \tag{3.7}$$

donde:

$$\begin{aligned}
V_i &= V_{X_i} [E_{\mathbf{X}_{\sim i}}(Y|X_i)], \\
V_{i,j} &= V_{X_i X_j} [E_{\mathbf{X}_{\sim ij}}(Y|X_i, X_j)] - V_i - V_j, \\
V_{i,j,l} &= V_{X_i X_j X_l} [E_{\mathbf{X}_{\sim ijl}}(Y|X_i, X_j, X_l)] - V_{i,j} - V_{i,l} - V_{j,l} - V_i - V_j - V_l \\
&\dots
\end{aligned}$$

El término $V_{X_i X_j} [E_{\mathbf{X}_{\sim ij}}(Y|X_i, X_j)]$ mide el efecto conjunto del par de factores input (X_i, X_j) sobre el output Y . De forma análoga, se pueden calcular los términos de interacción de órdenes superiores. La notación $\mathbf{X}_{\sim i}$ en $V_{X_i} [E_{\mathbf{X}_{\sim i}}(Y|X_i)]$ indica la matriz de todas los factores input exceptuando X_i . El operador $E_{\mathbf{X}_{\sim i}}$ denota la esperanza del factor output Y sobre todos los posibles valores de $\mathbf{X}_{\sim i}$ manteniendo fijo X_i . La varianza se calcula sobre todos los posibles valores de X_i .

Cuckier et al. (1973) y Sobol' (1993) estiman de forma analítica los términos de esta descomposición para el cálculo de los coeficientes de sensibilidad. Normalizando por $V(Y)$ la ecuación (3.7) se obtiene la siguiente identidad:

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{i,j} + \dots + S_{1,2,\dots,n} \quad (3.8)$$

donde cada término de la descomposición corresponde a los índices de sensibilidad definidos en el análisis de sensibilidad global.

Si además de cumplirse la independencia entre los factores input no existen interacciones entre ellos en el modelo, se cumple $\sum_{i=1}^n V_i = V(Y)$ y, por tanto, $\sum_{i=1}^n S_i = 1$.

Si el modelo no es aditivo, se calculan el resto de coeficientes de orden superior. Los coeficientes que miden el efecto total a la variación del output Y debido a un factor input X_i , es decir, debido a su coeficiente de primer orden y a los efectos de mayor orden provocados por las interacciones, se llaman **coeficientes de sensibilidad total**. Por ejemplo, supóngase que se dispone de $n = 3$ factores input independientes, los tres coeficientes de sensibilidad total se calculan como sigue:

$$S_{T1} = S_1 + S_{1,2} + S_{1,3} + S_{1,2,3} \quad (3.9)$$

donde $S_1 + S_{1,2} + S_{1,3} + S_{1,2,3}$ es la suma de todos los términos de la ecuación (3.8) para este caso.

De forma análoga,

$$\begin{aligned}
S_{T2} &= S_2 + S_{1,2} + S_{2,3} + S_{1,2,3} \\
S_{T3} &= S_3 + S_{1,3} + S_{2,3} + S_{1,2,3}
\end{aligned}$$

Sin embargo, no se suelen estimar los coeficientes de orden superior calculando todos los términos de la fórmula (3.8) puesto que en un modelo con n factores el número total de coeficientes a estimar es $2^n - 1$ (incluidos los de primer orden S_i).

Por tanto, se define el **coeficiente de sensibilidad total del factor X_i** como sigue (Saltelli et al., 2008):

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}(V_{X_i}(Y|\mathbf{X}_{\sim i}))}{V(Y)} = 1 - \frac{V_{\mathbf{X}_{\sim i}}(E_{X_i}(Y|\mathbf{X}_{\sim i}))}{V(Y)}$$

Para un factor dado X_i , toda diferencia significativa entre S_i y S_{Ti} indica que la interacción entre factores input es importante para el factor output Y . El cálculo de S_{Ti} se puede aplicar a cualquier modelo independientemente del grado de correlación entre los factores input, sin embargo la ecuación (3.9) sólo es válida cuando existe independencia en los factores inputs.

Para un modelo no aditivo en el que existen interacciones entre los factores input, los coeficientes de sensibilidad de primer orden, o efectos principales, no recogen la variabilidad del output del modelo Y y, por tanto, $\sum_{i=1}^n S_i \leq 1$.

Los coeficientes de sensibilidad definidos mediante el cálculo de varianzas satisfacen el objetivo del GSA que es identificar cuáles son los factores input con mayor efecto sobre la construcción del factor output Y .

Análisis de incertidumbre

En primer lugar, para poder aplicar un análisis de incertidumbre se deben identificar las fuentes de incertidumbre en los inputs y simular diferentes escenarios según éstas. Para realizar el análisis de incertidumbre en el campo de los ICs se sugiere la aplicación de la técnica Monte Carlo (Nardo et al., 2008) que consiste en perturbar todas las fuentes de incertidumbre y analizar los efectos derivados de dichas variaciones sobre el modelo, de tal forma que sea posible estimar una función de distribución para $Ranking(IC_j)$ y/o \bar{R}_S . La técnica Monte-Carlo se basa en los siguientes pasos:

- 1) Sin pérdida de generalidad, supóngase que se plantean las siguientes tres fuentes de variabilidad (inputs) con un número finito de alternativas a considerar:
 - a. X_1 : técnica de normalización de los datos
 - b. X_2 : técnica de ponderación de los indicadores simples
 - c. X_3 : técnica de agregación

En primer lugar, se considera una función de densidad de probabilidad para cada factor input según sus posibles alternativas. Supóngase que el factor X_i , pudiendo ser i una de las

tres fuentes de variabilidad anteriores, se compone de n alternativas a ser elegidas, por tanto $X_i \sim U(0,1)$. Sea $\alpha \in [0,1]$ el número aleatorio del que partirá el algoritmo. Por tanto, se selecciona una de las n opciones según α quede incluido dentro de los intervalos siguientes:

α	X_i
Si $\alpha \in \left[0, \frac{1}{n}\right]$	Se elige la “opción 1” para X_i
Si $\alpha \in \left[\frac{1}{n}, \frac{2}{n}\right]$	Se elige la “opción 2” para X_i
...	...
Si $\alpha \in \left[\frac{n-1}{n}, 1\right]$	Se elige la “opción n ” para X_i

- 2) Se genera un número N de muestras compuestas por las combinaciones de factores $X^k, 1 \leq k \leq N$, con $X^k = \{X_1^k, X_2^k, X_3^k\}$, siendo X_i^k la opción obtenida para X_i en la simulación k . De esta forma, se obtiene para cada muestra el valor del IC y se calcula el escalar Y^k que puede ser cualquiera de las variables output del análisis de incertidumbre, *Ranking*(IC_j) ó \bar{R}_s
- 3) Se calcula $Y^k, 1 \leq k \leq N$ para todas las combinaciones. La secuencia de $Y^k, 1 \leq k \leq N$ proporciona la distribución de probabilidad estimada de la variable output. Las características de esta distribución, tales como la media, la varianza y momentos de orden superior se pueden estimar con un nivel arbitrario de precisión relacionado con el tamaño de la simulación N . Estos valores son los que se analizan para constatar el grado de incertidumbre del IC frente a los cambios considerados.

La generación de muestras puede realizarse empleando diferentes procedimientos tales como el muestreo aleatorio simple, el muestreo cuasialeatorio, el muestreo estratificado u otro que se considere conveniente (Saltelli et al., 2008).

Algunos ejemplos de ICs sobre los que se ha aplicado un análisis de incertidumbre y/o un análisis de sensibilidad basado en el uso de la varianza son: índice de avance tecnológico (Saisana, Saltelli y Tarantola, 2005; Cherchye et al., 2008); índice que mide la disposición de medios para el comercio electrónico (*e-business*) (Tarantola et al., 2006); índice que mide el aprendizaje en la educación superior (Saisana, 2008); índice para evaluar la gobernanza en países africanos (Saisana, Annoni y Nardo, 2009); ranking de Shanghai y ranking THES en educación superior (Saisana y D’Hombres, 2008); índice de desempeño ambiental (Saisana y Saltelli, 2008); índice de competitividad regional (Annoni y Kozovska, 2010); índice de capacitación de los consumidores (Nardo et al., 2011).

3.4.8. Conexión con otros indicadores

Frecuentemente se construyen ICs que tratan de medir conceptos que están muy relacionados con otros indicadores ya existentes. El objetivo de conectar el IC construido con otros indicadores de gran similitud es corroborar que los resultados obtenidos con el índice son coherentes. La mejor forma de ilustrar estas conexiones es mediante diagramas simples de dispersión.

Para una mejor interpretación de esta etapa se va a exponer un ejemplo con el Índice de Rendimiento Tecnológico (TAI) desarrollado por las Naciones Unidas (2001). Se supone que los países con un alto índice de rendimiento tecnológico son aquellos que tienen un PIB per cápita más elevado. Para comprobar dicha afirmación se realiza un diagrama de dispersión entre ambos indicadores y un análisis de correlación para determinar la relación existente entre ambos (Nardo et al., 2008). Si no se obtiene el resultado previsto se debe justificar cualquier anomalía. En el ejemplo expuesto se debería justificar el comportamiento de Corea y Noruega.

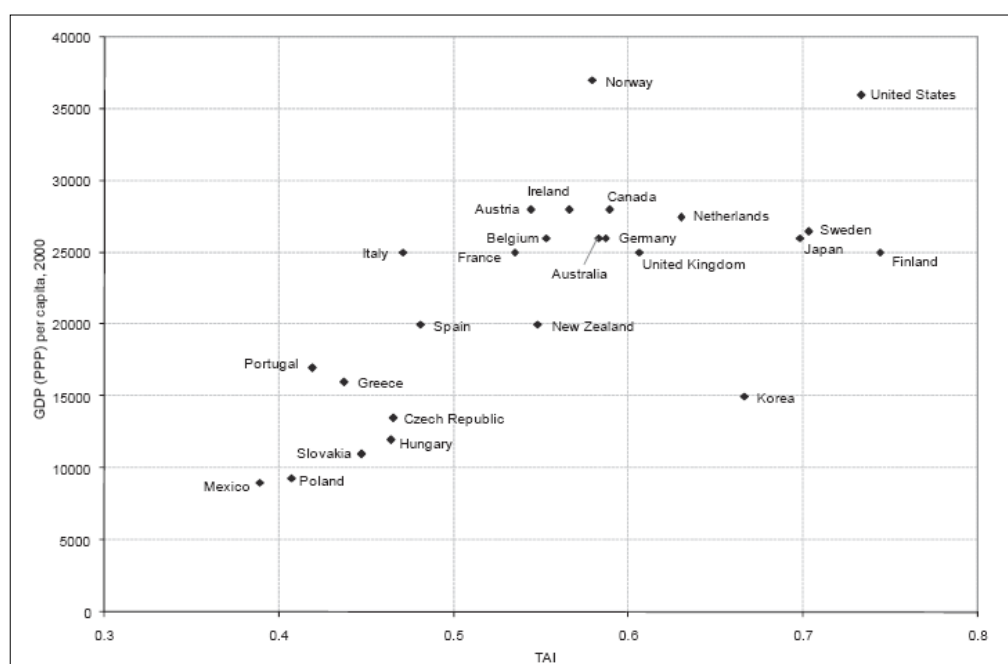


Figura 3.9. TAI vs. GDP (PPP)

3.4.9. Identificación de los indicadores simples en el Indicador Compuesto

Los ICs representan un punto de partida para el análisis de las conclusiones. Pero, aparte de poder interpretar el valor final de éstos también se pueden descomponer para identificar la

contribución de las dimensiones y de los indicadores simples a su formación con el fin de definir políticas de mejora para cada unidad de análisis.

A continuación se muestra un ejemplo de las posibles representaciones que se podrían realizar para identificar las contribuciones de los indicadores simples al Índice de Rendimiento Tecnológico (TAI) (Nardo et al., 2008).

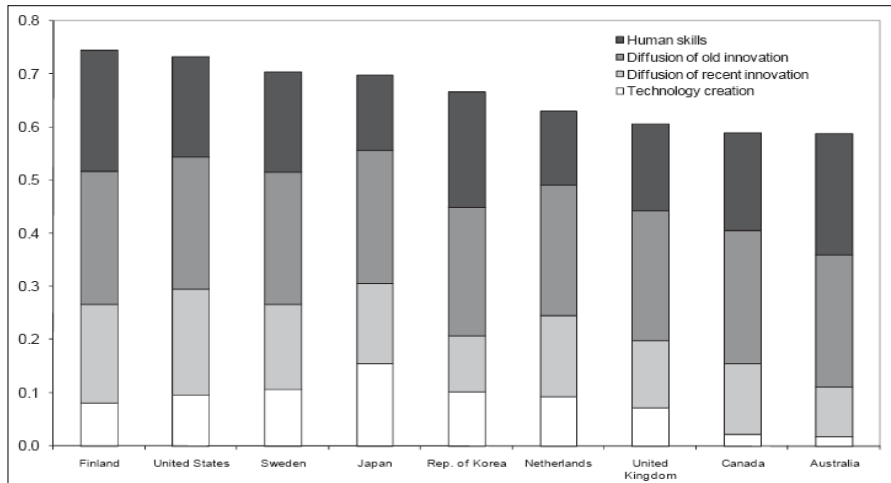


Figura 3.10. TAI por dimensiones

En la figura 3.10 se representa la descomposición del IC TAI según las diferentes dimensiones. Las barras se han construido estandarizando los valores de las dimensiones y multiplicándolos por sus pesos. La suma de las cuatro dimensiones es el valor del IC.

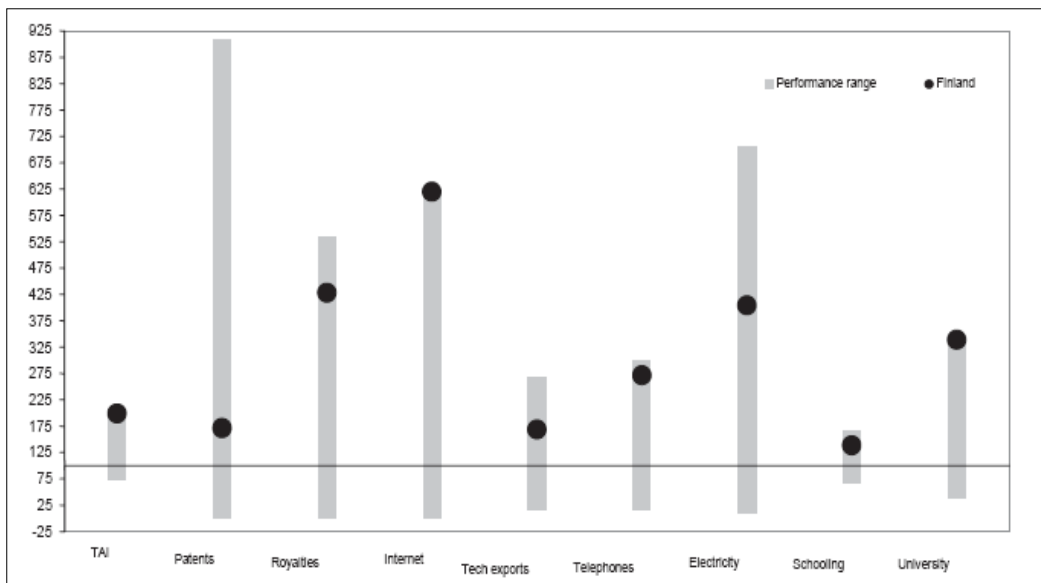


Figura 3.11. TAI (Finlandia)

En la figura 3.11 se muestra una única unidad de análisis, Finlandia. Las barras grises representan el rango de variación de cada indicador simple y el punto negro la situación de Finlandia en cada caso. De esta forma, se identifican los indicadores en los que Finlandia presenta una puntuación más elevada y en los que tiene un valor más bajo con respecto al resto de países.

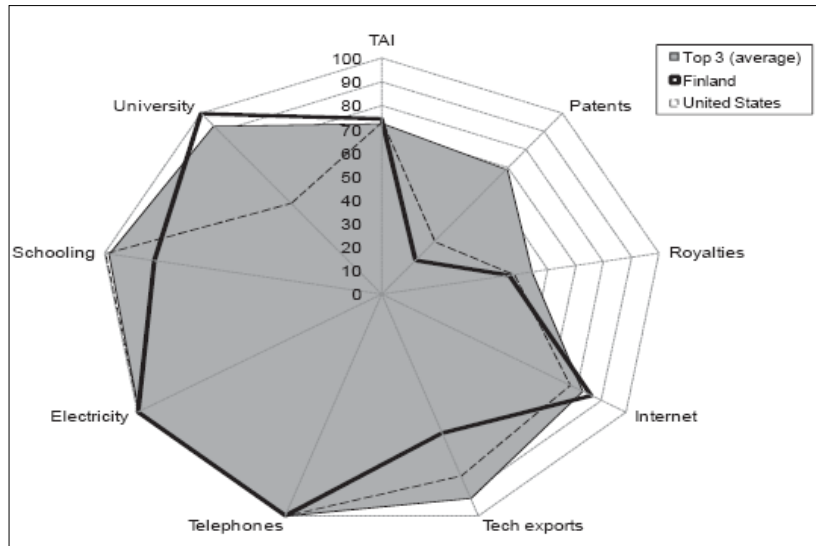


Figura 3.12. TAI (Finlandia) vs. TAI (EE.UU.) vs. TAI (Valor medio)

Por último, se compara Finlandia con la media de las tres unidades de análisis con mayor puntuación en el índice TAI y con los Estados Unidos. En la figura 3.12 la unidad de análisis con mejor rendimiento toma el valor de 100 para cada indicador, mientras que la de peor rendimiento toma el valor de 0.

El objetivo principal de esta etapa es estudiar el rendimiento de las unidades de análisis a través de los indicadores simples y de las dimensiones con el fin de revelar cuáles son los más influyentes en la puntuación final del IC.

3.4.10. Visualización de los resultados

La forma de representar un IC no es una tarea fácil. Es muy importante transmitir y diseminar de la manera más entendible posible los resultados obtenidos. Los grupos de interés a los que van dirigidos los ICs suelen ser diversos por lo que es necesario considerar una estrategia comunicacional con el fin de favorecer su amplia diseminación.

Los tomadores de decisiones suelen invertir poco tiempo en el análisis del estudio demandando, por tanto, que los resultados sean presentados de la manera más concisa, sintética y sencilla posible. Por lo que el diseño visual de su presentación debe proveer señales claras que

alerten y expongan situaciones extraordinarias que permitan identificar posibles áreas de intervención.

A continuación se exponen las diferentes formas de representación de un IC más empleadas (Nardo et al., 2005).

PRESENTACIÓN EN TABLAS

La presentación en tablas es la forma más simple de mostrar la información. Para cada unidad de análisis se presenta su IC y sus indicadores simples. Uno de los aspectos a tener en cuenta cuando se muestran datos tabulados es la importancia de adoptar algún criterio para enseñar la información de manera ordenada. Normalmente se suelen agrupar las unidades de análisis según comportamientos similares entre sí con respecto a algún indicador simple de interés.

MONITORING HUMAN DEVELOPMENT: ENLARGING PEOPLE'S CHOICES . . .										
1 Human development index										
HDI rank ^a	Life expectancy at birth (years) 2002	Adult literacy rate (% ages 15 and above) 2002 ^b	Combined gross enrolment ratio for primary, secondary and tertiary schools (%) 2001/02 ^c	GDP per capita (PPP US\$) 2002	Life expectancy index	Education index	GDP index	Human development index (HDI) value 2002	GDP per capita (PPP US\$) rank minus HDI rank ^d	
High human development										
1	Norway	78.9	.. ^e	98 ^f	36,600	0.90	0.99	0.99	0.956	1
2	Sweden	80.0	.. ^e	114 ^{g,h}	26,050	0.92	0.99	0.93	0.946	19
3	Australia	79.1	.. ^e	113 ^{g,h}	28,260	0.90	0.99	0.94	0.946	9
4	Canada	79.3	.. ^e	95 ^f	29,480	0.90	0.98	0.95	0.943	5
5	Netherlands	78.3	.. ^e	99 ^f	29,100	0.89	0.99	0.95	0.942	6
6	Belgium	78.7	.. ^e	111 ^{f,g}	27,570	0.90	0.99	0.94	0.942	7
7	Iceland	79.7	.. ^e	90 ^f	29,750	0.91	0.96	0.95	0.941	1
8	United States	77.0	.. ^e	92 ^h	35,750	0.87	0.97	0.98	0.939	-4
9	Japan	81.5	.. ^e	84 ^h	26,940	0.94	0.94	0.93	0.938	6
10	Ireland	76.9	.. ^e	90 ^f	36,360	0.86	0.96	0.98	0.936	-7
11	Switzerland	79.1	.. ^e	88 ^f	30,010	0.90	0.95	0.95	0.936	-4
12	United Kingdom	78.1	.. ^e	113 ^{f,g}	26,150	0.88	0.99	0.93	0.936	8
13	Finland	77.9	.. ^e	106 ^{f,g}	26,190	0.88	0.99	0.93	0.935	6
14	Austria	78.5	.. ^e	91 ^f	29,220	0.89	0.96	0.95	0.934	-4
15	Luxembourg	78.3	.. ^e	75 ^{f,i}	61,190 ^j	0.89	0.91	1.00	0.933	-14
16	France	78.9	.. ^e	91 ^f	26,920	0.90	0.96	0.93	0.932	0
17	Denmark	76.6	.. ^e	96 ^f	30,940	0.86	0.98	0.96	0.932	-12
18	New Zealand	78.2	.. ^e	101 ^{g,h}	21,740	0.89	0.99	0.90	0.926	6
19	Germany	78.2	.. ^e	88 ^h	27,100	0.89	0.95	0.94	0.925	-5
20	Spain	79.2	97.7 ^{k,l,k}	92 ^h	21,460	0.90	0.97	0.90	0.922	5
21	Italy	78.7	98.5 ^{k,l,k}	82 ^f	26,430	0.89	0.93	0.93	0.920	-3
22	Israel	79.1	95.3	92	19,530	0.90	0.94	0.88	0.908	5
23	Hong Kong, China (SAR)	79.9	93.5 ^{l,k}	72	26,910	0.91	0.86	0.93	0.903	-6
24	Greece	78.2	97.3 ^{k,l,k}	86 ^f	18,720	0.89	0.95	0.87	0.902	5
25	Singapore	78.0	92.5 ^l	87 ^m	24,040	0.88	0.91	0.92	0.902	-3

Figura 3.13. Presentación del Índice de Desarrollo Humano en tablas

GRÁFICOS DE BARRAS

En los gráficos de barras las unidades de análisis se representan en el eje vertical y los valores del IC en el eje horizontal tal y como muestra la siguiente figura.

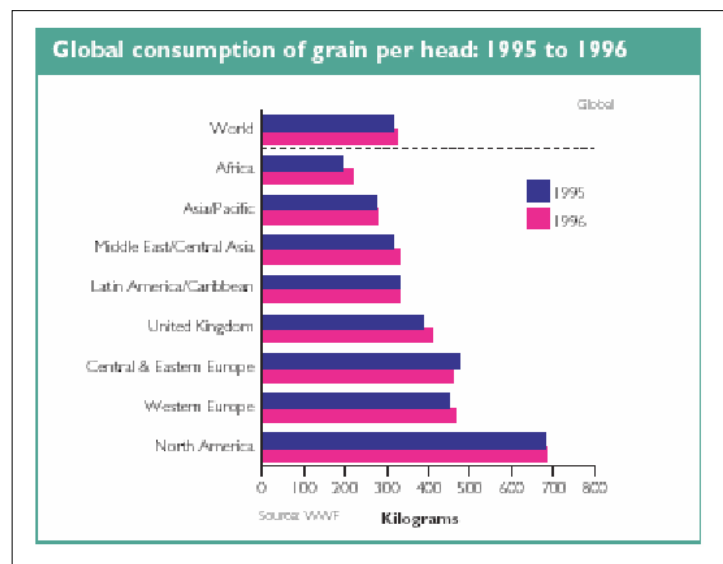


Figura 3.14. Presentación del Índice de Consumo mundial de cereales per cápita en gráfico de barras

Una de las ventajas que presenta el gráfico de barras es que es posible expresar los valores del indicador en diferentes periodos de tiempo. Es importante ordenar las unidades de análisis de mayor a menor (o al revés) y puede ser conveniente fijar líneas de corte en las que se exprese el valor medio sobre toda la población. Este formato es claro y de fácil comprensión.

GRÁFICOS DE LÍNEAS

Los gráficos de líneas se suelen usar para presentar el rendimiento de las unidades de análisis en varios momentos de tiempo. Sin embargo, uno de los inconvenientes más importantes de este formato es que la base de datos del análisis contenga muchas unidades de referencia.

En este tipo de gráficos se puede representar la evolución de los niveles absolutos del indicador, los porcentajes o tasas de variación en relación al año anterior, la evolución del índice con respecto a un año base, etc.

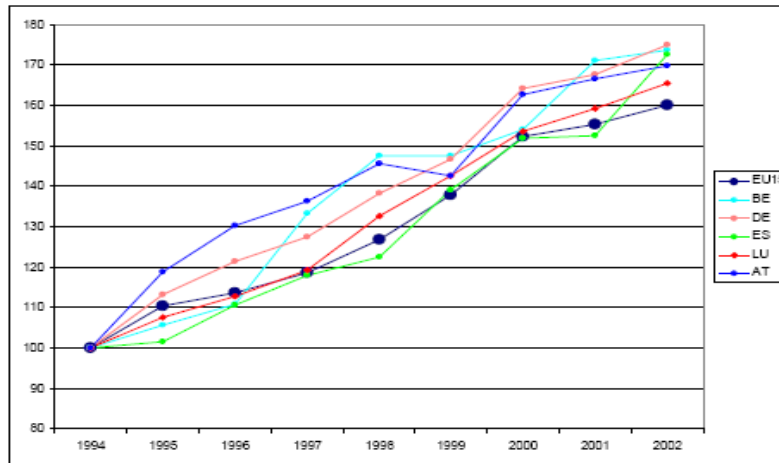


Figura 3.15. Presentación del Índice de Mercado Interno en gráfico de líneas

SEÑALES DE TRÁFICO

Se representan los resultados de cada valor del IC, e incluso si se considera necesario de cada valor de los indicadores simples, mediante señales de tráfico o emoticonos. Esto se puede realizar cuando la información está muy ligada al desempeño de metas concretas o cuando interesa expresar los cambios acontecidos. Cuando se usa este formato es necesario establecer un criterio en el que se explique el significado de cada color o emoticono.

Assessment for indicator against objective		Change since 1970	Change since 1990	Change since Strategy ¹
H1	Economic output	✓	✓	✓
H2	Investment	≈	≈	≈
H3	Employment	≈	≈	✓
H15	Waste	***	***	***
	All arisings and management	***	***	***
	Household waste	***	✗	✗

Figura 3.16. Presentación del Índice de Sostenibilidad desarrollado por el gobierno británico en señales de tráfico

Key	
✓	Significant change, in direction of meeting objective
≈	No significant change
✗	Significant change, in direction away from meeting objective
***	Insufficient or no comparable data

Figura 3.17. Leyenda de las señales de tráfico

RANKINGS

Es una forma rápida de expresar los resultados. Normalmente se colocan las unidades de análisis de forma decreciente según las puntuaciones del IC (a la unidad de análisis con mayor puntuación en el IC se le asigna la primera posición en el ranking y así se va completando la tabla sucesivamente). Este formato de presentación carece de mucha información puesto que no se pueden visualizar las diferencias entre unidades de análisis. Si se dispusiera de información por años, sería conveniente mostrar los cambios en términos de posiciones ganadas o perdidas durante el intervalo de tiempo.

GROWTH COMPETITIVENESS INDEX RANKINGS			
Country	Growth Competitiveness ranking 2003	Growth Competitiveness ranking 2003 among GCR 2002 countries	Growth Competitiveness ranking 2002*
Finland	1	1	1
United States	2	2	2
Sweden	3	3	3
Denmark	4	4	4
Taiwan	5	5	6
Singapore	6	6	7
Switzerland	7	7	5
Iceland	8	8	12
Norway	9	9	8
Australia	10	10	10
Japan	11	11	16
Netherlands	12	12	13
Germany	13	13	14
New Zealand	14	14	15
United Kingdom	15	15	11
Canada	16	16	9
Austria	17	17	18
Korea	18	18	25
Malta	19	—	—
Israel	20	19	17
Luxembourg	21	—	—
Estonia	22	20	27

Figura 3.18. Presentación del Índice de Competitividad en forma de ranking

GRÁFICOS RADAR O DIAGRAMAS DE TELARAÑA

Es una herramienta muy útil para mostrar visualmente las brechas entre el estado actual y la situación ideal a alcanzar. En el gráfico se ubican los valores numéricos aumentando en valor desde el centro hasta el perímetro que es donde alcanza su valor máximo permitido. Para poder hacer comparaciones entre los diferentes indicadores es conveniente normalizar todas las variables sobre la base de una misma escala. Pero la principal limitación de este formato de presentación es que sólo se puede mostrar el desempeño de una unidad de análisis en un instante particular.

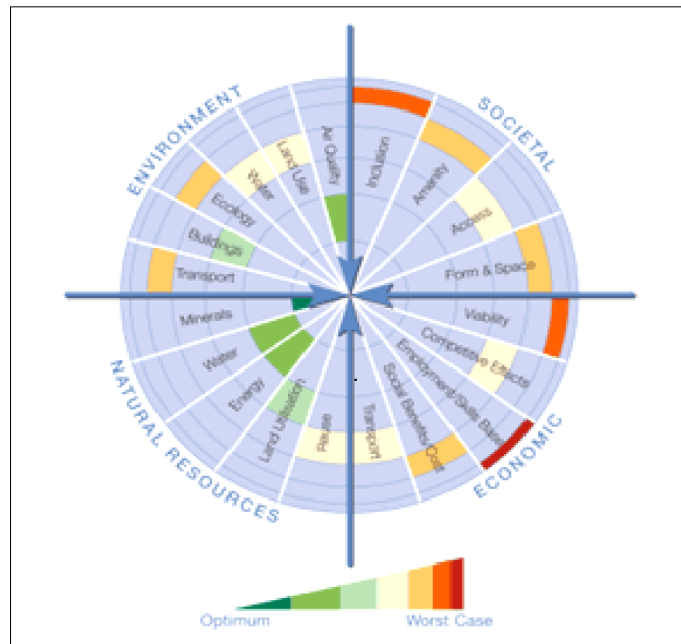


Figura 3.19. Presentación del Índice de Sostenibilidad del grupo SPEar en gráfico de radar

Capítulo 4

El papel de los indicadores en el contexto académico de la educación superior

4.1. El concepto de ‘calidad universitaria’

4.1.1. Diferentes enfoques del concepto ‘calidad’ y ‘calidad universitaria’

En la actualidad, y cada vez de forma más acentuada, se ha constatado un interés creciente por la calidad de servicios y productos en cualquier actividad económica. En países como Japón y Estados Unidos está muy presente la tendencia a la medición de la calidad, mientras que en España es un concepto relativamente reciente pero cada vez más integrado en la cultura y en la práctica tanto para los responsables de empresas e instituciones como para políticos y la sociedad en general.

Desde hace muchos años existe un gran debate en la literatura sobre la definición del concepto ‘calidad’. Sin embargo, no se ha llegado a una única definición del concepto debido a la dificultad de encontrar un consenso por su concreción o relatividad del término reflejada en la conocida y utilizada cita del filósofo Pirsing (1974):

“Quality...you know what it is, yet you don’t know what it is. But that’s self contradictory. But some things are better than others, that is, they have more quality. But when you try to say what the quality is, apart from the things that have it, it all goes proof! There’s nothing to talk about. But if you can’t say what Quality is, how do you know what it is, or how do you know that it even exist? If no one knows what it is, then for all practical purposes it really does exist. What else are the grades based on? Why else would people pay fortunes for some things are better than other... but what’s the «betterness»?... So round and round you go spinning mental wheels and nowhere finding anyplace to get traction. What the hell is Quality? What is it?”

En la línea de Pirsing, muchos autores han aportado diferentes concepciones del término 'calidad'. Deming (1989) define 'calidad' como la traducción de las necesidades futuras de los usuarios en características medibles, de modo que solo así un producto puede ser diseñado y fabricado para dar satisfacción a un precio que el cliente pagará. Al mismo tiempo, Juran (1990) expone que la palabra 'calidad' tiene múltiples significados entre los que destaca dos de ellos: la calidad consiste en aquellas características del producto que se basan en las necesidades del cliente y que por eso brindan satisfacción del producto y la calidad consiste en libertad después de las deficiencias.

De manera sintética, según Ishikawa (1986), el término 'calidad' significa calidad del producto. Más específicamente, el concepto 'calidad' incluye calidad del trabajo, calidad del servicio, calidad de la información, calidad del proceso, calidad del sistema, calidad de la compañía, calidad de los objetivos, etc. Así pues, trabajar en calidad consiste en diseñar, producir y servir un producto o servicio que sea útil, lo más económico posible y siempre satisfactorio para el usuario. Por otro lado, según la norma UNE-EN ISO 9000:2005, la 'calidad' se define como el grado en que un conjunto de características inherentes cumple con los requisitos.

Harvey y Green (1993) describen cinco diferentes concepciones del concepto 'calidad' que pueden aplicarse en el campo de la educación superior en el sentido que pueden permitir su medida y evaluación. Estas cinco concepciones son: calidad como fenómeno excepcional, calidad como perfección o coherencia, calidad como ajuste a un propósito, calidad como valor agregado y calidad como transformación (cambio cualitativo).

El término 'calidad' es usado en todos los campos de la vida moderna vinculándose a todos los sectores de economía, nacionales e internacionales, y en los que no se puede excluir a las universidades y a otras instituciones de educación. La garantía de la calidad es una necesidad inevitable por la responsabilidad social de las universidades y de los gobiernos (Mora, 2011).

Sin embargo, la preocupación por la 'calidad universitaria' no siempre ha derivado en una mayor precisión a la hora de definirla e interpretarla. Según Westerheijden (1990), el concepto de 'calidad universitaria' se engloba en un marco de referencia definido por las siguientes proposiciones:

- La 'calidad universitaria' es un concepto relativo y multidimensional en relación con los objetivos y actores del sistema universitario
- No existe una única percepción del concepto 'calidad universitaria'. El único posible nivel de concreción será el que adopte a partir del enfoque evaluativo desde el que se valore la calidad

Se podrían describir muchas más definiciones de autores con diferentes enfoques del término calidad en las instituciones de educación superior. A partir de diversos trabajos se han extraído

varias acepciones de lo que se ha entendido a lo largo de los años por 'calidad universitaria'. Mora (1991b), referenciando los trabajos de George (1982) y Astin (1985), resume la conceptualización de la 'calidad universitaria' en cinco enfoques evaluativos.

En primer lugar se distingue el concepto de **calidad como reputación**. En este caso, la calidad se evalúa mediante encuestas realizadas al profesorado de otras instituciones a los que se les supone la capacidad de juzgar sobre la calidad de cada institución desde un punto de vista subjetivo. Sin embargo, el hecho de que la evaluación de las instituciones se base en juicios de valor subjetivos hace que este enfoque reciba serias críticas sobre su uso.

Otro punto de vista sobre la calidad de una institución es medir la **disponibilidad de recursos** para la realización de las funciones básicas de la universidad que son la docencia, la investigación y la gestión. Se trata, por tanto, de un enfoque basado en medidas objetivas. Cuatro tipos básicos de recursos son habitualmente considerados en este enfoque: personal, medios físicos, recursos financieros y estudiantes. Sin embargo, si se sigue este enfoque se tiende a evaluar mejor a las universidades grandes con mayores recursos, lo cual no significa que sean mejores universidades.

La **calidad a través de los resultados** se basa en evaluar a los graduados una vez formados y preparados para su incorporación al mundo laboral. Esta visión sería la más adecuada para evaluar lo que las universidades aportan a los estudiantes que pasan por ella. Sin embargo, existen diversos problemas en el uso de este enfoque. Por una parte, las universidades con mayor reputación atraen a los estudiantes más preparados y, por consiguiente, producen a estudiantes con mejores posibilidades de empleo. Además, las mediciones a través de los resultados se hacen con demasiado retraso y, por tanto, se ralentiza uno de los objetivos prioritarios que es la mejora de la evaluación de la calidad.

Otra manera de interpretar la calidad de una institución es considerar la **calidad del contenido** de la institución, es decir, la calidad de lo que se enseña: nivel docente, currículum, sistema pedagógico, clima de la universidad, etc. Este enfoque presenta la dificultad de evaluar los conceptos implicados en el proceso de aprendizaje.

Finalmente, el último enfoque definido por los autores para el concepto de calidad universitaria es la **calidad por el 'valor añadido'**. Este enfoque se refiere a la contribución de la institución a la formación del estudiante, es decir, la diferencia entre lo que el estudiante ha aprendido al terminar la universidad y lo que sabía al ingresar en ésta. Esta visión del concepto de calidad parece que sea la más adecuada, sin embargo, el proceso de medición presenta un problema al intentar aislar los efectos de una institución de otros efectos ajenos a ésta.

El problema del enfoque óptimo que mejor define la calidad de las universidades sigue sin resolverse. Por tanto, podría afirmarse que la búsqueda de la calidad tiene sentido a través de todas las medidas indicadas y no a través de cada una de ellas de forma independiente, ya que

cada una contiene información válida y puede contribuir a la mejora de la definición de la calidad de la educación superior como un todo.

En esta línea, Grao y Winter (1999) afirman que hay que medir la calidad de las universidades entendiendo esta calidad como un objetivo global de una organización y no como la suma de distintas unidades independientes.

Según De la Orden et al. (1997), la búsqueda de la excelencia de la calidad es un argumento utópico, ya que no es un período sino un proceso continuo de mejora donde no hay un último escalón de la calidad y donde todo se puede optimizar.

Por tanto, todas las acepciones para el concepto de calidad expuestas anteriormente, entre otras muchas más, conllevan a la consideración de la calidad como un concepto multidimensional y complejo, ya que tal y como se muestra en la figura 4.1 existen múltiples enfoques hacia ella. Las distintas visiones ofrecen un enfoque completo acerca de los modos, no excluyentes sino complementarios, de acercarse a la comprensión de la calidad universitaria.



Figura 4.1. Enfoques de la calidad en la universidad

Las múltiples definiciones e interpretaciones del concepto de calidad reflejan la ambigüedad del concepto y dificultad de establecer un consenso único en su definición. Pérez (2000) expone varias razones de esta dificultad: el reduccionismo en la definición, las diversas concepciones de partida del servicio a evaluar y los distintos modelos de referencia que pueden estar orientados a funciones diferentes (certificación, mejora continua, acreditación,...). Aunque como señala Quintanilla (1998) se trata de un concepto que todo el mundo entiende a pesar de no existir un consenso único en su definición. Junto a esta manifestación, Bricall (2000) afirma que el hecho de que la calidad sea un concepto multidimensional y complejo no puede servir de excusa para no intentar comprenderlo en sus distintas acepciones y en su cambiante formulación a lo largo de la historia de cualquier institución y, en concreto, la Universidad.

4.1.2. Hacia una nueva concepción de ‘calidad universitaria’

La Declaración de Bolonia, firmada en 1999 por los ministros europeos de educación, produjo un cambio importante en el panorama europeo en lo que se refiere al fomento de la movilidad de los estudiantes, la estructura de las titulaciones, el énfasis de los procesos de enseñanza-aprendizaje, la cooperación entre instituciones, la comparabilidad y reconocimiento de los grados entre los diferentes países miembros y, sobre todo, la preocupación por la garantía de la calidad de las universidades y adopción de una dimensión europea en la enseñanza superior, el Espacio Europeo de Educación Superior (EEES).

En la siguiente afirmación realizada en las primeras líneas de la Declaración de Bolonia ya se mostraba la preocupación por la instauración de políticas de garantía de calidad y rendición de cuentas de las Instituciones de Educación Superior (IES) en el proyecto de construcción del EEES.

“A Europe of Knowledge is now widely recognised as an irreplaceable factor for social and human growth and as an indispensable component to consolidate and enrich the European citizenship, capable of giving its citizens the necessary competences to face the challenges of the new millennium, together with an awareness of shared values and belonging to a common social and cultural space.” (The Bologna Declaration, 1999).

En particular, el debate sobre las políticas de garantía de calidad y de transparencia de las universidades españolas constituye un importante elemento para el éxito del EEES. Por ello, aparece como un objetivo clave tanto del Proceso de Bolonia como de la agenda de la Unión Europea para la modernización de las universidades.

Los esfuerzos continuos realizados por los países para garantizar la calidad y excelencia de la oferta universitaria son una respuesta a diversos cambios que ha habido a nivel nacional e internacional en el ámbito educativo (Haug, 2011):

- La búsqueda de una mayor eficiencia económica y social de las universidades con vistas a los problemas presupuestarios y de empleabilidad
- La diferenciación de una oferta formativa que responda a los desafíos del aprendizaje a lo largo de la vida, el crecimiento de la educación superior total o parcialmente virtual
- El crecimiento de la enseñanza privada en un buen número de países
- El aumento de la educación importada y la aparición de ‘fábricas de diplomas’
- El fuerte impacto de los *rankings* universitarios internacionales

Sin embargo, las políticas universitarias de calidad y la rendición de cuentas en las universidades se van adquiriendo conforme éstas van logrando una mayor autonomía en su actividad docente, investigadora y gestora. Es por eso que las recientes políticas de calidad y

rendición de cuentas se encuentran aún en vías de desarrollo y se van aplicando en cada país a velocidades diferentes, aunque siempre con referencia a las dos agendas europeas que configuran el EEES. Por una lado, la agenda del Proceso de Bolonia con la incorporación en 2005 de los Criterios y Directrices para la Garantía de la Calidad en el EEES y la creación en 2007 de agencias de acreditación y/o certificación de la calidad, que se han ido generando a través de la red “European Association for Quality Assurance in Higher Education” (ENQA). Y, por otro lado, la agenda del marco de la Unión Europea donde las universidades son consideradas motores del conocimiento y de la innovación (Haug, 2011).

En este contexto, Mora (2011) define unas nuevas estrategias o enfoques para la búsqueda de la calidad en las IES.

La primera estrategia se basa en la evaluación institucional para el logro de los objetivos perseguidos. De esta forma, mediante la evaluación se genera una movilización real por la calidad y se proporciona información a la sociedad a través de la publicidad de los informes finales.

La segunda estrategia en búsqueda de la calidad es la apertura a la demanda social. El cambio de una universidad de élites a una universidad de masas, la competitividad y la diversificación a nivel internacional y nacional hacen que los sistemas universitarios deban inclinarse más hacia la satisfacción de las demandas sociales. De esta forma, se pondrían de manifiesto las preferencias de los usuarios directos (estudiantes) e indirectos (empleadores) generándose un clima de competencia por los recursos y por los clientes estimulante para la calidad de las instituciones.

La tercera estrategia se basa en la utilización de los mecanismos de financiación de la educación superior como instrumentos que impulsen la eficacia del sistema universitario. Así, los incentivos financieros pueden influir en el comportamiento y las actitudes del profesorado y de la institución en general en la búsqueda de la calidad.

Finalmente, la cuarta estrategia para la mejora de la calidad y, según Mora (2011), la más importante es la mejora de la gobernanza universitaria que permita ir adelante con políticas de calidad.

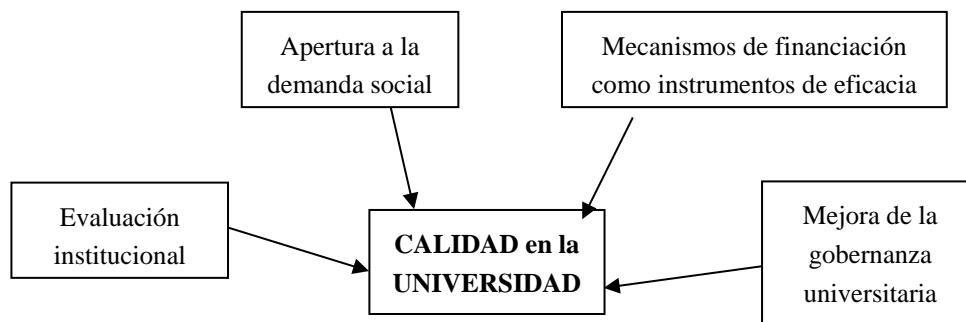


Figura 4.2. Nuevas estrategias para la mejora de la 'calidad universitaria'

En conclusión, las IES y, en particular, la universidad española deberían prestar atención a dos de los objetivos prioritarios y más relevantes en los últimos años a raíz de la crisis económica y financiera de Europa. Estos objetivos son: la mejora de la calidad de los procesos de enseñanza y aprendizaje, y la transparencia informativa sobre lo que las universidades ofertan. Este último objetivo es necesario puesto que las personas dedicadas a la gestión de las universidades, los políticos responsables del funcionamiento del sistema de educación superior y los académicos que analizan el sistema universitario son conscientes de la falta de una información rigurosa sobre los insumos, los procesos, los resultados y el impacto de las universidades (Mora, 2011).

4.2. La evaluación como estrategia de mejora de la calidad educativa

A pesar de los numerosos esfuerzos dedicados a la definición de una única interpretación para el concepto de 'calidad universitaria', la idea del término sigue siendo un tanto confusa y ambigua. Sin embargo, tal y como afirma George (1982):

“The most important point is that those persons involved must be reasonably clear about what is meant before assessment techniques and procedures are development. To do otherwise is almost to guarantee confusion, frustration and contention”.

Es por eso que aún sin conocer una definición exacta y común del concepto de calidad universitaria, es posible aproximarse al problema de la evaluación de la calidad de las universidades si se toman las debidas precauciones.

Desde finales del siglo XX y como consecuencia del interés por la calidad en la educación superior, en los países desarrollados se generó una situación denominada 'estado evaluativo' cuyo objetivo principal era insertar una forma particular de ética competitiva definida

externamente como la primera fuerza directriz del desarrollo de la educación superior (Neave, 1991). Desde entonces, la predisposición a la evaluación de la calidad universitaria se ha manifestado en numerosos países desarrollados.

Una universidad sólo puede alcanzar un nivel de calidad aceptable cuando los elementos humanos, financieros y físicos; la enseñanza y la investigación; y la organización y la dirección son en conjunto apropiados para los fines que la institución persigue. Por tanto, la evaluación de la calidad de una universidad sólo es posible plantearla de forma global evaluando cada una de sus piezas, pero dedicando una especial atención al funcionamiento en su conjunto. Este es el enfoque que adopta la evaluación institucional (Mora, 1998).

Según Mora (1998), la **evaluación institucional** se define como un intento de medir cómo se están cumpliendo las metas preestablecidas de una institución universitaria. Además, dicha evaluación exige haber definido previamente y con claridad cuáles son las metas que persigue la institución, lo que en algunos casos puede ser una primera dificultad para su puesta en práctica. Se podría afirmar, por tanto, que la calidad no se mide de forma exacta, pero se crean mecanismos y herramientas de aproximación a ella.

Cualquier proceso de evaluación supone identificar cuáles son los aspectos básicos que se consideran importantes para la consecución de los objetivos de la institución. Según Miller (1979), las cuestiones³ básicas que deben ser objeto de estudio en toda evaluación institucional son:

- Metas y objetivos de la institución
- Aprendizaje de los estudiantes
- Rendimiento del profesorado
- Programas académicos
- Servicios
- Gerencia
- Aspectos financieros
- Consejo de gobierno
- Relaciones externas
- Grado de preocupación por la mejora

Mora (1998) clasifica las cuestiones básicas a evaluar según las tres grandes dimensiones existentes en la organización de una universidad (enseñanza, investigación y servicios) y afirma que la evaluación se puede realizar o bien conjuntamente o bien por separado, pudiendo usar mecanismos y criterios de evaluación distintos en cada uno de estos tres aspectos:

³ Véase la referencia (Miller, 1979) para una descripción exhaustiva de los 45 elementos a evaluar dentro de cada uno de los 10 aspectos básicos enumerados en este trabajo

Enseñanza

- Aprendizaje de los estudiantes
- Rendimiento docente del profesorado
- Programas académicos
- Medios docentes disponibles (espacio, biblioteca, ordenadores, etc.)

Investigación

- Productividad investigadora
- Calidad de la investigación
- Dotaciones para la investigación

Servicios

- Gobierno de la institución
- Gerencia y administración
- Servicios disponibles para estudiantes y profesores

Respecto al tipo de evaluación institucional, en la literatura se pueden encontrar distintas clasificaciones según el agente evaluador, el enfoque, la obligatoriedad y el objeto de evaluación. En este trabajo se va a describir la clasificación más común de evaluación según el agente evaluador: evaluación interna y evaluación externa.

La **evaluación interna** es aquella que se lleva a cabo en la propia institución con el objetivo de mejorar su calidad. La comunidad universitaria, mediante un proceso de reflexión participativa, describe y valora su realidad permitiendo identificar los puntos fuertes y áreas de mejora para las universidades. Este tipo de evaluación se ha utilizado ampliamente en Estados Unidos ya que forma parte del proceso de acreditación del país (Mora, 1991a).

La **evaluación externa** se apoya en agencias u organismos especializados y ajenos a la universidad para realizar la evaluación. Este tipo de evaluación tiene la ventaja de contar con la imparcialidad de los expertos que evalúan la institución, pero la limitación, según observa Mora (1991a), de que los expertos no sean capaces de analizar desde fuera un organismo tan complejo como es la Universidad. Los fines propios de este segundo tipo de evaluación son (Cerdán et al., 1992):

- Obtener información contrastable y objetiva de la organización, funcionamiento y procesos didácticos que se dan en las universidades que permita complementar la evaluación interna
- Conocer la situación inicial de cada universidad en los ámbitos docentes y no docentes y poner así de manifiesto sus logros y dificultades, sus puntos fuertes y sus puntos débiles y sugerir orientaciones útiles que ayuden a tomar decisiones oportunas para resolver los problemas detectados

- Contribuir a que las universidades establezcan sus propios mecanismos de autoevaluación. Pueden proporcionar métodos e instrumentos, así como favorecer la reflexión y la cooperación
- Difundir la cultura de la evaluación entre todos los sectores de la comunidad educativa

La preocupación por la garantía de la calidad de las universidades españolas tiene ya una cierta historia. Son numerosas las iniciativas aplicadas que inculcan la ‘filosofía de la calidad’ de las actividades universitarias medidas a través de la evaluación.

En 1992 se puso en marcha el Programa Experimental de Evaluación de la Calidad del sistema universitario (Programa ESMU). En este plan se ensayó con la metodología típica de evaluación; una autoevaluación seguida de visitas externas por pares y con la publicación de informes finales complementados con la utilización de indicadores cuantitativos. Poco después algunos de los déficits del Programa Experimental fueron cubiertos en el Proyecto Piloto Europeo de la Evaluación de la Enseñanza (1994-95). La metodología aplicada para la evaluación fue prácticamente la misma que la del proyecto Experimental, si bien la evaluación se centró en la enseñanza y sólo incluía temas de investigación y gestión universitaria directamente relacionadas con la calidad de la docencia en las titulaciones evaluadas. Estos proyectos evidenciaron un alto grado de aceptación de la evaluación por parte de la comunidad universitaria. El punto crítico fue la motivación y dedicación de los miembros de los comités de evaluación, la falta de disponibilidad de datos requeridos y la definición imprecisa de los objetivos de la evaluación.

En 1995 se estableció el Plan Nacional de Evaluación de la Calidad de las universidades (PNECU) con el objetivo de promover la evaluación institucional de la calidad de las universidades, elaborar metodologías homogéneas para la evaluación de la calidad interna integradas en la Unión Europea y proporcionar información objetiva para la toma de decisiones. Tras cinco años de duración del Plan Nacional de Evaluación, en 2001 se estableció el II Plan Nacional de Evaluación de la Calidad de las Universidades (PCU) que incorporó la definición de un sistema, mediante el uso de indicadores, de información pública sobre las titulaciones; el establecimiento de un sistema de certificación de los servicios, siguiendo los mecanismos de certificación de la calidad que son usuales en las empresas (Normas ISO, Modelo EFQM,...) y de un sistema de acreditación de titulaciones.

Por su parte, uno de los objetivos prioritarios de la Ley Orgánica de Universidades (LOU) creada en 2001 y modificada por la Ley Orgánica 4/2007 LOMLOU (Ley Orgánica de Modificación de la ley Orgánica de Universidades) es la promoción y la garantía de la calidad universitaria mediante la evaluación, certificación y acreditación (véase artículo 31. Garantía de la Calidad). Para llevar a cabo la evaluación, certificación y acreditación se empezaron a crear agencias de calidad en las comunidades autónomas y en 2002 se fundó la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) como organismo responsable de llevar a cabo los procesos de evaluación y promoción de la calidad en las instituciones de enseñanza

superior. La misión de ANECA es aportar garantía externa de calidad al sistema universitario y contribuir a su mejora constante mediante el desarrollo de diferentes programas para llevar a cabo su actividad (evaluación, certificación y acreditación) (Agencia Nacional de Evaluación de la Calidad y Acreditación, 2012):

Programas de evaluación de enseñanzas e instituciones:

- Programa VERIFICA. Evalúa las propuestas de los planes de estudio diseñados en consonancia con los objetivos establecidos para la construcción del Espacio Europeo de Educación Superior
- Programa MONITOR. Efectúa el seguimiento de un programa verificado hasta que debe presentarse de nuevo para renovar su acreditación
- Programa AUDIT. Dirigido a los centros universitarios para orientarles en el establecimiento de sistemas de garantía interna de calidad
- Programa DOCENTIA. Da apoyo a las universidades para que diseñen mecanismos propios para valorar la calidad de la actividad docente de su profesorado
- MENCIÓN DE CALIDAD A PROGRAMAS DE DOCTORADO. Supone un reconocimiento a la solvencia científico-técnica y formadora de determinados programas de doctorado

Programas de evaluación del profesorado:

- Programa de evaluación del profesorado para la contratación PEP. Evalúa la actividad docente e investigadora y la formación académica de los solicitantes para el acceso a las figuras de profesor universitario contratado establecidas en la LOMLOU
- Programa de acreditación nacional ACADEMIA. Evalúa el perfil de los solicitantes para el acceso a los cuerpos de funcionarios docentes universitarios

4.3. Obtención de información en evaluación: el uso de indicadores

4.3.1. Contexto y orígenes de los indicadores en el ámbito de la educación superior

Con más o menos antecedentes tanto en Estados Unidos como en Europa, se reconoce que los indicadores de desempeño son fundamentales como un método estratégico para la evaluación de las IES. Así se manifiesta en las primeras líneas del informe titulado “Resultados de la educación superior: su identificación, medida y evaluación”, producto del seminario realizado en la Universidad de California, Berkeley, en 1970. En concreto, en este informe se expone la necesidad de justificar el presupuesto invertido en educación superior, así como de

demostrar el merecimiento de cada uno de los recursos adicionales concedidos. De forma literal los autores lo expresan como sigue:

“Our mandate is clear... We are going to have to prove that we deserve the dollars spent on higher education and justify our asking for each additional dollar” (Lawrence, Weathersby y Patterson, 1970).

De acuerdo con algunos analistas, a medida que la carga de los costes se desplaza cada vez más directamente al consumidor de los servicios de educación superior, el interés de éstos por demostrar el valor de la educación universitaria también crece. En este sentido, para Borden y Bottrill (1994), los indicadores de desempeño han surgido en gran parte como respuesta a estas presiones. Otro hecho que así lo denota es que organizaciones como la OECD, entre otras, han hecho de este concepto uno de los pilares de su trabajo en la gestión de la educación superior (Kells, 1993).

En el caso de los Estados Unidos, el registro de las primeras actividades de este tipo se remonta a las clasificaciones sobre reputación de las universidades a principios de 1910. Desde entonces se han producido innumerables listas y tablas de clasificación de las mejores universidades de acuerdo con una variedad de criterios. Entre las principales publicaciones destacan:

- *The American Council on education's ranking of doctoral programs*
- *The Carnegie Classification of all institutions*
- *The Gourman Report rankings of graduate and undergraduate programs*

Estos esfuerzos representan los intentos por establecer una amplia base de criterios para comparar las universidades y sus programas. Esta posibilidad de establecer semejanzas y diferencias o de posibles relaciones entre diversas IES que comparten una o varias características ha propiciado también una lucha por ‘ser’ o ‘estar’ entre los ‘mejores’ e intentar imitar los objetivos/resultados de las instituciones mejor posicionadas en las distintas tablas de clasificación.

El interés por la calidad de la educación superior tampoco es nuevo en el caso europeo, ya que desde sus orígenes las instituciones en este continente han tratado de evaluar sus funciones sustantivas. La universidad medieval contaba ya con sus propios modelos y enfoques hacia el control de la calidad demostrando desde entonces su capacidad de adaptación en entornos cambiantes (Kerr, 1982). En ese contexto, se puede hablar básicamente de dos modelos: el primero representado por las instituciones francesas y su práctica de conferir el control a una autoridad externa (Cobban, 1988) y el de las IES italianas como Bologna o Salerno. Ambos casos pueden ser considerados como el prototipo del control de calidad en términos de la rendición de cuentas del gremio de profesores hacia el Rector en el caso francés y de los órganos de control sobre los contenidos de la enseñanza en el caso italiano. El segundo es el

típico modelo inglés de comunidades académicas autónomas, expresión de lo que se conoce actualmente como aseguramiento de la calidad de revisión por pares. Los profesores deciden entre ellos lo que se debe enseñar y quién lo debe hacer.

A comienzos del siglo XIX, las hasta entonces instituciones dominantes de la enseñanza de educación superior comienzan su declive aferradas por largo tiempo a planes de estudio tradicionales (Francia) o debido al control de la iglesia sobre la enseñanza (Oxford y Cambridge) impidiéndoles modernizarse o adaptarse a los nuevos cambios.

Sobresale en esta fase de transición el dominio de las instituciones alemanas, siendo la Universidad de Berlín el ejemplo más destacado de la modernización de los gobiernos europeos, quienes gradualmente se convirtieron en los principales fuentes de financiamiento de las IES públicas definiendo además sus marcos legales y administrativos. Según Maassen (1997), el nacimiento de la universidad moderna orientada a la investigación puede ser visto como el principal impulso hacia el control del gobierno europeo en la gestión y financiamiento de sus instituciones educativas.

Actualmente es un hecho que las personas dedicadas a la gestión de las IES, los políticos responsables del funcionamiento de los sistemas universitarios y los académicos que analizan el sistema universitario son conscientes de la falta de información sobre los insumos, los procesos y, muy especialmente, sobre los resultados y el impacto de las IES. Esta situación, aunque está en constante cambio, ha sido una deficiencia generalizada que ha formado parte de las IES. En este nuevo contexto se revelan como elementos clave aspectos como la transparencia, la rendición de cuentas y la disponibilidad de la información (tanto para los gestores de las instituciones como para sus usuarios); transparencia para facilitar el cambio hacia un nuevo modelo, disponibilidad de información para ayudar a tomar decisiones de la mejor forma posible y rendición de cuentas para generar confianza entre sociedad y Universidad (Carot et al., 2012). La puesta en marcha de dichos aspectos clave conlleva a la definición de un sistema de indicadores que permita conocer la eficiencia, equidad y calidad del sistema educativo a través de información cuantitativa. Tal y como expone Schwartzman (1988), en este contexto emerge el ‘mercado’ de la calidad, ansioso de información cualitativa y cuantitativa que dé cuenta de los resultados de las IES y, además, que éstos puedan ser comparables y públicos.

4.3.2. Controversia sobre el uso de indicadores en la educación superior

A pesar de que actualmente los indicadores están presentes en la evaluación de las actividades de la mayoría de las universidades, en las últimas décadas se ha discutido abiertamente sobre su uso con el fin de orientar y justificar la toma de decisiones en la enseñanza de educación superior (Vidal, 1999). Hay quienes están absolutamente en contra del

uso de los indicadores como herramienta de medición de aspectos relacionados con la educación superior y, por el contrario, quienes están a favor de su uso para la toma de decisiones educativas. Pero lo más común es encontrar reflexiones en las que se hacen advertencias sobre el uso que se les debe dar y sobre las precauciones que se deben tener:

“Los indicadores de rendimiento son como los explosivos, en que pueden ser usados de forma destructiva o constructiva. Si fuesen productos químicos habría que calificarlos de inestables, no en el sentido que se cambien con cierta facilidad, sino en el sentido de que deben ser usados con extrema precaución, dado que una manipulación indebida puede provocar efectos no deseados” (Higher Education Quality Council, 1996).

Algunas de las razones de esta controversia son la discrepancia en la definición de unos objetivos claros sobre el uso que se le va a dar al indicador, la dificultad en la definición de indicadores diseñados para la posterior comparación entre instituciones o la preocupación de convertirlos en una meta por sí mismos cuando simplemente son una manifestación parcial de una realidad compleja (Vroeijerstijn, 1995). Además, suele ser muy tentador medir lo que es fácil de medir en lugar de lo que es relevante para la evaluación y esto acaba produciendo sesgos entre lo que se mide y lo que se pretende medir. Otro de los argumentos que se suele utilizar en contra del uso de indicadores y que es una de las razones por las que se ha avanzado tan poco en el desarrollo de información sobre las universidades es el deber de proteger la solidez de unas instituciones centenarias de las rápidas fluctuaciones del mundo exterior (Mora, 1999a).

A pesar de las limitaciones que presentan por encontrarse todavía en fase de perfeccionamiento e incluso de creación, lo cierto es que los indicadores facilitan la toma de decisiones, aportan información relevante acerca de fenómenos educativos y permiten analizar su evolución en el tiempo; siempre y cuando se definan claramente los objetivos de su uso y de su interpretación, y vayan acompañados de juicios y estimaciones cualitativas bien argumentadas y centradas en la singularidad de la institución, programa o persona evaluados.

Según Mora (1999a), el uso de indicadores como herramienta de medición puede ayudar a establecer un cierto orden y a hacer previsiones en un sistema tan complejo como es en la actualidad cualquier institución universitaria. No cabe duda de que cada vez es más común su uso en el ámbito de la educación superior, tanto para el funcionamiento a nivel interno de las universidades (modelos de evaluación de la calidad universitaria basados en el uso de indicadores: norma ISO, modelo EFQM,...; la evaluación del profesorado de la universidad: programa DOCENTIA; etc.) como para la realización de comparaciones entre instituciones o países (ARWU, 2003; QS, 2005; SCImago, 2007; Aguillo, Ortega y Fernández, 2008; Torres-Salinas et al., 2011; Instituto de Evaluación, 2011a; Instituto de Evaluación, 2011b) (véase sección 4.5).

Como consecuencia del incremento de su uso, se han aumentado los recursos para su elaboración y se ha mejorado la metodología empleada para su construcción (Vidal, 1999). Dada la creciente demanda parece haber acuerdos generales sobre ciertas propiedades y requisitos que deben cumplir los indicadores en su definición (Yorke, 1995; Carot et al., 2012). Así pues, son muchas las organizaciones dedicadas a la definición de indicadores de educación y a la recopilación de datos, como por ejemplo la OCDE, EUROSTAT, el Instituto de Estadística de la UNESCO (UIS), el Sistema Estadístico Europeo (SEE) que incluye el Instituto Nacional de Estadística (INE) de todos los estados miembros, el Instituto de Evaluación (IE) del Ministerio de Educación, Cultura y Deporte o la CRUE.

Una de las conclusiones a las que se ha llegado sobre el uso de indicadores es que por sí solos son una herramienta limitada, pero utilizados dentro de los actuales procesos de calidad en las universidades pueden convertirse en una herramienta auxiliar y valiosa para la mejora, además de servir como medio de información para las administraciones públicas y la sociedad sobre el funcionamiento interno de las IES (rendición de cuentas) (Mora, 1999a). Desde el mismo punto de vista, Olivera (1997) afirma que no se debe magnificar su utilidad ni condenarlos, sino utilizarlos como instrumentos que faciliten el análisis para la medición de complejos fenómenos educativos y su planificación.

4.4. El papel de los indicadores de educación superior en proyectos de investigación internacionales

A continuación se describen brevemente los objetivos y fundamentos de cuatro proyectos de investigación internacionales basados en la definición de indicadores para la medición de las distintas actividades de educación superior. Estos son: el proyecto “European indicators and ranking methodology for university third mission” (E3M), el proyecto “Sistema integral de información sobre las instituciones de educación superior de América Latina para el área común de educación superior con Europa” (INFOACES), el proyecto “The european classification of higher education institutions” (U-MAP) y, finalmente, el proyecto “Multi-dimensional global ranking of universities, a feasibility project” (U-MULTIRANK).

Existen muchos más proyectos o iniciativas en los que se definen y usan los indicadores como herramienta de medición, pero que no se van a ver con detalle en este trabajo. Ejemplos de éstos son: el proyecto Internacional de Indicadores de la OCDE (Proyecto INES-*International Indicator for Education Systems*), el proyecto MESALC (Mapa de la Educación Superior en América Latina y en el Caribe), el proyecto Internacional para la producción de indicadores de rendimiento de los alumnos de la OCDE (Proyecto PISA-*Programme for Indicators of Student Achievement*) y el proyecto PROFLEX (El PROfesional FLEXible en la sociedad del conocimiento).

De distintos modos, estos proyectos han tratado de avanzar en la definición de indicadores que contemplen y evalúen las diferentes actividades relevantes de las IES, teniendo en cuenta la diversidad de instituciones y entornos.

4.4.1. El proyecto “European indicators and ranking methodology for university third mission” (E3M)

“European indicators and ranking methodology for university third mission – E3M” es un proyecto financiado por la Comisión Europea en el marco del programa Lifelong Learning y con la participación de universidades de ocho países europeos, coordinado por el Centro de Gestión de la Calidad y del Cambio de la UPV. Es un proyecto que se desarrolló durante tres años, iniciado en enero de 2009 y finalizado en abril de 2012.

Tradicionalmente, la Universidad ha ido desarrollando de forma paralela dos actividades principales, la enseñanza (primera misión) y la investigación (segunda misión). En los últimos años, una llamada ‘tercera misión’ ha comenzado a tomar importancia bajo la premisa general de conectar la Universidad con las necesidades de la sociedad.

En la actualidad, existe una gran variedad de rankings que abarcan las dos primeras misiones de la Universidad y proporcionan gran cantidad de indicadores, pero la realidad es que no existe un conjunto de indicadores o una metodología de ranking que evalúe las actividades de esta tercera misión. El proyecto E3M se centra en cubrir esta necesidad. En este sentido, se llevó a cabo una reflexión sobre la naturaleza de la tercera misión de las universidades definiendo los procesos y las actividades que la integran e identificando los indicadores que mejor pueden representar el grado de desarrollo que una institución universitaria tiene en este tema.

El objetivo principal del proyecto es crear una herramienta útil que mida la actividad de la tercera misión y sirva a su vez para comparar las IES en Europa. En esta línea, se definen los objetivos específicos del proyecto:

- Crear un conjunto de indicadores de tercera misión, de ámbito europeo, para evaluar la efectividad de las instituciones en el cumplimiento de esta tercera misión
- Crear una metodología de ranking que permita a las instituciones realizar comparaciones, *benchmarking*, con sus pares en lo que se refiere a la tercera misión
- Definir un conjunto de buenas prácticas que sirva de base para el diálogo entre instituciones en el EEES

La primera etapa de la metodología de trabajo propuesta en el proyecto E3M fue el desarrollo de un marco conceptual y de trabajo para medir las actividades de tercera misión. De

esta forma, se definieron las tres principales dimensiones de la tercera misión: educación continua, compromiso social y transferencia de tecnología e innovación.

En segundo lugar, se identificaron los procesos de cada dimensión, entendiendo por proceso al conjunto de pasos que se ejecutan de forma secuencial con el fin de producir productos o servicios basados en los inputs específicos.

En la siguiente etapa del proyecto se diseñó un conjunto de indicadores con el objetivo de medir el resultado de los procesos definidos para cada una de las tres dimensiones. Se definieron varios tipos de indicadores: indicadores objetivos para medir actividades, indicadores de rendimiento e indicadores para medir la percepción subjetiva de los agentes implicados.

Posteriormente se seleccionó del conjunto de indicadores un subconjunto de indicadores ‘relevantes’ para medir de forma válida, precisa y universal las actividades de tercera misión. Se seleccionaron mediante una combinación de métodos estadísticos y cualitativos, se aplicó un análisis numérico para reducir el número de variables utilizando la información disponible y se empleó la metodología Delphi para llegar a un consenso sobre el conjunto de indicadores. El método Delphi es una técnica de consenso aplicada para llegar a un acuerdo, en base a la opinión de un conjunto de expertos, en este caso sobre el conjunto de indicadores relevantes para medir las actividades de tercera misión.

El último conjunto de indicadores definido es el conjunto de indicadores llamados ‘factibles’ que son un subconjunto de indicadores relevantes seleccionados de acuerdo a la accesibilidad y disponibilidad de la información. La batería de indicadores permite a las instituciones realizar ejercicios de benchmarking, centrados en la tercera misión, para conocer su propia situación y ubicarse a sí misma en relación con sus pares.

Finalmente, en la última etapa del proyecto se realizó un estudio de casos de buenas prácticas para cada una de las dimensiones. Se seleccionaron IES identificadas como excelentes en este ámbito a las que se visitó para realizar un estudio con el objetivo de conocer sus actividades de tercera misión en algunas de las tres dimensiones. Mediante esta etapa se validaron los resultados de las fases anteriores. Para un estudio más detallado de los resultados producidos en el proyecto E3M véase www.e3mproject.eu.

4.4.2. El proyecto “Sistema integral de información sobre las instituciones de educación superior de América Latina para el área común de educación con Europa” (INFOACES)

El proyecto INFOACES está cofinanciado por la Comisión Europea a través de su programa ALFA y está en actual desarrollo (2011-2013). Este proyecto se inició con el consorcio de 32

IES de 23 países, entre las que la UPV (Centro de Gestión de la Calidad y del Cambio) es coordinadora; 8 entidades y un grupo de expertos internacionales seleccionados en base a su experiencia en diversos campos de la educación superior.

INFOACES surge como una iniciativa de un grupo de instituciones con el objetivo general de crear un sistema integral de información sobre las IES de América Latina (AL) que permita el desarrollo institucional, la cooperación académica entre las instituciones participantes y que sirva de soporte al desarrollo del Área Común de Educación Superior (ACES) en sinergia con la Unión Europea. Asociado al objetivo general se han definido los siguientes objetivos específicos:

- Definir un sistema básico de indicadores para definir tipologías de las IES de AL utilizando criterios compatibles a los que se están desarrollando en la UE y en otras regiones
- Definir un sistema estratégico de indicadores para los distintos tipos de IES de AL. Los indicadores tendrán un enfoque holístico y recogerán información de las tres misiones de la Universidad (enseñanza, investigación y transferencia-extensión)
- Definir e implementar un sistema de obtención de información de resultados institucionales
- Proporcionar información a los distintos grupos de interés (estudiantes y sus familias, gestores universitarios, empleadores, autoridades públicas) sobre las IES, sus distintos tipos y sus características básicas
- Proporcionar a la comunidad internacional información sobre las IES de AL con criterios compatibles a los de otras regiones para facilitar el intercambio de estudiantes y la movilidad de graduados
- Servir de apoyo a la toma de decisiones institucionales en las IES favoreciendo las prácticas de benchmarking que estimulen la mejora de la calidad institucional
- Contribuir con estos indicadores al aseguramiento y la mejora de la calidad de la educación superior en AL cooperando con el sistema regional de evaluación de la calidad
- Dejar establecido, una vez finalizado el proyecto, un sistema de información sobre las IES de AL mediante la creación de una red estable asociada al sistema de información en la que estén representadas todas las IES que lo deseen, así como las asociaciones más importantes de IES nacionales y regionales de AL

La metodología propuesta en el proyecto INFOACES se basa en un modelo evaluativo global de tipo CIPP (siglas de Contexto, Insumo, Proceso, Producto). El modelo define la evaluación como el procedimiento orientado a obtener y proporcionar informaciones útiles para evaluar las posibles alternativas y, en consecuencia, resulta valioso para adoptar decisiones adecuadas en función de los objetivos considerados en la fase de diagnóstico y planificación, pasando por la puesta en marcha de los procesos y actuaciones hasta la fase de resultados finales.

Los elementos que componen el sistema integral de información son: el sistema básico de indicadores, el sistema de tipologías y el sistema estratégico de indicadores.

Sistema básico de indicadores

De acuerdo con el modelo evaluativo CIPP, este sistema está integrado por cuatro tipos de indicadores referidos a los cuatro tipos de aspectos: contexto, insumos, procesos y producto. El análisis de contexto trata de identificar las características del entorno en el cual se encuentra la institución de educación superior. El análisis de insumos pretende determinar la cantidad y el uso de los recursos necesarios para el logro de los objetivos operativos de la organización. El análisis de procesos genera información sobre la eficiencia y la eficacia de las diversas estrategias. Finalmente, el análisis del producto permite medir el grado de avance hacia los objetivos programados.

En esta primera fase se ha discutido abiertamente sobre cómo debería medirse, cómo son, qué hacen y qué resultados obtienen las IES. Además, el proyecto INFOACES apunta que la información resultante de las IES debe tratarse con datos fiables pero manejables, suficientemente precisos para que refleje lo que se pretende medir, pero cuya recolección sea factible y, sobre todo, útil para los usuarios que serán diversos y tendrán diferentes criterios y objetivos.

Es por eso que para abordar este primer objetivo específico, se ha definido el sistema básico de indicadores que refleja los resultados de las IES en sus distintas misiones (enseñanza, investigación y transferencia-extensión), teniendo en cuenta cómo son y dónde están. Estos indicadores han sido consensuados con mucho trabajo por casi un centenar de instituciones. Para un estudio más profundo del sistema de indicadores básico véase (Carot et al., 2012).



Figura 4.3. Marco conceptual teórico para la definición del sistema básico de indicadores

Sistema de tipologías

La diversidad de IES existente en AL hace necesario el establecimiento de un sistema de categorización de las instituciones en tipologías basado en perfiles institucionales que, a su vez, estén basados en las funciones reales y diversas que realiza cada institución.

Sistema estratégico de indicadores

Este sistema recoge los indicadores de resultados más relevantes que facilitan la comparación de calidad y el análisis de casos de buenas prácticas, ayudando a la definición de políticas basadas en el análisis, en base a evidencia, de la realidad de la educación superior en América Latina.

Estos tres elementos se integrarán en una plataforma de información/comunicación a la que pueda accederse.

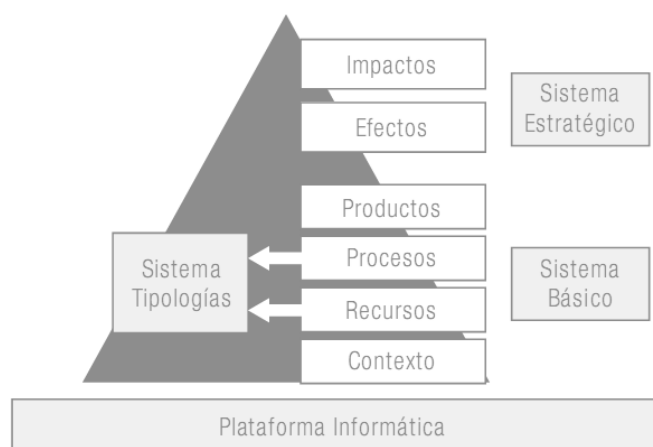


Figura 4.4. Fases de desarrollo del proyecto INFOACES

Se ha diseñado todo un sistema de indicadores, a través de un trabajo consensuado por un gran número de instituciones, asumiendo los siguientes principios:

- **Adecuación a los usuarios.** El sistema de información debe de ser capaz de ofrecer información adecuada, clara y accesible a cada uno de los colectivos a los que se dirige
- **Adaptación al marco conceptual.** Una definición clara del marco conceptual, con una delimitación precisa de las dimensiones y sub-dimensiones que se pretende medir y su alcance, permitirá tener mejores variables e indicadores
- **Enfoque holístico.** El modelo propuesto es un modelo de evaluación global (considera todas las actividades que desarrollan las instituciones)

- **Articulación de la información.** El sistema de información debe contar con mecanismos que aseguren que la información es obtenida, almacenada, manejada y mostrada de forma óptima
- **Mejora continua.** Se realizarán revisiones periódicas para incorporar modificaciones que generarán nuevas versiones mejoradas del sistema de indicadores
- **Escalabilidad.** El sistema se ha diseñado con un núcleo inicial de información que permita incorporar sucesivamente otros indicadores cuando se tenga capacidad de generar y recoger los datos necesarios
- **Información frente a ranking.** Es muy importante subrayar que un sistema de información no presupone, en absoluto, la elaboración de un ranking. El objetivo esencial de un sistema de información es establecer perfiles institucionales y elaborar indicadores que sean factibles para todas las instituciones participantes, obtenerlos en sus propias instituciones e introducirlos en un conjunto clasificado por tipologías de instituciones

Para un estudio más detallado de los resultados producidos en el proyecto INFOACES véase www.infoaces.org.

4.4.3. El proyecto “The european classification of higher education institutions” (U-MAP)

“The european classification of higher education institutions – U-Map” es un proyecto financiado por la Comisión Europea en el marco del programa Lifelong Learning que se desarrolló desde mediados de 2005 hasta principios de 2010 y fue coordinado por el “Centre for Higher Education Policy Studies (CHEPS)” de la Universidad de Twente. U-Map es una herramienta multi-dimensional de clasificación de distintos perfiles de IES, según las distintas misiones.

En la primera etapa del proyecto se identificaron seis grandes dimensiones para medir las distintas actividades de las IES y llevar a cabo la clasificación en distintos perfiles:

- Docencia y aprendizaje
- Perfil del estudiante
- Investigación
- Transferencia del conocimiento
- Orientación internacional
- Compromiso regional

En la siguiente etapa del proyecto se definió un conjunto de indicadores factibles para medir las distintas dimensiones identificadas. Un estudio detallado del conjunto de indicadores definidos se puede encontrar en (Van Vught et al., 2012).

La herramienta de clasificación definida proporciona dos resultados diferentes. Un primer resultado denotado en la plataforma on-line del proyecto como “*Profile finder*” presenta un conjunto de instituciones que responden a los criterios marcados por el usuario. Por otra parte, un segundo resultado denotado como “*Profile viewer*” proporciona una representación visual del perfil de una institución o de un conjunto de instituciones preseleccionadas.

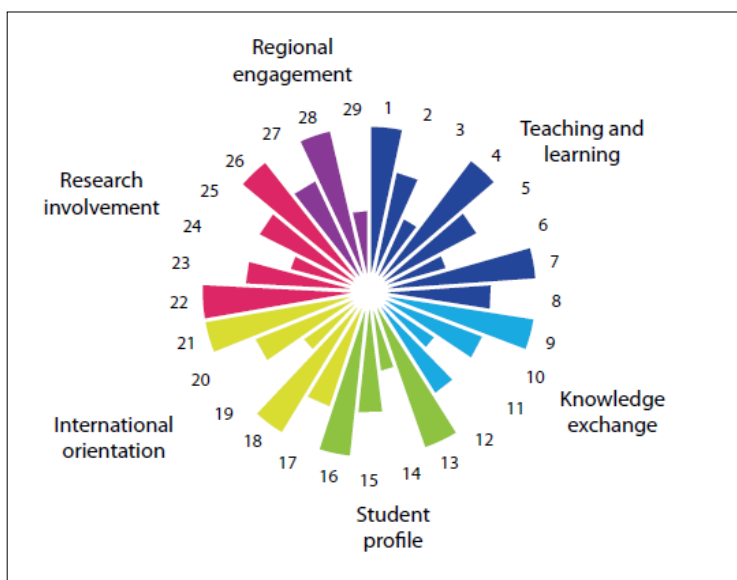


Figura 4.5. El “Profile viewer” para una institución

Los seis colores representan cada una de las dimensiones definidas y cada uno de los triángulos representa la posición de la institución con respecto al indicador evaluado. Las herramientas de clasificación U-Map se pueden encontrar en la página web del proyecto www.u-map.eu.

4.4.4. El proyecto “Multi-dimensional global ranking of universities, a feasibility project” (U-MULTIRANK)

“Multi-dimensional global ranking of universities, a feasibility project – U-Multirank” es un proyecto financiado por la Comisión Europea (DG de Educación y Cultura) y desarrollado por un consorcio de organizaciones de investigación bajo el nombre CHERPA durante los años 2009-2011.

El objetivo principal del proyecto es crear una nueva herramienta de transparencia internacional de las IES que sea multidimensional, multinivel y basada en el usuario. Se trata de una herramienta de medición que, según sus creadores, se diferencia sustancialmente de todos los rankings de universidades existentes y se dirige a las necesidades de las diversas partes interesadas en la educación superior.

Los criterios en los que se basa el proyecto para la elaboración de un nuevo concepto de ranking son (Van Vught y Ziegele, 2011) :

- **Multidimensional**, con el objetivo de cubrir las distintas misiones que llevan a cabo las IES
- **Transparencia**, con el fin de mostrar los resultados del ranking de una forma clara y entendible a los distintos usuarios y ofrecer la posibilidad de realizar una ordenación de ranking de acuerdo a sus necesidades
- **Global**, con el objetivo de cubrir instituciones dentro y fuera de Europa (en particular instituciones de Asia, Australia y EE.UU.)

El ranking U-Multirank incluye un conjunto de indicadores que permite a los usuarios comparar el rendimiento de las distintas IES mediante cinco dimensiones diferentes:

- Docencia y aprendizaje
- Investigación
- Transferencia del conocimiento
- Orientación internacional
- Compromiso regional

En este contexto, el proyecto U-Multirank produce dos tipos de rankings diferentes:

- **Ranking focalizado en las instituciones.** Este tipo de ranking permite realizar comparaciones entre instituciones para cada una de las dimensiones anteriormente enumeradas. Las puntuaciones en cada una de las cinco dimensiones no se agregan en un único índice
- **Ranking basado en los distintos campos de estudio.** En este caso, se proporciona un ranking multidimensional que permite realizar comparaciones de un conjunto de programas de estudio de instituciones con el mismo perfil

El proyecto U-Multirank ha incorporado los resultados de clasificación de instituciones del proyecto U-Map para realizar las comparaciones entre instituciones con el mismo perfil. Para un estudio más exhaustivo del conjunto de indicadores que maneja el proyecto U-Multirank véase (Van Vught y Ziegele, 2011).

Para un estudio más detallado de los resultados producidos en el proyecto U-MULTIRANK véase www.u-multirank.eu.

El objetivo común de estos cuatro proyectos es la elaboración de un sistema de indicadores que refleje las diversas actividades relevantes de las IES con el fin de establecer comparaciones entre las diferentes instituciones con un perfil semejante (teniendo en cuenta la estructura de las instituciones y sus principales misiones, la estructura económica y educativa del entorno, etc.).

En este sentido, no se pretende realizar tablas de clasificaciones internacionales en las que se le asigna una puntuación final a cada institución, realizando posteriormente una ordenación, sino favorecer el acceso a numerosos criterios y, de esta manera, satisfacer las necesidades de los diferentes perfiles de usuarios.

4.5. Un panorama de los rankings universitarios

En los últimos años se ha producido una rápida difusión de los rankings de universidades. El nuevo entorno global en el que se mueven, el proceso de internacionalización de la educación superior o la aparición de la llamada sociedad del conocimiento son algunas de las causas que han generado la necesidad de establecer comparaciones entre IES (Carot, Bas y García-Gutiérrez, 2009).

Los rankings o las tablas de clasificación tratan de ser una herramienta de ayuda para la toma de decisiones de los distintos grupos de interés, como los estudiantes y sus familias, investigadores, profesores, empresas o las propias universidades. La cuestión fundamental que abordan es la determinación de cuáles son las ‘mejores’ universidades. Pero, lógicamente, la respuesta puede ser muy distinta si lo que se trata es que un estudiante decida dónde estudiar, un investigador decida dónde desarrollar su actividad o una empresa decida qué universidad podría ser un buen lugar donde invertir en I+D; por citar algunos ejemplos. Es decir, una universidad puede ser muy buena en todos los aspectos o, quizá, sólo en alguno de ellos aunque en ese sea mucho mejor que otras. Por esta razón, los rankings globales que incluyen entre sus indicadores distintas dimensiones como docencia, investigación o transferencia son frecuentemente cuestionados.

Pero a pesar de todas sus limitaciones y de las críticas que han suscitado, es un hecho que los rankings tienen cada vez mayor impacto. La publicación de los resultados de algunos de los rankings más importantes como el ARWU o el THES son un acontecimiento mediático de relevancia que inmediatamente genera un amplio debate. En los medios de comunicación se pueden encontrar noticias que se limitan a informar sobre los resultados de los rankings sin agregar valoración alguna, otras noticias los critican o ensalzan, pero la mayoría de ellas se usan para cuestionar políticas en base a sus resultados y para destacar la posición de alguna universidad.

Sin embargo, la interpretación que hacen los medios de comunicación y las propias universidades de los resultados de los rankings, en ocasiones dista mucho del objetivo por el que fueron construidos. Muchas veces son usados con otros objetivos o por un tipo de usuarios distinto a sus destinatarios iniciales. En este caso, la utilización de los rankings supone un

enorme riesgo, ya que, cada vez más, la información que generan es considerada como referente en el diseño de políticas y en la toma de decisiones.

Por otra parte, los rankings crean opinión sobre cuáles son las ‘mejores’ universidades, promoviendo el establecimiento de modelos a seguir y generando una cierta tendencia a la homogeneización en torno a esos modelos de universidad. A pesar de ello, no resulta adecuado comparar con los mismos indicadores a universidades con finalidades distintas.

Los rankings y el uso que de ellos se hace han recibido numerosas críticas a pesar de su popularidad. Estas críticas se podrían resumir globalmente en tres tipos: las que hacen referencia a los indicadores que contienen, las asociadas a la metodología usada para su construcción y las que cuestionan el propio concepto de ranking. Un ranking, en definitiva, no es más que una ordenación de acuerdo a algún criterio. Pero ¿qué miden realmente los rankings?, ¿cuáles son los criterios para generar esta ordenación?, ¿reflejan realmente la calidad de las universidades?

Todos los rankings se construyen en base a una colección más o menos pequeña de indicadores con los que supuestamente puede recogerse el nivel de calidad de una universidad. Estos indicadores pueden estar referidos a alguna de las misiones de la Universidad: docencia, investigación o la llamada tercera misión; pero también a otras cuestiones como tamaño, entorno, etc. Así pues, los resultados obtenidos en el ranking estarán directamente relacionados con las dimensiones medidas. Algunos rankings pretenden medir y generar ordenaciones sólo sobre una de estas dimensiones. Esto ocurre en mayor grado con la investigación, cuyos resultados son más fácilmente medibles. Otros, sin embargo, dicen medir calidad global aunque en realidad en sus indicadores se ven más reflejadas unas dimensiones que otras. Normalmente, la dimensión que genera un mayor sesgo en los rankings es la de investigación.

La dificultad que entraña una correcta identificación y selección de los indicadores sobre los que construir un ranking ha sido recogida por muchos autores y, probablemente, constituye la mayor limitación de los rankings. Además del tipo de actividad que recoja, cada indicador puede ser: de recursos, de proceso, de resultados o de percepción. Generalmente, los indicadores de recursos y proceso son más fácilmente medibles y están disponibles con más frecuencia y, quizá, por ello son más usados aunque no está claro que reflejen directamente la calidad de una institución. En cuanto a los indicadores de resultado, son mucho más difíciles de obtener. En investigación se han desarrollado medidas bibliométricas que, a pesar de sus imprecisiones, son razonablemente válidas y fiables. Sin embargo, es más difícil obtener indicadores reales de resultados de la dimensión docente que no sean puramente de proceso. Y es todavía mayor la dificultad en las actividades de la tercera misión. Por último, en cuanto a los de percepción, que introducen una componente de subjetividad, deben ser obtenidos con encuestas sólidas y muestras suficientemente grandes y bien diseñadas.

Otro problema que presentan los indicadores es la selección de la fuente de información y la calidad de los datos en cuanto a precisión y exactitud. Una mala calidad en los datos limita la

elaboración de indicadores fiables. Es por eso que la metodología utilizada para la construcción de los indicadores es muy importante y debe ser lo más transparente posible.

La metodología de los rankings más conocidos se basa en obtener un valor numérico final asignando un peso a cada uno de los indicadores incluidos. Además de que el procedimiento es en sí mismo cuestionable, lo que suele recibir más críticas es la asignación de los pesos. En ocasiones, estos pesos se obtienen por medio de métodos estadísticos, pero en la mayoría de los casos se asignan con un cierto grado de arbitrariedad. Como consecuencia, la clasificación final puede variar mucho dependiendo de los pesos de los indicadores. Algunos autores han mostrado como la metodología, en particular el esquema de pesos utilizados, provoca más cambios en los resultados de la ordenación final que la inclusión o eliminación de algunos indicadores. Es por eso que se debe justificar cualquier elección metodológica elegida y aplicar análisis de robustez para evaluar la calidad del ranking.

Con la puntuación numérica obtenida se construye una escala ordinal que coloca a cada universidad en un orden que busca reflejar la ‘calidad’ de la institución. De esta forma la calidad (o lo que el ranking entienda por calidad) se convierte en un atributo unidimensional que además se expresa en forma de orden jerárquico. Pero, ¿es la calidad de una universidad un concepto unidimensional o multidimensional?, ¿tiene sentido reducir todos los indicadores medidos a uno sólo? Las críticas dirigidas a la idea de ranking en sí misma se basan en este razonamiento, cuestionando el resultado global del ranking, aunque a veces se acepte de buen grado la consideración del conjunto de indicadores pero con un uso distinto, por ejemplo el benchmarking.

El número de rankings que se pueden encontrar en la actualidad es considerable. Aunque se han usado varias formas de clasificarlos, de modo general se suelen agrupar en rankings nacionales e internacionales.

	Rankings nacionales	Rankings internacionales
Estados Unidos	<ul style="list-style-type: none"> - US News & World Report (USNWR) - Business Week - Gourman - The Wall Street Journal - The Center (investigación) - AUTM (transferencia de tecnología e innovación) - Journal of Blacks in Higher Education (compromiso social) - New Mobility Magazine (compromiso social) - NewsWeek International 	<ul style="list-style-type: none"> - The Wall Street Journal
Reino Unido	<ul style="list-style-type: none"> - The Times - The Sunday Times - The Guardian University Guide - NUBS (transferencia de tecnología e innovación) - EUGEN's Top Universities -homosexualidad y no discriminación 	<ul style="list-style-type: none"> - The Guardian - The Philosophical Gourmet - The Economist - The Financial Times - Times Higher Education Supplement (THES - QS World University Ranking)
Australia	<ul style="list-style-type: none"> - Melbourne Institute - The Good University Guide - Hobsons 	
Canadá	<ul style="list-style-type: none"> - The Maclean's Guide to Canadian Universities 	
Alemania	<ul style="list-style-type: none"> - CHE-Die Zeit - B-School - Center of Excellence Women and Science - Ranking Humbolt 	<ul style="list-style-type: none"> - CH Excellence Ranking
Francia	<ul style="list-style-type: none"> - Le Nouvel Observateur - Le Point 	<ul style="list-style-type: none"> - Professional Ranking (ParisTech-Mines)
Italia	<ul style="list-style-type: none"> - Il Score 24 ORE 	
España	<ul style="list-style-type: none"> - Ranking universidades españolas (Gaceta Universitaria) - Periódico El Mundo (50 carreras) - InternetLab - indexación artículos científicos - COTEC (investigación) - Ranking de productividad en investigación de las universidades públicas españolas (Universidad de Granada) - Índice IAIF de calidad universitaria (investigación y docencia) 	<ul style="list-style-type: none"> - Scimago (RI3) - Webometrics Rankings of World Universities (CIndoc,CSIC) - 4iCU- 4 International colleges and universities- (popularidad en la Web)
Asia-Pacífico		<ul style="list-style-type: none"> - Asia Week
Latino América		<ul style="list-style-type: none"> - Qué pasa - America Economía
China		<ul style="list-style-type: none"> - University Jiao Tong of Shanghai - ARWU
Taiwan		<ul style="list-style-type: none"> - The Performance Ranking of Scientific Papers for World Universities
Holanda		<ul style="list-style-type: none"> - Ranking of Leiden
Bélgica		<ul style="list-style-type: none"> - University System Ranking (Lisbon Council)
Polonia		<ul style="list-style-type: none"> - Perspektywy – global

Tabla 4.1. Rankings nacionales e internacionales según el país donde se elaboran

A modo de ejemplo ilustrativo de las bondades y limitaciones de los rankings, a continuación se describen cuatro de los que actualmente tienen más impacto (Rauhvargers, 2011).

Academic Ranking of World Universities - ARWU

Este ranking lo elabora la “Shanghai Jiao Tong University” desde 2003. Es un ranking que ordena 1.000 universidades de todo el mundo aunque sólo publica los resultados de las 500 primeras. Ofrece tres tipos de clasificaciones: a nivel global, según campos de estudio y según materias. Los indicadores y pesos que usa son:

- Alumnos de una institución que han conseguido un Premio Nobel o han recibido galardones oficiales [10%]
- Profesores de una institución que han conseguido un Premio Nobel u otros premios oficiales [20%]
- Investigadores más citados en 21 categorías por el Thomson ISI (ciencias, medicina, física, ingeniería y ciencias sociales) [20%]
- Artículos publicados en Nature y Science [20%]
- Rendimiento académico en relación al tamaño de la institución (número total de docentes a tiempo completo) [10%]

El ARWU es, quizá, el más citado de todos los rankings que se elaboran y la publicación anual de sus resultados produce siempre un considerable impacto mediático. Pero, debido a su forma de cálculo, la ordenación que genera el ARWU está claramente sesgada hacia los resultados de investigación y favorece a las universidades grandes. Aunque este ranking se diseñó para comparar universidades atendiendo a su rendimiento investigador, se usa cada día más como un indicador de la calidad global de la institución. Otro problema que presenta este ranking es que no evalúa muchas áreas de conocimiento, perjudicando en este caso a las humanidades y ciencias sociales.

Ranking THE

El diario “The Times” publica anualmente desde 2004 en su suplemento educativo un ranking mundial conocido como THES. Este ranking pretende reflejar la parte docente y la investigadora usando como indicadores y pesos:

- Opinión de expertos académicos sobre la lista de las 30 universidades líderes a nivel mundial (divididas en cinco áreas) [40%]
- Opiniones de empleadores [10%]
- Ratio profesor/alumno [20%]
- Citas de investigación por profesor [20%]
- Proporción de estudiantes extranjeros [5%]
- Proporción de profesores extranjeros [5%]

Observando estos indicadores, pueden hacerse varias críticas a los resultados que se obtienen en el ranking. En primer lugar, la reputación que una universidad tiene a nivel internacional es

una fuerte influencia en varios de los indicadores, pero no siempre está asegurado que esos niveles de reputación correspondan exactamente con los niveles de calidad. Más aún cuando esto es evaluado por expertos, el número y la selección de los cuales puede también sesgar los resultados en función, por ejemplo, de la composición por países, áreas o tipo de universidades de donde procedan.

Ranking de Leiden

Este ranking es elaborado por el “Centre for Science and Technology Studies” de la Universidad de Leiden y está alcanzando un considerable prestigio en Europa. Se construye en base a un conjunto de indicadores bibliométricos desarrollado por el propio centro que llevan a identificar cuatro indicadores generales calculados para todas las universidades europeas:

- Amarillo (ranking por tamaño): mayor número de publicaciones
- Verde (ranking por tamaño-independiente): número de citas por publicación dividido por el impacto medio en el campo
- Naranja (ranking por tamaño-dependiente, fuerza bruta): número total de publicaciones multiplicadas por el impacto relativo en el campo (impacto medio normalizado)
- Azul (ranking por citas): número de citas por publicación, sin autocitas

De nuevo, se trata de un ranking centrado en resultados de investigación y medido, exclusivamente, a través de artículos científicos. De esta forma, salen beneficiadas las universidades grandes y aquellas enfocadas hacia la investigación.

CHE-Ranking

Es, quizá, el ranking europeo que goza de mejor consideración elaborado por el “Center for Higher Education Development” desde 1998. El ranking usa, para cada universidad, un buen número de indicadores desagregados para cada rama de estudios. Los resultados no pueden ser agregados, por lo que no es posible calcular un ranking absoluto. Debido a su carácter multidimensional, las fuentes de información de donde se obtienen sus indicadores son diversas:

- Universidades
- Departamentos
- Control de los profesores
- Control de los estudiantes
- Análisis bibliométrico
- Análisis de patentes
- Graduados

Entre sus indicadores contiene algunos de opinión, lo que le ha valido algunas críticas. Por otra parte, contiene también indicadores que se usan únicamente a nivel informativo. En

concreto, este ranking se basa en 34 indicadores distintos por cada universidad, resumidos a su vez en 9 módulos por temas. El prestigio que se ha ganado este ranking se debe, por una parte, a la forma que tiene de ordenar (sin un ranking global y por tramos) y, por otra, al conjunto de indicadores de carácter multidimensional.

Los rankings de universidades están tomando una posición relevante en la comparación de universidades, provocando cada vez mayor competitividad y contribuyendo a la creación de la imagen que las universidades proyectan. Con sus efectos positivos y negativos, son una realidad que no se puede obviar y que hace necesaria una evaluación crítica y prudente de sus resultados.

En el nuevo contexto mundial de creciente movilidad, internacionalización, globalización y competitividad, los rankings pueden desempeñar un papel importante de ayuda a la toma de decisiones. Pero, para ello, deben diseñarse con un objetivo claro, la selección de los indicadores debe ser la adecuada para lograr esos objetivos, los datos recopilados deben tener la calidad requerida, la metodología debe basarse en hechos contrastados y, finalmente, debe usarse por los destinatarios para los que pensó.

Por otra parte, los rankings han potenciado la preocupación por la evaluación de la calidad y la rendición de cuentas de las IES, lo cual puede redundar directa o indirectamente en una mejora de su calidad. A raíz de la proliferación de los rankings, se han impulsado iniciativas para definir más y mejores indicadores, para mejorar los sistemas fiables de adquisición de información y para incluir entre los sistemas de medición y evaluación aspectos no recogidos hasta el momento o recogidos de forma parcial (actividades de tercera misión, por ejemplo).

En definitiva, de lo que se trata es de usarlos como una herramienta con su campo de aplicación y sus limitaciones, y de mejorarlos y construir nuevos rankings que cumplan los dos requisitos de calidad básicos exigibles: validez (que midan aquello que pretenden medir) y fiabilidad (que lo hagan con un error pequeño).

Capítulo 5

Evaluación de la actividad docente y académica del profesorado universitario

5.1. Introducción

La universidad española se plantea alcanzar muchos retos relevantes con el fin de conseguir una mayor calidad en el sistema universitario español y competitividad en relación con el resto de universidades europeas. Este trabajo de investigación se centra en uno de los retos en particular, la evaluación del profesorado universitario. La transformación surgida como consecuencia de la implantación de las directrices del Espacio Europeo de Educación Superior (EEES) tiene una influencia directa en la actividad académica del profesorado, observándose la importancia cada vez mayor de las actividades de evaluación que aseguren la calidad de los procesos formativos que afecten a los estudiantes, a la comunidad universitaria y a la sociedad en general (European Commission Education and Training, 2009). Los resultados de la evaluación permitirán a las universidades el diseño y la ejecución de planes de mejora de las actividades académicas del profesorado. Por tanto, es importante disponer de herramientas de medición que permitan evaluar la actividad del profesorado universitario en sus tres dimensiones: docencia, investigación y gestión.

Sin embargo, la falta de una definición consensuada para el concepto de ‘buen profesor’ no ha permitido desarrollar una metodología de evaluación de la calidad docente común para todas las IES. En palabras de Altbach (2006):

“There are in fact, no widely accepted methods for measuring teaching quality, and assessing the impact of education on students is so far an unexplored area as well”.

Algunas de las metodologías aplicadas en distintas universidades a nivel europeo para la evaluación del profesorado son: el uso de cuestionarios a estudiantes, la evaluación por pares, los autoinformes, el uso de indicadores para la mejora de la calidad docente, etc. Para una revisión exhaustiva de cada uno de los métodos de evaluación anteriores véase (Henard, 2010).

A continuación se exponen dos modelos de evaluación, mediante el uso de indicadores compuestos, propuestos por la Universitat Politècnica de València para la medición de la actividad docente y académica del profesorado universitario y que van a servir de base para este trabajo de investigación.

5.2. Modelo de evaluación de la actividad docente del profesorado aplicado al contexto académico universitario español⁴

En el año 2006 nace el Programa DOCENTIA, promovido por la ANECA y en colaboración con las Agencias Autonómicas con el objetivo prioritario de proporcionar un marco de referencia, un modelo y unos procedimientos para abordar la evaluación de la actividad docente que se desarrolla en las universidades. Este programa se ha diseñado teniendo en cuenta el papel cada vez más relevante que las universidades han de jugar en la evaluación de su profesorado. Por ello, DOCENTIA persigue favorecer la mejora de la enseñanza a partir del desarrollo de procesos de evaluación de la actividad docente realizados por las propias universidades y ajustados, por tanto, a sus necesidades en el convencimiento de que las decisiones que las universidades adopten tras la evaluación (formación del profesorado, incentivos económicos, reconocimiento, promoción, etc.) contribuyan a mejorar la calidad de la actividad docente.

Desde ANECA se han establecido ya dos convocatorias en las que se ha podido apreciar la gran acogida que ha tenido el programa, tanto en las universidades públicas como en las privadas, con una elevada participación que supone el 85% del total de universidades existentes en el sistema universitario español (Pozo, 2010).

Sin embargo, y aunque todos los diseños han tenido como marco de referencia las pautas establecidas por ANECA en el Programa DOCENTIA, la diversidad de procedimientos e instrumentos de medida diseñados; la heterogeneidad existente en la definición de indicadores; la ausencia, en muchas ocasiones, de documentos marco de política de profesorado; la falta de concreción de los criterios de evaluación y, por señalar algún ejemplo más, el desconocimiento actual sobre la capacidad discriminante de la mayoría de los manuales de evaluación, requieren de un análisis exhaustivo y pormenorizado de los diseños existentes hasta la fecha y de los aspectos más significativos que han surgido durante el proceso de implantación.

En cualquier caso, sea cual sea la adaptación efectuada a cada contexto universitario, el programa DOCENTIA tiene entre sus objetivos prioritarios apoyar individualmente al

⁴ Esta sección está basada en el informe del programa DOCENTIA diseñado por ANECA (2006) y en el Manual de Evaluación de la Actividad Docente del Profesorado diseñado por la UPV (2010)

profesorado, proporcionándole evidencias que puedan serle de utilidad para su uso en el proceso de acreditación, según lo establecido en el Real Decreto 1312/2007 de 5 de octubre.

Las instituciones universitarias españolas han acogido el modelo DOCENTIA con el compromiso de elaborar su propio sistema de evaluación, sin olvidar las directrices propias del programa. En la tabla 5.1 se recoge, por comunidades autónomas, aquellas universidades españolas que utilizan el programa DOCENTIA para evaluar la labor docente de su profesorado universitario (Pozo, 2010). Tras una revisión exhaustiva de los procedimientos adaptados por las distintas instituciones, en general, se observa uniformidad en la filosofía del proceso adoptado para la evaluación de la calidad docente.

Comunidad Autónoma	Universidades públicas españolas que evalúan a través del programa DOCENTIA	Universidades públicas españolas con un método de evaluación alternativo
ANDALUCÍA	Almería Cádiz Córdoba Granada Huelva UNIA Jaén Málaga Pablo de Olavide	Sevilla
ARAGÓN		Zaragoza
ASTURIAS	Oviedo	
BALEARES	Islas Baleares	
CANARIAS	La Laguna Las Palmas	
CANTABRIA	Cantabria	Internacional Menéndez Pelayo
CASTILLA LA MANCHA		Castilla la Mancha
CASTILLA Y LEÓN	Burgos Salamanca Valladolid León	
CATALUÑA	Autónoma de Barcelona Barcelona Girona Lleida Politécnica de Cataluña Pompeu Fabra Rovira i Virgili	Abat Oliba CEU
EXTREMADURA	Extremadura	
GALICIA	La Coruña Santiago de Compostela Vigo	
LA RIOJA	La Rioja	
MADRID	Alcalá de Henares Autónoma de Madrid Carlos III Complutense de Madrid Politécnica de Madrid UNED Rey Juan Carlos	
MURCIA	Murcia Politécnica de Cartagena	
NAVARRA	Pública de Navarra	
C. VALENCIANA	Alicante Jaume I Miguel Hernández Politécnica de Valencia Valencia	
PAÍS VASCO	País Vasco	

Tabla 5.1. Evaluación de la labor docente en las universidades públicas españolas

5.2.1. Modelo de evaluación de la actividad docente del profesorado de la Universitat Politècnica de València

IMPLANTACIÓN DEL PROGRAMA DOCENTIA EN LA UPV. POLÍTICA INSTITUCIONAL DE LA UPV EN LA EVALUACIÓN DEL PROFESORADO

La Universitat Politècnica de València ha venido desarrollando a lo largo de muchos años la evaluación de la actividad realizada por su personal docente e investigador, como fruto de su espíritu crítico y de su vocación de innovación y mejora continua. La realización del proceso de encuesta a los estudiantes sobre la actividad docente fue aprobada por la Junta de Gobierno de la UPV en 1994 con carácter definitivo, tras un periodo de diseño, experimentación y análisis previo.

La finalidad de la evaluación de la actividad docente va más allá de la asignación de complementos retributivos específicos y se entronca en una voluntad de impulso de una cultura organizativa fundamentada principalmente en la dedicación y en los méritos de sus miembros. La evaluación de la actividad docente se entiende como una evaluación interna que la universidad realiza de su profesorado para garantizar el cumplimiento de los objetivos de las enseñanzas que imparte.

Así, la UPV ha desarrollado sistemas, ampliamente aceptados en la comunidad, con criterios objetivos. Un ejemplo muy destacado es el sistema de evaluación de la actividad investigadora del profesorado mediante indicadores (Índice de Actividad Investigadora - IAI) y valoraciones personalizadas (Valoración de la Actividad Investigadora Personalizada - VAIP), que han servido de base para priorizar actuaciones, asignar recursos, etc. actuando, por consiguiente, como motivación para que el profesorado incremente la calidad y cantidad de su actividad y producción.

Si se quisiera enumerar resumidamente las características fundamentales de la política de profesorado de la UPV habría que señalar primeramente la intención de ser una política con voluntad globalizadora, en el sentido de que intenta articular sistemas que permitan incorporar todas las vertientes de la actividad académica del PDI. Seguidamente, hay que destacar el objetivo de ser una política integradora que intenta desarrollar de forma articulada los procesos de selección, de promoción, evaluación y de reconocimiento relacionados con el PDI.

La política institucional de la UPV se ha plasmado recientemente con extremo detalle en el Plan Estratégico UPV 2007/2014 donde se han analizado pormenorizadamente las circunstancias y aspectos fundamentales para el devenir de la institución. La Comisión del Plan Estratégico de la UPV aprobó la declaración de la Misión de la misma, así como los elementos constitutivos de la Visión de la misma. Dicha declaración estableció que es Misión de la UPV (Universitat Politècnica de València, 2007):

- La formación integral de los estudiantes a través de la creación, desarrollo, transmisión y crítica de la ciencia, de la técnica del arte y de la cultura, desde el respeto a los principios éticos con una decidida orientación a la consecución de un empleo acorde con su nivel de estudios
- La contribución al desarrollo cultural, social y económico de la sociedad valenciana y española mediante el apoyo científico, técnico y artístico
- El desarrollo de un modelo de institución caracterizada por los valores de excelencia, internacionalización, solidaridad y eficacia; una institución abierta que incentiva la participación de instituciones, empresas y profesionales en todos los aspectos de la vida universitaria.

Igualmente se establecieron los elementos constitutivos de la Visión, es decir el propósito con el que se concretará en el futuro inmediato el cumplimiento de la Misión, caracterizado en ocho puntos. Así, el objetivo de la UPV es llegar a ser:

- Una universidad capaz de atraer y formar buenos estudiantes y convertirse en su opción preferible estando a la cabeza de la calidad docente y siendo exigente con la aplicación de los estándares del Espacio Europeo de Educación Superior
- Un centro de excelencia reconocido internacionalmente que propicie el intercambio y la movilidad de alumnos, profesores e investigadores
- Una universidad con investigación consolidada de calidad con una política definida de excelencia
- Una universidad innovadora y emprendedora, líder a nivel mundial
- Una universidad de fuerte implicación social y que sirva de puente entre la Comunidad Valenciana y cualquier otra región del mundo
- Una universidad con una programación de postgrado y formación continua que esté vinculada a las necesidades y demandas de la sociedad y que sea generadora de nuevas oportunidades profesionales y personales
- Una universidad con un capital humano altamente cualificado, fuertemente motivado e ilusionado con el desempeño de su trabajo
- Una universidad que gestiona sus recursos y servicios con eficacia y eficiencia

Para alcanzar la Visión formulada, la UPV ha decidido orientar su actividad a la consecución de 16 objetivos estratégicos, articulados en 5 ejes:

1. FORMACIÓN Y APRENDIZAJE

- **Promover la excelencia docente**
- Atraer a los mejores docentes
- Mejorar el rendimiento académico de los estudiantes
- Conseguir la inserción laboral de calidad de los egresados

2. INVESTIGACIÓN, DESARROLLO TECNOLÓGICO E INNOVACIÓN

- Ampliar la actividad y potenciar la excelencia investigadora
- Incrementar y mejorar la calidad de las actividades de transferencia tecnológica e innovación

3. COMPROMISO SOCIAL Y VALORES

- Orientar la UPV a la obtención de los resultados que mejoren su utilidad social a corto, medio y largo plazo
- Fortalecer el compromiso social de la UPV
- Potenciar los valores y la formación integral de las personas

4. PERSONAS

- Propiciar un capital humano comprometido con la excelencia docente, investigadora y de gestión
- Garantizar las mejores condiciones de trabajo a su personal

5. ORGANIZACIÓN

- Conseguir la eficiencia en el desarrollo de la gestión académica
- Conseguir la eficiencia en la gestión de la administración y los servicios
- Disponer de un personal formado al máximo nivel que el desempeño de excelencia de sus puestos de trabajo requiere
- Disponer de los recursos tecnológicos y las infraestructuras que el desarrollo de servicios universitarios de calidad requiere
- Alcanzar niveles organizativos de calidad que proporcionen plena satisfacción a las expectativas de los usuarios

La UPV ha obtenido la Certificación del Modelo de Evaluación de la Actividad Docente, en el marco del programa DOCENTIA, de la ANECA-AVAP. La Certificación emitida con fecha 19 de diciembre de 2012 tiene una vigencia de cuatro años, válido hasta el 19 de diciembre de 2016.

CONSTRUCCIÓN DEL ÍNDICE DE ACTIVIDAD DOCENTE (IAD)

Marco conceptual teórico para la evaluación de la actividad docente del profesorado

La evaluación de la actividad docente forma parte inequívoca de los objetivos estratégicos formulados en el Plan Estratégico de la UPV y debe tener las siguientes características:

- La finalidad principal de la evaluación y del seguimiento de la calidad es favorecer las actuaciones orientadas a la mejora de la calidad y el reconocimiento de la dedicación y los resultados alcanzados
- La evaluación de la actividad del profesorado ha de tener carácter global integrando, en la medida de lo posible, el conjunto de las diferentes actividades académicas que se desarrollan en la Universidad
- Junto a la docencia y las responsabilidades de gobierno y dirección, se integran las actividades de investigación, transferencia de resultados de la investigación, formación continuada e implicación institucional en el marco de la planificación de los recursos de personal académico de los centros, departamentos e institutos de investigación

La Universitat Politècnica de València participó en la convocatoria DOCENTIA que puso en marcha la Agencia de Evaluación de la Calidad y Acreditación (2006), con el objetivo de atender a las demandas de las universidades para poder disponer de un mecanismo para la gestión de la calidad de la actividad docente del profesorado universitario. Dicho mecanismo se debe desarrollar mediante una metodología lo más eficiente posible con el fin de obtener unos resultados de acuerdo a los objetivos planteados.

En el marco conceptual teórico descrito en el programa DOCENTIA se define la ‘**actividad docente**’ como el conjunto de actuaciones, que se realizan dentro y fuera del aula, destinadas a favorecer el aprendizaje de los estudiantes con relación a los objetivos y competencias definidas en un plan de estudios y en un contexto institucional determinado. En consecuencia, la actividad docente implica la planificación y gestión de la enseñanza a clase vacía, el despliegue de métodos de enseñanza, actividades de aprendizaje y de evaluación a clase llena, así como la posterior revisión y mejora de las actuaciones realizadas (Agencia Nacional de Evaluación de la Calidad y Acreditación, 2006).

Como consecuencia, se debe entender por ‘**evaluación de la actividad docente**’ la valoración sistemática de la actuación del profesorado considerando su rol profesional y su contribución para conseguir los objetivos de la titulación en la que está implicado, en función del contexto institucional en que ésta se desarrolla.

La actividad docente del profesorado implica diferentes actuaciones dirigidas a organizar, coordinar, planificar y enseñar a los estudiantes, así como a evaluar el aprendizaje. Dichas actuaciones se despliegan en respuesta a los objetivos formativos y competencias que se pretenden favorecer en los estudiantes. En consecuencia, la evaluación de la actividad docente debe reconocer el conjunto de actuaciones que el profesor desarrolla en su quehacer profesional (dimensión cuantitativa), pero también debe recoger la valoración global de tal actividad en relación con los objetivos de la titulación y de la propia institución (dimensión cualitativa).

Para llevar a cabo el proceso de evaluación de la actividad docente, el programa DOCENTIA propone un modelo de evaluación con tres dimensiones:

- **Planificación de la docencia**
- **Desarrollo de la enseñanza**
- **Resultados**

Estas tres dimensiones tienen como transversalidad la dedicación docente que actuaría como una dimensión cero o condición previa, de modo que si el profesor no desarrolla las actuaciones que la describen no procedería determinar la calidad de la misma.

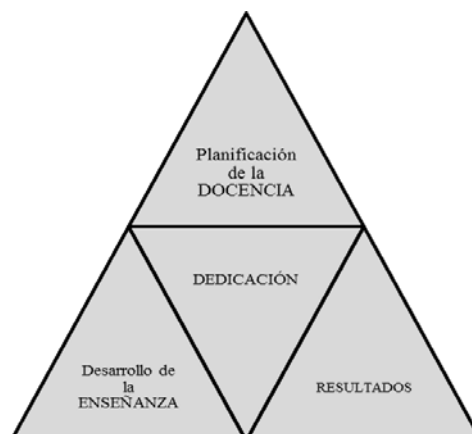


Figura 5. 1. Dimensiones del modelo de evaluación DOCENTIA

Los criterios de evaluación de la actividad docente, alineados con las dimensiones anteriores, en los que se apoya el modelo de evaluación son los siguientes:

- **Adecuación:** la actividad docente debe responder a los requerimientos establecidos por la universidad y el Centro en que se desarrolla, en relación a la organización, planificación, desarrollo de la enseñanza y a la evaluación de los aprendizajes de los estudiantes. Tales requerimientos deben estar alineados con los objetivos formativos y competencias recogidas en el plan de estudios y con los objetivos de la institución
- **Satisfacción:** la actividad docente debe generar una opinión favorable de los demás agentes implicados, en especial de estudiantes, colegas y responsables académicos
- **Eficiencia:** considerando los recursos que se ponen a disposición del profesor, la actividad docente debe propiciar en los estudiantes la consecución de los resultados previstos, en términos de objetivos formativos y competencias explicitadas en el plan de estudios
- **Orientación a la innovación docente:** la actividad docente debe abordarse desde una reflexión sobre la propia práctica que favorezca el aprendizaje del profesorado, a través de la autoformación o de la formación regulada por otras instancias, y debe desarrollarse desde una predisposición a introducir los cambios que mejoren el proceso de enseñanza-aprendizaje y que, por tanto, afecten al modo en que se planifica y se desarrolla la enseñanza o se evalúan los resultados de la misma

El proceso de evaluación de la actividad docente será de aplicación en todas las titulaciones oficiales de la UPV y todos los profesores de la UPV que imparten docencia en cualquiera de

estas enseñanzas serán objeto de evaluación, cualquiera que fuera su situación contractual y categoría profesional. Por lo tanto se define como un proceso individual y obligatorio.

El proceso de evaluación tiene carácter anual y se inicia a instancia de la universidad, a propuesta de la Comisión para la Evaluación de la Docencia de la UPV, con la remisión individual a cada profesor de su IAD (Índice de Actividad Docente) correspondiente anual.

A continuación se enumeran las finalidades de la evaluación de la actividad docente del profesorado que servirán para la toma de decisiones en diferentes ámbitos:

- La mejora de la calidad de la docencia
- El reconocimiento de la actividad docente
- Determinar planes de formación e innovación de la docencia
- Servir de herramienta para introducir elementos de objetividad, equidad y transparencia en la definición y la ejecución de las políticas de PDI, tales como la formación o la promoción del profesorado y en especial en todos los procesos que requieren una valoración de la actividad docente de las personas y las unidades académicas
- Disponer de indicadores cuantitativos comparables con el sistema de indicadores de la actividad investigadora que permitan conformar una posterior integración en un indicador general de la actividad académica personal
- Permitir incorporar las dimensiones cuantitativa y cualitativa de la actividad realizada
- La asignación de recursos económicos propios
- La asignación de incentivos, tales como los complementos retributivos existentes (anuales, quinquenios,...) o los que se pudieran dotar

La evaluación de la actividad docente del profesorado, de carácter universal y obligatorio, produce efectos en:

- Distribución y asignación de recursos propios de la universidad
- Reconocimiento de quinquenios: a tal efecto se exigirá para su reconocimiento que el profesor solicitante disponga de al menos cuatro años con resultado de la evaluación favorable o muy favorable
- Reconocimiento de otros complementos e incentivos económicos

Es voluntad de la universidad que la obtención de posibles resultados desfavorables en la evaluación de la docencia no implique necesariamente la inmediata separación de la labor docente. En sintonía y desarrollo con los criterios de la “European Association for Quality Assurance in Higher Education” (ENQA), la universidad proporcionará a los profesores de bajo rendimiento oportunidades para que mejoren sus habilidades mediante el diseño de planes de mejora individualizados, adecuados a las necesidades detectadas y en función de la dimensión en la que se detecten las carencias -planificación, desarrollo o resultados-, de tal forma que su seguimiento les permita mejorar.

Estos planes establecerán necesidades de formación específica, propuestas de aplicación de métodos de innovación docente, etc. La Comisión para la Evaluación de la Docencia realizará el seguimiento de las diferentes acciones diseñadas en los planes de mejora que se derivaran de la evaluación docente. La universidad podrá desarrollar, no obstante, en un futuro los mecanismos necesarios que permitan dar finalmente de baja de sus funciones docentes al personal que, tras sucesivos planes de mejora individualizados, persista en la obtención de resultados desfavorables.

Selección de indicadores del IAD

Para dar respuesta a los criterios descritos en el marco conceptual teórico la UPV define, para cada una de las dimensiones contempladas en el programa DOCENTIA, un conjunto de indicadores que reflejan los diferentes aspectos de la actividad docente asociados a cada una de ellas:

PLANIFICACIÓN DE LA DOCENCIA

Formación: participación en actividades de formación, impartida y recibida, tanto pedagógica y para la innovación educativa como relativa a las materias que se imparten. Este indicador se referencia a un período de cinco cursos académicos, de tal forma que las actividades de formación propiamente dichas pueden suplementarse con la experiencia acreditada del profesor mediante los resultados de las encuestas de opinión del alumnado.

Para un período de cinco cursos académicos se estima adecuado un total de 600 horas de participación en actividades de formación (120 horas anuales). La experiencia del profesor se considera acreditada con un resultado en las encuestas de opinión del alumnado superior a 6.

Por tanto, el indicador *Formación* está muy relacionado con el indicador K_1 que como se expone más adelante es el indicador que refleja los resultados de la opinión de los alumnos mediante una encuesta. La combinación de ambos elementos se recoge mediante el siguiente factor:

$$Forma = \left[\frac{\sum_{i=cur_{eva}-4}^{cur_{eva}} HF_i}{600} + \frac{1}{5} * \sum_{i=cur_{eva}-4}^{cur_{eva}} a_i \right] \geq 1$$

$$a_i = 1, si \ encuesta_i > 6$$

$$a_i = 0, si \ encuesta_i \leq 6$$

HF_i : horas de participación en actividades de formación en el año i

cur_{eva} : curso objeto de evaluación

De esta forma, siendo $Puntos_{formación} = 15$ el valor máximo de puntos IAD asociados a este indicador, el valor particularizado para cada profesor se obtiene mediante la siguiente expresión:

$$IAD_{FORMA} = Puntos_{formación} * Forma$$

Guía docente: si un profesor participa en la docencia de n asignaturas y en cada una de ellas imparte $CrImp_i$ créditos respecto al total de créditos impartidos en la asignatura, $CrTot_i$, el factor *Guía* se define como:

$$Guía = \sum_{i=1}^n \frac{CrImp_i}{CrTot_i}$$

De esta forma, siendo $Puntos_{guía} = 5$ el valor en puntos IAD asignado a la realización de la guía docente de una asignatura, el valor particularizado para cada profesor se obtiene mediante la siguiente expresión:

$$IAD_{GUIA} = Puntos_{guía} * Guía$$

Información de tutorías y grupos de docencia: se asigna al profesor un valor, en puntos IAD, de $Puntos_{tutoría} = 1$ por disponer de forma pública del horario de tutorías desde el inicio del curso académico y otro valor, en puntos IAD, de $Puntos_{horario}$ por publicar horarios de grupos de docencia con especificación del nombre del profesor. Pero $Puntos_{horario}$ no se considerará en la valoración del IAD, en tanto en cuanto la posible ausencia de esta información no puede ser atribuida exclusivamente al PDI:

$$IAD_{INFORMA} = Puntos_{tutoría} + Puntos_{horario}$$

Materiales docentes: se consideran los siguientes:

- Publicaciones docentes (libros, apuntes y otros): $Puntos_{publi}$
- Materiales docentes en plataforma educativa PoliformaT: $Puntos_{POLIT}$

Para la valoración de $Puntos_{publi}$ ($Puntos_{publi} \neq 10$) se considerarán los siguientes materiales docentes, con la valoración que se indica:

- Publicaciones con depósito legal de venta en el servicio de publicaciones de la UPV en el curso objeto de valoración: 2,5 puntos IAD
- Publicaciones con ISBN, editadas en los últimos cinco cursos, incluido el curso objeto de valoración: 5 puntos IAD

- Objetos y módulos de aprendizaje valorados favorablemente en los últimos tres cursos, incluido el curso objeto de valoración con un máximo de 5 puntos IAD:
 - Objetos de aprendizaje: 0,2 puntos IAD
 - Módulos de aprendizaje: 0,5 puntos IAD
- Artículos en revistas docentes, publicadas en el curso objeto de valoración:
 - Revistas no indexadas: 2,5 puntos IAD
 - Revistas indexadas: 5 puntos IAD
 - Cuando una publicación sea firmada por varios autores se multiplicará el valor unitario antedicho por el siguiente coeficiente:
 - Hasta 4 coautores: 1 punto IAD
 - Para 5 ó más coautores: $4/(n^{\circ} \text{ de coautores})$

K_{POLT} es el factor de satisfacción del material docente dispuestos en la plataforma educativa PoliformaT, que estará comprendido entre 0 y 1. La información para evaluar K_{POLT} se obtendrá mediante consulta a los alumnos matriculados, ponderando los resultados correspondientes a las diferentes asignaturas impartidas por un mismo PDI $[K_{POLT}]_{asi}$ y teniendo en cuenta únicamente aquellos resultados de la consulta en que hayan respondido, al menos, un 20% de los alumnos matriculados.

$$K_{POLT} = \frac{\sum_{i=1}^n [K_{POLT}]_{asi} \frac{Cr_{imp_i}}{Cr_{Tot_i}}}{\sum_{i=1}^n \frac{Cr_{imp_i}}{Cr_{Tot_i}}}$$

$$Puntos_{POLT} = 5K_{POLT}$$

El total de puntos IAD correspondiente a este indicador es:

$$IAD_{MAT} = Puntos_{Publi} + Puntos_{POLT}$$

DESARROLLO DE LA ENSEÑANZA

El baremo de los apartados correspondientes al desarrollo de la enseñanza se efectúa en horas para posteriormente transformarlas en puntos IAD de acuerdo con la siguiente equivalencia:

$$25 \text{ horas} = 1 \text{ punto IAD}$$

Docencia Impartida: este indicador recoge en horas la actividad del profesor, correspondiente a las asignaturas impartidas, teniendo en cuenta la experiencia adquirida en cada una de ellas de acuerdo con la siguiente escala:

FA_i es el factor de años de impartición de la asignatura i

- Primer año de impartición de la asignatura: 40 horas
- Segundo año de impartición de la asignatura: 30 horas
- Tercer año y siguientes de impartición de la asignatura: 20 horas

FI_i es el factor de idioma de impartición de la asignatura i

- Si la docencia se imparte en valenciano, inglés u otra lengua extranjera el número de horas correspondientes se multiplicará por 1,50.

$$Horas_{DOC} = \sum_{i=1}^n Crimp_i * FA_i * FI_i$$

siendo $Horas_{DOC}$ el número total de horas resultante de la docencia impartida. Según la equivalencia establecida el resultado correspondiente en puntos IAD es:

$$IAD_{DOC} = \frac{Horas_{Doc}}{25}$$

Tutorías realizadas: este indicador refleja en horas la actividad del profesor relativa a las atenciones tutoriales llevadas a cabo de acuerdo con el siguiente baremo:

- Tutorías convencionales: 216 horas (6 h/semana, 45 semanas, 80% de dedicación media por año)
- Participación en el Plan de Acción Tutorial (PATU): $P = 30$ horas
- Participación como tutor en el Programa de Formación Inicial Pedagógica del Profesorado Universitario (FIPPU): $F = 30$ horas

$$Horas_{TUT} = 216 * \frac{tut_{con}}{6} + P + F$$

tut_{con} : horas de tutoría según el tipo de contrato.

siendo $Horas_{TUT}$ el número total de horas resultante de las tutorías realizadas. Según la equivalencia establecida el resultado correspondiente en puntos IAD es:

$$IAD_{TUT} = \frac{Horas_{TUT}}{25}$$

Evaluación: si un profesor participa en la docencia de n asignaturas y en cada una de ellas imparte $Crimp_i$ créditos respecto al total de créditos impartidos en la asignatura, $CrTot_i$, la correspondencia en horas dedicadas por el profesor a las tareas de evaluación se obtiene mediante la siguiente expresión:

$$Horas_{EVA} = 0,2 * \sum_{i=1}^n Crmat_i * \frac{Crimp_i}{CrTot_i}$$

siendo $Crmat_i$ los créditos matriculados en la asignatura i

$$IAD_{EVA} = \frac{Horas_{EVA}}{25}$$

Otras actividades docentes: se consideran las siguientes:

- Dirección de Proyecto Fin de Carrera, de Trabajo Fin de Carrera o de Tesina de Máster: 40 horas (con un máximo en este apartado de 400 horas)
- Dirección de Trabajo para DEA: 40 horas
- Dirección de Tesis Doctoral: 40 horas, en el curso de lectura y en los dos cursos siguientes
- Participación en tribunales de PFC, TFC, Tesinas de Máster y DEA: 2 horas por cada unidad leída
- Participación en tribunales de Tesis Doctorales en la UPV: 10 horas por cada unidad leída
- Participación en tribunales de Tesis Doctorales en otras universidades: 20 horas por cada unidad leída
- Coordinación docente de cinco o más grupos en una misma asignatura: 10 horas
- Cátedras en empresa: 1 hora por cada 200 euros de presupuesto, con un límite máximo total de 400 horas para todos los participantes, sin que ninguno de ellos pueda superar a nivel individual una asignación de 200 horas
- Tutoría académica de prácticas en empresa: 5 horas por cada alumno tutorizado
- Organización y realización de viajes de prácticas: (20 horas+10 horas* n° de días del viaje), por cada viaje. A repartir entre todos los participantes
- Asistencia a congresos de docencia: 30 horas por cada congreso
- Participación en actividades de mejora y renovación docente en otras universidades: 30 horas por cada participación
- Docencia en otras universidades o teaching staff: (5*n° de horas impartidas)
- Participación en proyectos de innovación o grupos de innovación docente UPV: 30 horas

Siendo $Horas_{OTRAS}$ el número total de horas resultante de otras actividades docentes realizadas (con un máximo de 800 horas), según la equivalencia establecida el resultado correspondiente en puntos IAD es:

$$IAD_{OTRAS} = \frac{Horas_{OTRAS}}{25}$$

RESULTADOS

Encuesta de opinión del alumnado sobre la actuación docente del profesor: al resultado de la encuesta de opinión, $Resen$, se le asocia un factor K_1 con el siguiente criterio:

- Si $Resen \leq 5, K_1 = 1$
- Si $Resen > 5, K_1 = 1 + \frac{Resen-5}{10}$

Rendimiento Académico: siendo $TRmed$ el valor medio de la tasa de rendimiento académico de las asignaturas de igual carácter (troncales, obligatorias, optativas), - en el mismo centro, en la misma titulación, en el mismo curso - a la tasa de rendimiento académico de cada asignatura, $TRasig$, se le asocia un factor K_2 con el siguiente criterio:

- Si la tasa de rendimiento académico $TRasig$ es igual o superior a ($TRmed-10\%$):
$$K_{2i} = 1,5$$
- Si la tasa de rendimiento académico $TRasig$ es inferior a ($TRmed -10\%$):

$$K_{2i} = 1,5 - \left(\frac{TRmed - TRasig}{60} \right) \geq 1$$

Si un profesor imparte docencia en n asignaturas y en cada una de ellas $Crmat_i$ representa el número de créditos matriculados y K_{2i} es el factor de rendimiento académico asociado, el valor del factor K_2 se obtendrá según la siguiente expresión:

$$K_2 = \frac{\sum_{i=1}^n K_{2i} * Crmat_i}{\sum_{i=1}^n Crmat_i}$$

Cumplimiento de Tutorías: al cumplimiento de las tutorías académicas se le asocia un factor K_3 con el siguiente criterio:

- Si existen tres PISADOS (Parte de Incidencias Sobre la Actividad Docente) durante el curso académico, por incumplimiento de tutorías, sin resolución positiva por el centro:
 $K_3 = 1$
- En caso contrario: $K_3 = 1,50$

Cumplimiento de los plazos de entrega de actas: al cumplimiento de los plazos de entrega de actas se le asocia un factor K_4 con el siguiente criterio:

- Entrega anterior a la fecha inicial: $K_{conv} = 1,50$
- Entrega posterior a la fecha inicial: $K_{conv} = 1$

La valoración global en el curso de una asignatura, en relación con la fecha de entregas de actas será:

$$K_{4i} = \frac{\sum_{conv=1}^{nconv_i} K_{conv}}{nconv_i}$$

Si un profesor imparte docencia en n asignaturas el valor del factor K_4 se obtendrá según la siguiente expresión:

$$K_4 = \frac{\sum_{i=1}^n K_{4i} * \frac{CrImp_i}{CrTot_i}}{\sum_{i=1}^n \frac{CrImp_i}{CrTot_i}}$$

siendo $nconv_i$: nº de convocatorias de la asignatura i

En la tabla 4.2 se muestran las relaciones existentes entre los indicadores anteriormente definidos y los criterios de evaluación de la actividad docente alineados con las tres grandes dimensiones definidas.

CRITERIOS	DIMENSIONES		
	PLANIFICACIÓN DE LA DOCENCIA	DESARROLLO DE LA ENSEÑANZA	RESULTADOS
ADECUACIÓN	Guía docente	Docencia	Encuesta
	Información	Tutorías	Rendimiento
	Materiales	Evaluación	Cumplimientos
		Otras actividades	
SATISFACCIÓN	Formación	Docencia	Encuesta
	Guía docente	Tutorías	Rendimiento
	Información	Evaluación	Cumplimientos
	Materiales	Otras actividades	
EFICIENCIA	Guía docente	Docencia	Encuesta
	Información	Tutorías	Rendimiento
	Materiales	Evaluación	
		Otras actividades	
ORIENTACIÓN A LA INNOVACIÓN DOCENTE	Formación	Docencia	Encuesta
	Guía docente	Tutorías	
	Materiales	Evaluación	
		Otras actividades	

Tabla 5.2. Relación de los indicadores simples con las dimensiones y criterios del modelo de evaluación

Un aspecto muy importante en la recogida de datos de los diferentes indicadores simples es la importancia de disponer de las herramientas necesarias y eficaces para facilitar la mecanización del proceso. En este caso, resulta evidente la necesidad de estas herramientas para la implantación de un sistema de evaluación de la actividad docente del profesorado como el que se propone en la UPV.

La UPV dispone en la actualidad de bases de datos y herramientas informáticas necesarias para el tratamiento de la mayor parte de los elementos que definen los indicadores propuestos. Las fuentes de información que se utilizan para llevar a cabo y documentar la evaluación de la actividad docente del profesorado son las siguientes:

1. Bases de datos institucionales
2. Autoinforme del profesor
3. Informes de responsables académicos
4. Encuestas de opinión de los estudiantes

Los procedimientos de recogida de la información relativa a la actividad docente del profesorado, en lo referente a la actualización de datos académicos (plan de ordenación docente, guías docentes, horarios de tutorías, etc.) así como a los resultados de las encuestas de opinión del alumnado, seguirán los procesos y los plazos marcados institucionalmente.

Durante el curso académico el profesorado podrá declarar e incorporar a las bases de datos correspondientes todos aquellos elementos que correspondan a sus actividades personales, tales como publicaciones docentes, participaciones en congresos de docencia, etc. Al finalizar el curso, cada profesor dará traslado al correspondiente responsable académico de la documentación acreditativa de dichas actividades para su validación e incorporación definitiva a la documentación personal del profesor.

Tras cerrarse el proceso de validación de datos, la primera etapa de la evaluación consistirá en el traslado, por parte de la Comisión para la Evaluación de la Docencia de la UPV, de la información disponible en las bases de datos institucionales a los interesados. Dicha información contendrá los datos personalizados relativos a cada uno de los indicadores del modelo de evaluación incluidos los resultados de las encuestas realizadas por los estudiantes. En el plazo previsto, el profesor procederá a completar o modificar, en el caso que sea necesario, dicha información dando traslado a los responsables académicos para la validación, si procede, de la información añadida o modificada y aportando si lo estima oportuno las valoraciones y reflexiones personales sobre cada una de las dimensiones: planificación de la docencia, desarrollo de la enseñanza y resultados de su actividad docente. El documento que recoja la información personalizada así como las reflexiones y valoraciones del profesor, constituirá su Autoinforme.

Los responsables académicos (directores de centros, de departamentos y de los órganos responsables de los másteres) emitirán informes personalizados referentes a todas aquellas cuestiones de la actividad docente del profesor que sean de su incumbencia. Dichos informes especificarán los datos concretos del profesor relativos a cada uno de los indicadores del modelo de evaluación que sean de su responsabilidad y, en su caso, aquellos datos que fueran relevantes de la actividad del profesor, tales como las felicitaciones o valoraciones meritorias sobre el desarrollo de la enseñanza de las que tenga constancia, las incidencias detectadas en el

desarrollo de la docencia tanto en el aula como en la realización de actividades prácticas o en las tutorías (quejas o reclamaciones de los estudiantes) y las incidencias detectadas en el desarrollo de las evaluaciones de los aprendizajes de los estudiantes: quejas o reclamaciones, incumplimiento en el plazo de entrega de actas, etc.

Normalización de los datos

Todos los indicadores definidos anteriormente se miden en un sistema de puntos IAD propuesto por la UPV, por tanto no se realiza ninguna normalización posterior a pesar de que los indicadores se miden en rangos de variación muy diferentes entre ellos.

Ponderación y agregación de los indicadores simples

El sistema de puntos IAD que se propone goza de dos particularidades que es necesario resaltar:

- Se trata de un sistema sumativo que no limita a priori el número de puntos que se pueden conseguir
- Las actividades ordinarias realizadas se reconocen mediante los correspondientes puntos IAD, pero las puntuaciones obtenidas pueden incrementarse en función del nivel de los resultados conseguidos

El resultado final de puntos IAD correspondiente a la dimensión Planificación de la Docencia es:

$$IAD_{PLAN} = IAD_{FORMA} + IAD_{GUIA} + IAD_{INFORMA} + IAD_{MAT}$$

El resultado final de puntos IAD correspondiente a la dimensión Desarrollo de la Enseñanza es:

$$IAD_{DESA} = IAD_{DOC} + IAD_{TUT} + IAD_{EVA} + IAD_{OTRAS}$$

Por tanto, se asigna la unidad como peso para cada uno de los indicadores dentro de las dos primeras dimensiones y se agrega con un método aditivo cada una de ellas.

Ponderando el IAD correspondiente a la Planificación de la Docencia en un 30% y el IAD correspondiente al Desarrollo de la Enseñanza en un 70%, se agregan las dos primeras dimensiones de la siguiente forma:

$$Puntos_{IAD} = 0,30 * IAD_{PLAN} + 0,70 * IAD_{DESA}$$

Los cuatro factores K_i asociados a los indicadores de Resultados configuran un coeficiente global de resultados, K_{res} , que se obtiene ponderadamente mediante la siguiente expresión:

$$K_{res} = 0,60 * K_1 + 0,30 * K_2 + 0,05 * K_3 + 0,05 * K_4$$

Con ello, el valor final del índice de actividad docente correspondiente al año X es:

$$IAD_X = Puntos_{IAD} * K_{res}$$

Los umbrales de cualificación del IAD se obtienen a partir de las siguientes expresiones, en función de la capacidad docente de cada profesor (CAPDOC):

- EXCELENTE: $U_E = 22 + 2,1 * (CAPDOC) - \left(\frac{CAPDOC}{8,25}\right)^2$
- NOTABLE: $U_N = 10 + 1,5 * (CAPDOC) - \left(\frac{CAPDOC}{8,64}\right)^2$
- ACEPTABLE: $U_A = 5 + 1,2 * (CAPDOC) - \left(\frac{CAPDOC}{8,75}\right)^2$

La capacidad docente del profesor la establece el Vicerrectorado de Profesorado y Ordenación Académica (definida como los créditos por contrato menos las reducciones). Las reducciones de créditos por cargos que afecten a la capacidad docente se determinarán, para cada curso académico, proporcionalmente al tiempo de ejercicio del mismo. Las bajas de duración superior a sesenta días afectarán, proporcionalmente, a la determinación de la capacidad docente.

Determinación del Índice de Actividad Docente Normalizado (IAD_N)

El IAD es un índice que se mide en puntos calculados con las expresiones y equivalencias descritas anteriormente. La forma en que se ha diseñado hace que el valor del IAD refleje el volumen de actividad docente y ésta dependa lógicamente de la denominada capacidad docente. Los umbrales para la calificación en las categorías de Excelente, Notable, Aceptable o Desfavorable, se establecen en función de la capacidad docente de cada profesor. Con el objetivo de proporcionar un índice relativo, comparable e independiente de la capacidad docente se define el Índice de Actividad Docente Normalizado (IAD_N) que también permite evaluar la evolución en el tiempo de la actividad docente.

El IAD_N se calcula asignando un valor de 50 puntos al valor del umbral entre Desfavorable/Aceptable, 60 puntos al umbral entre Aceptable/Notable y 100 puntos al umbral entre Notable/Excelente, transformando el valor del IAD de acuerdo con la siguiente expresión:

$$IAD_N = \begin{cases} 50 * \frac{IAD}{U_A} & \text{si } U_A \geq IAD \\ 50 + 50 * \frac{(IAD - U_A)}{(U_N - U_A)} & \text{si } U_A < IAD \leq U_N \\ 60 + 60 * \frac{(IAD - U_N)}{(U_E - U_N)} & \text{si } U_N < IAD \end{cases}$$

siendo U_A , U_N y U_E los umbrales de Aceptable, Notable y Excelente respectivamente.

De esta forma, la calificación de Desfavorable se corresponde con un valor inferior a 50 en el IAD_N, la de Aceptable a un valor de entre 50 y 60, la de Notable a un valor entre 60 y 100, y un valor de IAD_N superior a 100 significa que se ha obtenido la calificación de Excelente.

Procesos de evaluación y agentes que intervienen

La Comisión para la Evaluación de la Docencia (CED) establecerá los criterios y procedimientos de evaluación de la docencia que se harán públicos mediante comunicación personalizada a todos los interesados comunicando asimismo el calendario de aplicación.

Tras la comunicación al interesado de la resolución del proceso de evaluación de la docencia, éste dispondrá de dos semanas para solicitar la revisión de la evaluación realizada ante la Comisión de Garantías.

Una vez resueltas las reclamaciones, la Comisión para la Evaluación de la Docencia emitirá finalmente el Informe Anual. Las resoluciones desfavorables de la evaluación de la docencia se vincularán al mantenimiento de la actividad docente del profesor pudiendo suponer la redefinición del encargo docente.

En todo caso, las resoluciones desfavorables incorporarán la propuesta de planes de mejora personalizados que serán gestionados por el Instituto de Ciencias de la Educación y que abarcarán las tres dimensiones evaluadas. En dichos planes se contemplarán actividades de formación y asesoramiento del profesor, así como el seguimiento de la ejecución de estos planes de mejora individual y su evaluación. En última instancia el Instituto de Ciencias de la Educación emitirá un informe a la Comisión para la Evaluación de la Docencia, que resolverá en consecuencia.

Las reclamaciones se resolverán por la Comisión de Garantías en un plazo máximo de dos meses tras la finalización del plazo de solicitud de revisión posterior a la resolución inicial del proceso de evaluación de la docencia.

Igualmente se solicitará al interesado que aporte el nuevo Autoinforme, complementando el Autoinforme inicial con información adicional sobre los mismos aspectos anteriormente señalados.

El proceso establecido hace necesario conocer los niveles de actividad docente desarrollados por el profesorado. Los resultados de la evaluación de la actividad docente tendrán, como criterio general, carácter público. Dicho carácter deberá compaginar la protección necesaria de los datos personalizados con la oportunidad de la publicidad de los resultados de la evaluación docente. A tal fin la universidad, en aplicación de la legislación vigente, podrá, si lo estimara oportuno, hacer públicos los aspectos, datos y resultados del proceso.

La CED hará público el informe anual a través de los medios que se consideren convenientes. La información de detalle se proporcionará a los correspondientes responsables académicos, agregada al ámbito de su responsabilidad. La UPV y la CED garantizarán la confidencialidad en la gestión de los datos de carácter personal. Se deberá garantizar la transparencia en las actuaciones y las valoraciones que se realicen al amparo del proceso establecido.

Tendrán carácter confidencial los Autoinformes elaborados por los profesores, las evaluaciones individuales y singulares que se realicen en el proceso, así como las encuestas de los estudiantes y los resultados de éstas. Los informes de los responsables académicos tendrán igualmente carácter confidencial.

5.3. Modelo de evaluación de la actividad académica del profesorado propuesto por la UPV

El Plan Estratégico de la Universitat Politècnica de València en su eje IV. PERSONAS, propone como objetivo IV.1.: “Propiciar un capital humano comprometido con la excelencia docente, investigadora y de gestión, desarrollando las siguientes líneas estratégicas y planes de acción:

- IV.1. L1 La evaluación permanente de la actividad laboral del personal de la UPV como sistema de búsqueda de la excelencia. Esta evaluación dará lugar al reconocimiento de resultados por la organización, mediante diferentes instrumentos: incentivos, carrera profesional, etc.
- Plan VALORA PERSONAS: disponer de un sistema que permita valorar, mediante indicadores objetivados, la contribución anual de cada uno de los empleados al desarrollo de la actividad y a la obtención de resultados de la UPV en la totalidad de los ámbitos en los que desarrolle su actividad universitaria.

- Plan RECONOCIMIENTO: disponer de un sistema de incentivos que contemple el reconocimiento al rendimiento de las personas considerando conjuntamente la totalidad de los ámbitos de desarrollo de sus funciones (docencia, investigación y gestión), así como articular una carrera profesional basada en la permanencia continuada del rendimiento de las personas en niveles de excelencia.”

Sin embargo, está pendiente de desarrollo el Plan Valora Personas que permita objetivar y reconocer la contribución anual de cada uno de los PDI al conjunto de la actividad y los resultados académicos del profesorado: docencia, I+D+i y gestión académica. Para ello, la UPV se ha propuesto construir un indicador sintético de la actividad y los resultados del profesorado que se denomina Índice de Actividad Académica (IAA). Esta valoración, consolidada del esfuerzo del profesorado universitario, debe permitir gestionar más equitativamente los encargos de trabajo administrados por la universidad: obligaciones docentes, cargos académicos y reducciones lectivas, desde la disponibilidad de una información precisa sobre la actividad del PDI en todos sus ámbitos: docencia, I+D+i, y gestión.

El IAA pretende englobar en un único valor los diferentes tipos de actividad del profesorado. La actividad académica del profesorado de la universidad puede decirse que es la suma de la actividad de gestión, de docencia y de investigación. Podría incluirse la formación como una actividad más del profesorado, independiente de las anteriores, tarea que es muy importante especialmente en su etapa inicial como profesor. La dificultad que supone medir esta tarea y la falta en la UPV de un indicador que la evalúe, no permite incorporarla de forma independiente al resto de actividades del profesorado en la UPV, por lo que se mide a partir de la actividad investigadora, docente y de gestión.

La actividad investigadora se valora esencialmente por los resultados y, por tanto, debe considerarse que la actividad que produce un resultado puede extenderse hasta varios años anteriores. Por ello, resulta más adecuado valorar la actividad investigadora mediante el indicador VAIP “Valoración de la Actividad Investigadora Personalizada”. La metodología desarrollada para este indicador se puede encontrar en el reglamento para la evaluación de la actividad investigadora, desarrollo e innovación en la UPV (Universitat Politècnica de València, 2008).

La actividad docente resulta del conjunto de tareas que hay que realizar para que los alumnos adquieran las competencias y habilidades definidas en los diferentes planes de estudios. Estas tareas se extienden más allá de las clases lectivas y es por eso que resulta más adecuado utilizar el Índice de Actividad Docente (IAD) descrito anteriormente.

La actividad de gestión queda definida en la Normativa de Organización Docente (NOD) de la UPV. Las actividades principales son: la reducción de créditos lectivos a los cargos académicos y la dedicación del profesorado a las sesiones de las comisiones y demás órganos colegiados necesarios para la dirección y gestión de la universidad y sus unidades académicas.

Esta última actividad ha sido contabilizada en términos de créditos de gestión (Cr.Ges) asignando un número de horas de dedicación al año a cada tarea. Cada Cr.Ges equivale a $1.560/33=47,27$ horas y a un crédito lectivo reducido en la NOD (Universitat Politècnica de València, 2012). La notación empleada para esta dimensión es AG (Actividades de Gestión).

Índice de Actividad Académica del profesorado

El Índice de Actividad Académica (IAA_c) de un profesor c es la suma ponderada de las actividades de docencia, investigación y gestión medidos por sus respectivos indicadores ($IAD_c, VAIP_c, AG_c$). Los indicadores se multiplican por los porcentajes ($p_{doc}, p_{inv}, p_{ges}$) que representan la importancia asignada por el modelo de la UPV a cada una de las tres actividades y se dividen entre los valores medios de cada uno de los indicadores ($IAD_{med,UPV}, VAIP_{med,UPV}, AG_{med,UPV}$).

$$IAA_{UPV_c} = p_{doc} * \frac{IAD_c^*}{IAD_{med,UPV}} + p_{inv} * \frac{VAIP_c^*}{VAIP_{med,UPV}} + p_{ges} * \frac{AG_c^*}{AG_{med,UPV}}$$

El IAA además considera las siguientes restricciones en las puntuaciones de las dimensiones:

- Para la dimensión IAD, la puntuación se limita al valor de IAD superado por el 3% del profesorado de la UPV. Este valor corresponde a una puntuación de 68,5
- En la dimensión VAIP, la puntuación se limita al valor del indicador VAIP que es superado por el 3% del profesorado de la UPV. Este valor corresponde a una puntuación de 208
- Los créditos de la dimensión de gestión se limitan a 20

Para los profesores asociados, la eventual actividad de investigación y gestión no se considera, valorándose únicamente la actividad docente.

Para el cálculo del IAA del año 20XX, se utilizarán los créditos de gestión del curso 20(XX-1)/20XX, el IAD del año 20(XX-1) y el VAIP del año 20(XX-2).

Debido tanto a su carácter estratégico como a su dependencia de la estructura de plantilla de la UPV, los valores de p_{doc} , p_{inv} y p_{ges} serán establecidos por acuerdo de Consejo de Gobierno y revisados, como mínimo, cada 4 años.

En el último informe publicado por la UPV (2012) sobre el Índice de Actividad Académica se fija el valor de p_{ges} en un 6%, el valor de p_{doc} en un 67,60% y el valor de p_{inv} en un 26,40%.

Capítulo 6

Propuesta metodológica para la construcción de un Indicador Compuesto que evalúe la actividad docente en la UPV

6.1. Introducción

Este capítulo se ha centrado en el estudio de las diversas técnicas metodológicas para la construcción de un IC que evalúe la actividad docente del profesorado de la UPV. Para ello, se han desarrollado diferentes estrategias metodológicas basadas en estudios realizados por investigadores pioneros en el tema. La metodología se divide en etapas en las que se aplican las técnicas de análisis más adecuadas en base a los datos disponibles para el estudio. El desarrollo teórico de las etapas y los diversos análisis que se pueden aplicar en cada una de ellas se exponen con más detalle en el capítulo 3 de la Tesis Doctoral. Las técnicas de análisis aplicadas en algunas de las etapas que no se han visto en el capítulo 3, por ser muy particulares de cada conjunto de datos, se muestran en este capítulo.

A continuación se enumeran las etapas planteadas en este estudio:

1. **Desarrollo del marco conceptual teórico.** Definición del objetivo de estudio
2. **Selección del conjunto de indicadores simples** a partir del marco conceptual teórico definido
3. **Tratamiento de datos.** Análisis de los estadísticos básicos de los indicadores simples a nivel univariante
 - **Detección de valores atípicos a nivel univariante.** Aplicación de un umbral predeterminado sobre los datos estandarizados para la detección de valores atípicos a nivel univariante
 - **Detección de valores atípicos a nivel multivariante.** Cálculo de la distancia al cuadrado de Mahalanobis con el fin de identificar aquellas observaciones que tienen un comportamiento anómalo al resto en el conjunto de indicadores simples

- **Transformación de los indicadores simples.** Aplicación de la familia de transformaciones Box-Cox sobre los indicadores simples con el fin de mejorar los supuestos del análisis multivariante y reducir el número de valores extremos
- 4. **Análisis multivariante.** Análisis de la estructura interna de los datos
- 5. **Normalización de los datos.** Aplicación de las técnicas de normalización z-score y min-max para obtener un conjunto de indicadores comparables
- 6. **Ponderación y agregación de los indicadores simples.** Aplicación de los métodos de asignación presupuestaria y asignación igualitaria de pesos combinados con las técnicas de agregación lineal y geométrica
- 7. **Escenarios de trabajo.** Definición de alternativos escenarios de trabajo según las diferentes técnicas de normalización, ponderación y agregación empleadas
- 8. **Análisis de sensibilidad.** Estudio del análisis de sensibilidad global mediante la aplicación del método no paramétrico *State Dependent Parameter* para el cálculo del efecto principal o coeficiente de importancia de cada indicador simple sobre el IC
- 9. **Análisis de incertidumbre.** Estudio del análisis de incertidumbre de cada uno de los escenarios definidos con el fin de analizar la robustez de la metodología aplicada. Estudio de las distintas técnicas de visualización de los resultados

Para realizar los análisis se han empleado diferentes programas estadísticos dependiendo de las opciones y utilidades de cada uno de ellos. Para el tratamiento previo de los datos, la detección de valores atípicos y la parte de análisis multivariante se ha utilizado el software SPSS 16.0. La transformación de los indicadores mediante la familia de transformaciones Box-Cox se ha realizado con el software Minitab v. 15. Por otra parte, en las dos últimas etapas metodológicas en la que se aplica un análisis de sensibilidad y de incertidumbre a los indicadores compuestos construidos se ha utilizado el software libre SIMLAB 2.2.1 y el software Matlab 7.12.0.635 (R2011a).

A continuación se muestra el desarrollo de cada una de las etapas metodológicas enumeradas anteriormente para la construcción de un IC que evalúe la actividad docente del profesorado de la UPV.

6.2. Técnicas metodológicas para la construcción del Índice de Actividad Docente del profesorado de la UPV

6.2.1. Desarrollo del marco conceptual teórico

La evaluación de la actividad docente del profesorado de la UPV tiene como objetivo principal mejorar la calidad de sus enseñanzas y el desarrollo profesional de sus docentes. Con este fin, la UPV define un modelo basado en el Programa de Apoyo a la Evaluación de la Actividad Docente que recibe el nombre DOCENTIA. Este programa “*pretende satisfacer las*

demandas de las universidades y la necesidad del sistema educativo de disponer de un modelo y de unos procedimientos para garantizar la calidad del profesorado universitario y favorecer su desarrollo y reconocimiento” (Agencia Nacional de Evaluación de la Calidad y Acreditación, 2006).

El programa DOCENTIA se enmarca en el conjunto de actuaciones destinadas a construir un escenario para favorecer los principios de calidad y competitividad entre universidades europeas contribuyendo al Espacio Europeo de Educación Superior. Este programa se ha diseñado de modo que la evaluación de la actividad docente del profesorado esté alineada con el marco estratégico definido por cada universidad.

Basándose en los principios del programa DOCENTIA y en su marco de referencia, la **actividad docente** se define como el “conjunto de actuaciones, que se realizan dentro y fuera del aula, destinadas a favorecer el aprendizaje de los estudiantes con relación a los objetivos y competencias definidas en un plan de estudios y en un contexto institucional determinado”. Además, se entiende por **evaluación de la actividad docente** “la valoración sistemática de la adecuación del profesorado considerando su rol profesional y su contribución para conseguir los objetivos de la titulación en la que está implicado, en función del contexto institucional en el que se desarrolla” (Agencia Nacional de Evaluación de la Calidad y Acreditación, 2006).

Así pues, para lograr la evaluación de la actividad docente el modelo tiene en cuenta tres grandes aspectos de la actividad del profesorado que son las tres dimensiones que definen el concepto que se quiere medir y que, a su vez, están definidas por un conjunto de indicadores simples. Las dimensiones consideradas se exponen a continuación:

– **Planificación de la docencia**

Bajo la denominación de planificación de la docencia se recogen las siguientes actividades:

- Modalidades de organización de la enseñanza (clases prácticas y prácticas externas, seminarios, clases teóricas, tutorías,...)
- Coordinación con otras actuaciones docentes en el ámbito de la titulación y de acuerdo con la política del centro y los departamentos
- Planificación de la enseñanza y del aprendizaje con relación a las materias o asignaturas impartidas
- Resultados de aprendizaje previstos
- Actividades de aprendizaje previstas
- Criterios y métodos de evaluación
- Materiales y recursos para la docencia

– **Desarrollo de la enseñanza**

Esta dimensión recoge el conjunto de actividades para el desarrollo de la enseñanza y la evaluación del aprendizaje. En este sentido, las actuaciones que desarrolla el profesorado se pueden desagregar en actividades de enseñanza y aprendizaje realizadas, y procedimientos de evaluación aplicados

- Resultados

La dimensión 'Resultados' recoge información relativa a los resultados en términos de objetivos formativos logrados por los estudiantes y los resultados para una revisión y mejora de la actividad docente

6.2.2. Selección del conjunto de indicadores simples

La figura 6.1 muestra de forma esquemática el marco conceptual teórico definido para la construcción de un IC cuyo objetivo principal es medir la actividad docente del profesorado de la UPV. Para ello, se ha definido el concepto que se pretende estudiar mediante la desagregación en las dimensiones nombradas anteriormente de acuerdo con el programa DOCENTIA. A su vez, la UPV ha propuesto cuatro indicadores simples para la evaluación de cada dimensión (Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV, 2010). Una descripción más detallada de la definición de cada indicador y de su procedimiento de obtención se puede encontrar en el capítulo 5 de la Tesis Doctoral.

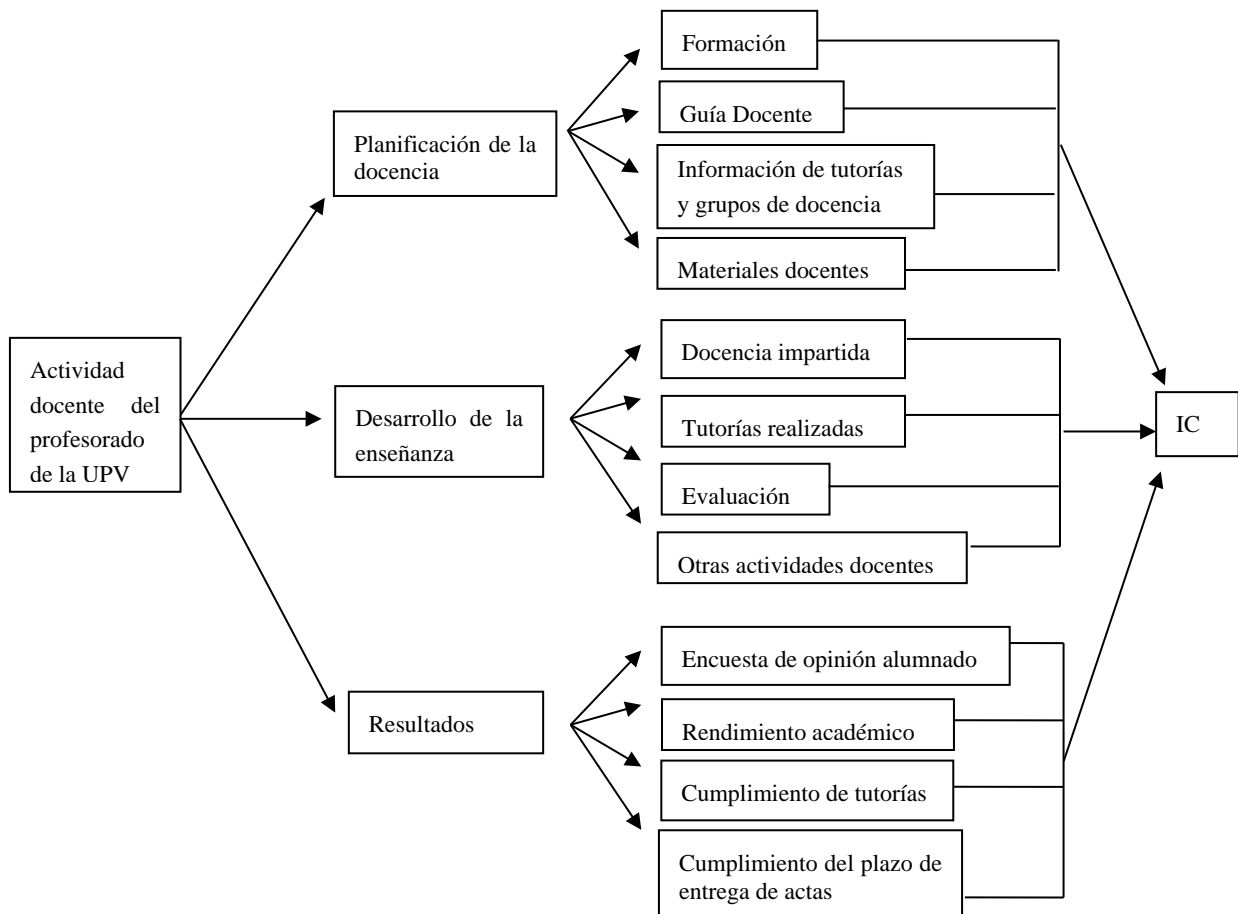


Figura 6.1. Diagrama de flujo del marco conceptual teórico del IAD

A continuación se muestra la codificación que se va a utilizar a lo largo de todo el documento para cada uno de los indicadores simples.

Planificación de la docencia	
IAD_{FORMA}	Formación
IAD_{GUIA}	Guía Docente
$IAD_{INFORMA}$	Información de tutorías
IAD_{MAT}	Materiales docentes
Desarrollo de la enseñanza	
IAD_{DOC}	Docencia impartida
IAD_{TUT}	Tutorías realizadas
IAD_{EVA}	Evaluación
IAD_{OTRAS}	Otras actividades docentes
Resultados	
K_1	Encuesta de opinión alumnado
K_2	Rendimiento académico
K_3	Cumplimiento de tutorías
K_4	Cumplimiento del plazo de entrega de actas

Tabla 6.1. Abreviaturas de los indicadores simples

6.2.3. Tratamiento de datos

Para el presente trabajo se dispone de una base de datos formada por 2.741 observaciones que corresponden a la población total del profesorado de la UPV, a los cuales se les ha medido los 12 indicadores simples durante el curso académico 2007/2008. Los datos se conocen de forma totalmente anónima y proceden del Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV.

La primera parte del análisis estadístico se ha centrado en el estudio de la calidad de los indicadores candidatos a definir la actividad docente del profesorado. Para ello, los indicadores se han analizado a nivel univariante con los siguientes objetivos: calcular estadísticos descriptivos básicos de los indicadores y obtener conclusiones de cada uno de ellos, identificar valores faltantes y aplicar las técnicas de imputación de datos si se considera adecuado, e identificar posibles valores atípicos y adoptar las decisiones convenientes.

6.2.3.1. Análisis descriptivo univariante de los indicadores

En la tabla 6.2 se muestran los resultados del estudio descriptivo del comportamiento de cada uno de los indicadores simples después de realizar un filtrado en el que se ha excluido al profesorado con unas puntuaciones que no son válidas para su evaluación según la normativa vigente de la UPV.

	N	Media	Desv. típ.	Rango de variación		Asimetría	Curtosis
				Mínimo	Máximo		
IAD_{FORMA}	2689	12,90	4,46	0,00	15,00	-2,00	2,65
IAD_{GUIA}	2689	5,95	4,78	0,00	28,35	0,93	0,68
$IAD_{INFORMA}$	2689	0,83	0,38	0,00	1,00		
IAD_{MAT}	2689	7,77	3,92	5,00	15,00	0,99	-0,68
IAD_{DOC}	2689	16,45	6,43	4,40	54,00	0,77	1,00
IAD_{TUT}	2689	8,41	1,21	2,88	11,04	-2,12	4,56
IAD_{EVA}	2689	5,69	3,40	0,34	30,61	1,50	3,74
IAD_{OTRAS}	2689	5,31	5,75	0,00	32,00	1,95	4,33
K_1	2689	1,23	0,12	1,00	1,50	-0,35	-0,60
K_2	2689	1,44	0,10	1,00	1,50	-2,02	3,76
K_3	2689	1,50	0,00	1,50	1,50		
K_4	2689	1,47	0,07	1,00	1,50	-2,54	7,17

Tabla 6.2. Estadísticos descriptivos de los indicadores

En general, se observa una gran diferencia entre las variables en cuanto al rango de variación y a la desviación típica. Este resultado aporta indicios del comportamiento heterogéneo de las variables de estudio, a pesar de estar medidas en la misma unidad (puntos IAD).

Los indicadores que alcanzan, en término medio, un número mayor de puntos IAD son IAD_{FORMA} ($\overline{IAD_{FORMA}} = 12,90$) e IAD_{DOC} ($\overline{IAD_{DOC}} = 16,45$). Estos indicadores reflejan la preparación del profesorado para afrontar su actividad docente y la cantidad de docencia impartida, respectivamente. Además, presentan una desviación típica elevada, $\sigma_{IAD_{FORMA}} = 4,46$ y $\sigma_{IAD_{DOC}} = 6,43$. Por otra parte, el indicador $IAD_{INFORMA}$ que mide la actividad del profesor relativa a las acciones tutoriales y los indicadores K_1 , K_2 , K_3 y K_4 , que evalúan los resultados derivados de la actividad docente desarrollada por cada profesor, presentan una media muy baja con respecto al resto de indicadores simples y una desviación típica prácticamente nula.

El indicador K_3 es constante para todas las observaciones, por tanto no tiene sentido incluir este indicador en algunos de los análisis que se muestran a lo largo de este capítulo, puesto que no aporta información adicional sobre las posibles diferencias entre el profesorado en términos de su actividad docente.

Teniendo en cuenta que una variable es asimétrica cuando el valor absoluto del coeficiente de asimetría es, aproximadamente, mayor que 1 (Peña, 2002), los indicadores IAD_{FORMA} , IAD_{TUT} , IAD_{EVA} , IAD_{OTRAS} , K_2 y K_4 presentan una clara asimetría.

Una característica deseable en un conjunto de datos es la homogeneidad, propiedad que puede medirse mediante el coeficiente de curtosis. Este coeficiente mide la relación entre la variabilidad de las desviaciones y la desviación media. Para una distribución normal el estadístico de curtosis toma el valor 0. Si hay unos pocos datos atípicos alejados del resto, la variabilidad de las desviaciones será grande y el coeficiente de curtosis será alto. En este estudio

prácticamente todos los indicadores alcanzan valores del coeficiente de curtosis alejados de cero y este resultado es indicativo de la heterogeneidad que presenta el conjunto de datos.

Una forma de detectar valores atípicos es identificar aquellas variables que presenten un valor en la curtosis mayor que 7 u 8 (Peña, 2002). En este caso, el coeficiente de curtosis del indicador K_4 alcanza el valor 7 y, por tanto, este resultado informa de la existencia de posibles valores atípicos en este indicador.

Adicionalmente al cálculo de los estadísticos básicos se ha realizado un estudio con el fin de diagnosticar la normalidad de las distribuciones de los indicadores que es un requerimiento para la correcta estimación de los parámetros del modelo, así como para los contrastes de significación asociados a estos parámetros. Los contrastes χ^2 de Kolmogorov-Smirnov, Shapiro-Wilks o los contrastes basados en coeficientes de asimetría y curtosis permiten realizar este estudio (Peña, 2002).

Al aplicar el test de Shapiro-Wilks y el test de Kolmogorov-Smirnov a los indicadores se obtiene un p-valor prácticamente nulo. Por tanto, asumiendo un nivel de significación de 0,05 se concluye que ninguna de las variables sigue una distribución normal. La asimetría de las distribuciones de cada uno de los indicadores es un indicativo del resultado obtenido con los test de normalidad. Mediante los contrastes basados en coeficientes de asimetría y curtosis se obtiene el mismo resultado (datos no mostrados).

No obstante, los test de significación son muy sensibles a las muestras de gran tamaño (superior a 1.000 observaciones) (Hair et al., 2007), por lo que también se deben contrastar estos resultados con los métodos gráficos como el histograma de cada uno de los indicadores o el gráfico de papel probabilístico normal. Estas representaciones muestran que las únicas variables que se aproximan a una distribución normal son IAD_{GUIA} , IAD_{DOC} y K_1 (datos no mostrados).

6.2.3.2. Imputación de los datos

En este estudio no ha sido necesario aplicar técnicas de imputación de datos, puesto que se han podido medir todos los indicadores sobre cada una de las observaciones. Según el Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV (2009), los sistemas y protocolos de adquisición de datos en la UPV han sido diseñados bajo el principio de que deben estar disponibles, de forma automática en la práctica, la totalidad de los indicadores para la fase inicial del desarrollo del IC sin solicitar al profesorado ninguna información que esté contenida en las bases de datos de la UPV y procurando que toda la información necesaria para la determinación del mismo haya sido previamente registrada.

6.2.3.3. Detección de valores atípicos a nivel univariante

En esta investigación se han identificado como datos atípicos aquellas observaciones que parecen haberse generado de forma distinta al resto de los datos debido a errores de medición o de transcripción.

De entre todas las posibles técnicas empleadas para la detección de valores atípicos a nivel univariante se ha recurrido a identificar aquellas observaciones cuyo valor estandarizado en los indicadores excede de 4 unidades en valor absoluto. Se ha considerado adecuado este umbral puesto que el número de observaciones del estudio es grande (Hair et al., 2007).

Los resultados muestran un total de 80 observaciones con valores extremos en algunos de los indicadores. Este conjunto de observaciones está formado con un 31,25% por profesores Asociados y profesores Ayudantes, un 53,75% por profesores Titulares de Escuela Universitaria, Titulares Universitarios y Catedráticos de Universidad, un 2,5 % por Catedráticos de Escuela Universitaria, un 5% por profesores Colaboradores Doctores y un 7,5% por profesores Colaboradores.

El primer grupo presenta, en su mayoría, valores extremos muy bajos en las variables K_2 y K_4 que corresponden al rendimiento académico y al cumplimiento de los plazos de entrega de actas. Sin embargo, el segundo grupo presenta valores extremos muy altos mayormente en las variables IAD_{EVA} e IAD_{OTRAS} . Se trata de profesores que dedican muchas horas a las tareas de evaluación, ya que el número de alumnos matriculados en sus asignaturas es elevado y, además, participan en otras actividades docentes aparte de las habituales. No se ha podido obtener ninguna conclusión general de los otros grupos, puesto que cada profesor presenta valores extremos en diferentes indicadores.

Se ha comprobado que los valores extremos, detectados hasta el momento, no se deben a errores de medición ni de transcripción. Por tanto, no se van a considerar como datos atípicos para este estudio, puesto que se ha decidido mantener los datos originales para definir estrategias metodológicas alternativas a la propuesta por la UPV en la que se consideran todas las observaciones.

Además de ayudar a detectar posibles anomalías en el comportamiento de algunas observaciones, el estudio de los valores extremos habitualmente permite mejorar y precisar la definición del sistema de medición utilizado.

6.2.4. Análisis multivariante

En primer lugar, se ha extendido el análisis de detección de valores atípicos a nivel multivariante evaluando cada observación a lo largo del conjunto de indicadores simples. Este

análisis se ha considerado necesario, ya que puede ocurrir que una observación sea atípica a nivel multivariante y no aparezca como tal en los análisis univariantes.

En segundo lugar, se ha realizado una transformación de los indicadores, ya que tal y como se ha visto en la sección anterior algunas de las distribuciones marginales son asimétricas. Mediante la familia de transformaciones Box-Cox se reduce el número de valores extremos y se consigue aumentar la simetría en las distribuciones, relaciones lineales entre los indicadores y normalidad multivariante (Peña, 2002).

Seguidamente, se ha aplicado una técnica multivariante de reducción de la dimensión con el fin de evaluar las relaciones existentes entre los indicadores simples definidos. La técnica estadística que se ha aplicado para identificar los grupos que se pueden formar con los indicadores simples, según el grado de correlación entre éstos, es el análisis factorial exploratorio seleccionando como método de factorización el análisis de componentes principales. Previamente, se analizan los supuestos que se deben cumplir para la aplicación del análisis factorial y, posteriormente, se exponen los resultados obtenidos.

Finalmente, se ha realizado un estudio para definir diferentes perfiles de profesorado según características comunes con respecto a los indicadores y su grado de dedicación a la actividad docente. Para ello, se ha aplicado un análisis de conglomerados en dos fases sobre los individuos.

6.2.4.1. Detección de valores atípicos a nivel multivariante

Un valor atípico se caracteriza por estar alejado del resto de valores y, por tanto, por tener una distancia con respecto al resto grande. Alternativamente, se puede definir un valor atípico como la distancia entre éste y el centro de los datos utilizando como medida la distancia euclídea. No obstante, esta medida de distancia no es razonable, puesto que puede ocurrir que la distancia euclídea de un valor observado al centro de los datos no sea grande, que ese valor no se haya detectado como dato atípico a nivel univariante y, sin embargo, que sí se trate de un dato atípico al verlo representado gráficamente junto con el conjunto de datos. En este caso, la distancia euclídea no tiene en cuenta la estructura de correlación de los datos.

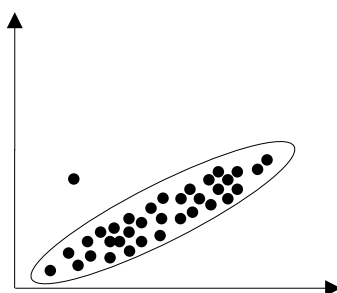


Figura 6.2. Representación de un valor atípico a nivel multivariante

Para solucionar el problema de correlación se estandarizan previamente los datos a nivel multivariante para que tengan media cero y matriz de covarianzas la identidad. De este modo, la distancia euclídea entre las variables incorreladas equivale a la distancia de Mahalanobis (Peña, 2002):

$$d_i = [(x_i - \bar{x})' S^{-1} (x_i - \bar{x})]^{1/2}$$

La distancia al cuadrado de Mahalanobis, d_i^2 , es una medida de la distancia de cada observación en un espacio multidimensional respecto del centro medio de las observaciones. Debido a que proporciona una medida común de centralidad multidimensional, también tiene propiedades estadísticas que tienen en cuenta las pruebas de significación.

Para comprobar si una observación se trata de un valor atípico se plantea un contraste de hipótesis en el que la hipótesis alternativa confirma la conjetura planteada. Para realizar el test se calcula la distancia al cuadrado de Mahalanobis que se distribuirá si H_0 es cierta, para muestras grandes, como una χ_p^2 al 99,9% de confianza siendo p el número de indicadores del estudio (Uriel, 1995; Hair et al., 2007).

Posteriormente al cálculo de la distancia al cuadrado de Mahalanobis para cada una de las observaciones, se han seleccionado aquellas que no cumplen que $d_i^2 < \chi_{11}^2$ siendo $\chi_{11}^2 = 31,2635$ con un nivel de significación de 0,001. La mayoría de las observaciones atípicas seleccionadas ya se habían detectado a nivel univariante excepto un 0,44% del total que son las que se han estudiado con más detenimiento. Se trata de profesores que presentan un conjunto de puntuaciones en los indicadores diferente al resto. Sin embargo, este comportamiento no es debido a errores de medición ni de transcripción. Si se asume la retención de estas observaciones en el estudio multivariante, se asume que pueden ser causantes de errores al aplicar las técnicas multivariantes. En el caso concreto del cálculo de las componentes principales la presencia de valores atípicos puede provocar un aumento en la varianza o una distorsión en las covarianzas. Por otra parte, si se eliminan los casos atípicos se limita la generalidad del estudio.

6.2.4.2. Transformaciones

Las transformaciones de los datos se pueden aplicar para corregir el incumplimiento de los supuestos estadísticos subyacentes a las técnicas multivariantes y para mejorar la correlación entre las variables (Hair et al., 2007). En el análisis de datos, las transformaciones se realizan con el fin de obtener datos más simétricos, más lineales y más constantes en varianza. Las transformaciones son monótonas, mantienen el orden de las observaciones y tienen el efecto de expansión y contracción de las distancias a las observaciones extremas, haciendo que la distribución del indicador sea más simétrica alrededor de su media.

La medida que indica si una variable presenta asimetría en su distribución es el coeficiente de asimetría κ calculado en el análisis descriptivo univariante. Los histogramas de cada uno de los indicadores proporcionan información adicional sobre la necesidad de aplicar una escala de transformación a aquellos indicadores claramente asimétricos. En el campo de los ICs se define un umbral de aproximadamente $|\kappa| > 1$ para aplicar una transformación sobre los indicadores simples (Annoni, 2010). Por tanto, siguiendo este criterio los indicadores sobre los que se ha realizado una transformación son IAD_{FORMA} , IAD_{TUT} , IAD_{EVA} , IAD_{OTRAS} , K_2 , K_4 .

Para ello, se han aplicado transformaciones pertenecientes a la familia Box-Cox que son un conjunto de transformaciones exponenciales que se suelen utilizar para datos asimétricos, incluyendo como casos particulares la transformación logarítmica, la raíz cuadrada y la inversa.

Considerando que $x_i > 0 \forall x_i \in X$, siendo X el indicador que se pretende transformar, la transformación sugerida por Box y Cox (1964) se define de la siguiente forma:

$$x_i \rightarrow T(x_i) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(x_i) & \text{si } \lambda = 0 \end{cases}$$

Se trata de un conjunto de transformaciones continuas, monótonas creciente, cóncavas si $\lambda < 1$ o convexas si $\lambda > 1$. La estimación del parámetro λ depende del valor del coeficiente de asimetría y, consecuentemente, del tipo de asimetría, positiva o negativa. Cuando $\lambda > 1$ la transformación produce una mayor separación o dispersión de los valores grandes del indicador, mientras que cuando $\lambda < 1$ los valores grandes tienden a concentrarse y los valores pequeños a dispersarse (Peña, 2002). Cuando se dispone de un conjunto de datos con un comportamiento similar o homogéneo se aconseja elegir un único parámetro de transformación λ que sea aplicable a todos los indicadores con el mismo tipo de asimetría (Annoni, 2010).

En este estudio no se ha podido encontrar un único parámetro de transformación para todos aquellos indicadores con el mismo tipo de asimetría debido a la naturaleza tan diversa de los indicadores. Por tanto, se han realizado las transformaciones con la elección de un parámetro adecuado para cada indicador individual.

Existen muchos métodos propuestos para la estimación de los valores λ (véase (Box y Cox, 1964)). En este caso, la estimación de los parámetros se ha obtenido mediante el software Minitab v.15 en el que se aplica un algoritmo que estima el valor de λ de tal forma que la transformación de la variable X , con $x_i > 0 \forall x_i \in X$, se define como sigue:

$$x_i \rightarrow T(x_i) = \begin{cases} x_i^\lambda & \text{si } \lambda \neq 0 \\ \log(x_i) & \text{si } \lambda = 0 \end{cases}$$

Con el fin de encontrar el valor óptimo de λ , el algoritmo de cálculo modifica el valor original de la variable mediante la ecuación:

$$x_i \rightarrow W(x_i) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda G^{\lambda-1}} & \text{si } \lambda \neq 0 \\ G \log(x_i) & \text{si } \lambda = 0 \end{cases}$$

donde x_i son los datos originales y G es la media geométrica de los datos. El proceso de optimización se basa en encontrar una estimación del parámetro λ que minimice la desviación estándar de W .

El parámetro de transformación utilizado, la media, la desviación típica y el coeficiente de asimetría de cada uno de los indicadores transformados se muestran en la siguiente tabla:

	Parámetro de transformación	Media	Desviación típica	Asimetría
$IAD_{FORMA_BOX-COX}$	$\lambda = 0,12$	1,26	0,32	-3,46
$IAD_{TUT_BOX-COX}$	$\lambda = 4,51$	16751	6888	0,02
$IAD_{EVA_BOX-COX}$	$\lambda = 0,22$	1,41	0,18	-0,03
$IAD_{OTRAS_BOX-COX}$	$\lambda = 0,08$	1,02	0,29	-2,88
$K_2_BOX-COX$	$\lambda = 5,00$	5,50	1,69	-1,49
$K_4_BOX-COX$	$\lambda = 5,00$	6,94	1,24	-1,98

Tabla 6.3. Parámetros λ empleados en las transformaciones Box-Cox

Para cada indicador transformado se obtiene un p-valor menor que 0,005 asociado a un estadístico Anderson-Darling muy grande, por lo que se concluye que ningún indicador transformado sigue una distribución normal, exceptuando el indicador IAD_{EVA} (datos no mostrados).

Otra de las ventajas de la aplicación de las transformaciones Box-Cox es conseguir más correlación entre los indicadores. Sin embargo, la diferencia obtenida entre las nuevas correlaciones y las correlaciones obtenidas con los datos originales es mínima (datos no mostrados).

Se recomienda aplicar una transformación sobre los datos en el caso de usar modelos donde se asuma normalidad. En este estudio se han transformado los datos para comprobar si se cumplen los supuestos subyacentes del análisis factorial como la normalidad, homocedasticidad y linealidad. Sin embargo, estos supuestos se pueden obviar desde el punto de vista estadístico, pero su incumplimiento produce una disminución en las correlaciones observadas (Hair et al., 2007). De hecho, es deseable que exista cierto grado de multicolinealidad para detectar las relaciones entre los indicadores.

A pesar de que las transformaciones se suelen usar en la construcción de indicadores compuestos, se debe tener en cuenta que cualquier transformación altera los datos originales y

puede cambiar la interpretación de los indicadores (Annoni y Kozovska, 2010). Por consiguiente, es conveniente reflexionar sobre la necesidad de aplicar una transformación a los indicadores.

En el caso de estudio el uso de transformaciones Box-Cox no ha aportado una mejora significativa en los resultados del análisis factorial. Por tanto, se ha considerado adecuado realizar los análisis estadísticos sobre la base de datos original.

6.2.4.3. Información agrupada con respecto a los indicadores individuales

Previamente a la aplicación del análisis factorial exploratorio para identificar las dimensiones subyacentes del bloque de indicadores que definen la actividad docente, se han contrastado algunos supuestos básicos para analizar la pertinencia de aplicación del análisis al conjunto de indicadores observados. En este análisis no se han incluido los indicadores K_3 e $IAD_{INFORMA}$ por tratarse de variables constante y categórica binaria, respectivamente. El análisis factorial se basa en las correlaciones entre indicadores y, en este caso, no es apropiado el cálculo de las correlaciones entre un indicador constante o binario y el resto de indicadores numéricos.

SUPUESTOS PARA LA EVALUACIÓN DEL GRADO DE MULTICOLINEALIDAD ENTRE LAS VARIABLES OBSERVADAS

Matriz de correlación

Se ha calculado la matriz de correlación entre los indicadores simples para detectar si existe multicolinealidad entre ellos. Un posible umbral para considerar la presencia de multicolinealidad, indicativo de la buena aplicación del análisis factorial, es una correlación entre los indicadores superior a 0,3 (Hair et al., 2007). Según la matriz de correlaciones, a simple vista no existe multicolinealidad entre los indicadores.

	IAD_{FORMA}	IAD_{GUIA}	IAD_{MAT}	IAD_{DOC}	IAD_{TUT}	IAD_{EVA}	IAD_{OTRAS}	K_1	K_2	K_4
IAD_{FORMA}	1									
IAD_{GUIA}	0,149	1								
IAD_{MAT}	0,126	0,162	1							
IAD_{DOC}	0,004	0,224	0,05	1						
IAD_{TUT}	0,097	0,21	0,24	0,334	1					
IAD_{EVA}	-0,067	-0,066	-0,005	0,545	0,274	1				
IAD_{OTRAS}	0,124	0,245	0,232	0,007	0,243	-0,044	1			
K_1	0,655	0,124	0,121	0,005	0,072	-0,058	0,118	1		
K_2	0,076	-0,006	0,027	-0,006	0,028	-0,046	-0,002	0,079	1	
K_4	0,117	0,051	0,065	0,082	0,092	0,032	-0,001	0,108	0,013	1

Tabla 6.4. Matriz de correlación de los indicadores simples

La relación más fuerte se produce entre los indicadores IAD_{FORMA} y K_1 . Esta correlación se debe a la definición de estos indicadores. El indicador IAD_{FORMA} está definido de tal forma que su valor se ve influenciado por el valor del indicador K_1 , que recoge la opinión de los estudiantes sobre la actividad docente del profesorado. Se considera que si un profesor tiene una buena evaluación por parte de sus alumnos, es decir, si obtiene una nota alta en la encuesta de opinión del alumnado sobre su actuación docente está totalmente cualificado para su posición y no necesita una preparación adicional para afrontar su actividad docente. En estos casos se le asigna la puntuación máxima al indicador IAD_{FORMA} .

También existe correlación, aunque más débil, entre IAD_{DOC} y los indicadores IAD_{TUT} , IAD_{EVA} , puesto que el volumen de docencia impartida depende de la capacidad docente de cada profesor y ésta última está directamente relacionada con la actividad del profesor relativa a las atenciones tutoriales llevadas a cabo y, consecuentemente, con las tareas o acciones de evaluación realizadas.

En general, se observa que no existe un alto grado de correlación entre los indicadores. A partir de la matriz de correlaciones se podría pensar que casi todos los indicadores definidos son imprescindibles, ya que cada uno de ellos aporta algo de información no recogida en el resto. Por consiguiente, se concluye que el conjunto de indicadores simples definidos por la UPV parece adecuado para evaluar las diferentes actividades que engloban la actividad docente del profesorado de acuerdo con el marco conceptual definido.

Contraste de esfericidad de Barlett

El contraste de esfericidad de Barlett tiene como objetivo comprobar la pertinencia de aplicación del análisis factorial a un grupo de variables observables. La hipótesis nula a contrastar es que todos los coeficientes de correlación teóricos entre cada par de variables son nulos (Uriel, 1995):

$$H_0: |R_p| = 1$$

$$H_1: |R_p| \neq 1$$

donde R_p se define como la matriz de correlación poblacional que recoge la relación existente entre cada par de variables. La diagonal principal de esta matriz está formada por unos, mientras que los elementos de fuera de la diagonal principal r_{jh} son coeficientes de correlación entre cada par de variables. Si todos los coeficientes r_{jh} son nulos la matriz R_p será igual a la matriz identidad con lo que su determinante será igual a la unidad.

El estadístico para contrastar la hipótesis se basa en la matriz de correlación muestral de los residuos llamada R .

$$\chi^2_{0,5(K^2-K)} = - \left[n - 1 - \frac{1}{6}(2K + 5) \right] \ln R$$

siendo K el número de variables del estudio y n el número de casos analizados.

En caso de que se acepte la hipótesis nula se debería abandonar el enfoque multivariante y aplicar el análisis de la varianza por separado a cada una de las variables dependientes. Al aplicar el contraste de esfericidad de Barlett al conjunto de indicadores del estudio, se obtiene un p-valor de 0,00 y un valor elevado para la Chi-cuadrado igual a 4412,329. Por tanto, se rechaza la hipótesis nula y se concluye que la aplicación del análisis factorial para el conjunto de indicadores del estudio es adecuada según este contraste de hipótesis.

La principal limitación del estadístico χ^2 es que está influenciado por el tamaño muestral, es decir, que a mayor número de casos analizados (a mayor n) el valor de la χ^2 tiende a aumentar, tal y como se puede observar en el estadístico definido, por lo que cuanto mayor sea la muestra más fácil será rechazar la hipótesis nula. Por tanto, el incremento del tamaño muestral da lugar a que la prueba de contraste de Barlett sea más sensible a la detección de correlaciones entre los indicadores (Hair et al., 2007).

Matriz de correlaciones anti-imagen

Una alternativa al contraste de Barlett que permite analizar la adecuación del análisis factorial es comprobar si las correlaciones parciales son bajas con el fin de asegurarse que existen factores subyacentes al conjunto de indicadores. Para ello, se ha calculado la matriz de correlaciones anti-imagen que es simplemente el valor negativo de la correlación parcial.

	IAD_{FORMA}	IAD_{GUIA}	IAD_{MAT}	IAD_{DOC}	IAD_{TUT}	IAD_{EVA}	IAD_{OTRAS}	K_1	K_2	K_4
IAD_{FORMA}	0,553									
IAD_{GUIA}	-0,063	0,549								
IAD_{MAT}	-0,031	-0,068	0,732							
IAD_{DOC}	0,012	-0,280	0,014	0,541						
IAD_{TUT}	-0,044	-0,101	-0,176	-0,177	0,720					
IAD_{EVA}	0,032	0,229	0,026	-0,528	-0,167	0,517				
IAD_{OTRAS}	-0,026	-0,183	-0,154	0,076	-0,193	0,028	0,660			
K_1	-0,636	-0,019	-0,038	-0,005	0,011	0,014	-0,036	0,549		
K_2	-0,030	0,031	-0,014	-0,019	-0,035	0,055	0,019	-0,039	0,649	
K_4	-0,056	-0,009	-0,038	-0,050	-0,056	0,010	0,041	-0,040	0,000	0,780

Tabla 6.5. Matriz de correlaciones anti-imagen de los indicadores simples

Las correlaciones anti-imagen elevadas son indicativas de una matriz de datos que no es adecuada para el análisis factorial (Hair et al., 2007). En este caso, se observa que la mayoría de las correlaciones anti-imagen son menores en valor absoluto que 0,3, por tanto, se concluye que el análisis factorial es apropiado teniendo en cuenta este supuesto.

Los valores de la diagonal principal de la matriz de correlaciones anti-imagen son los valores MSA (Medida de Suficiencia de Muestreo) para los indicadores individuales. Según Hair et al. (2007) se deben excluir del análisis factorial aquellos indicadores que caen en la gama de inaceptables, es decir, aquellos indicadores con valor MSA por debajo de 0,5. En este caso, todos los valores de la diagonal principal son mayores que 0,5. Una vez que los indicadores individuales logran un nivel aceptable, se puede calcular el MSA general o también llamado coeficiente KMO para tomar una decisión sobre la aplicación del análisis factorial.

Coeficiente KMO (Kaiser-Meyer-Olkin)

El coeficiente Kaiser-Meyer-Olkin es una medida de adecuación muestral al análisis factorial basada en los coeficientes de correlación observables y parciales entre variables originales (Uriel, 1995):

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

donde r_{jh} son los coeficientes de correlación observados entre las variables originales y a_{jh} son los coeficientes de correlación parcial entre variables originales.

Un coeficiente KMO muy próximo a la unidad indica que existe multicolinealidad entre las variables y que, por tanto, es adecuada la aplicación del análisis factorial. Según indica Kaiser (1974), un coeficiente KMO por debajo de 0,5 indica que no es adecuado aplicar un análisis factorial sobre los datos.

En el estudio el coeficiente KMO resultante tiene un valor igual a 0,58. Este resultado refleja la presencia de multicolinealidad entre las variables, aunque no es muy fuerte.

En general, y a pesar de obtener valores cercanos al límite de los umbrales de aceptación predefinidos en cada uno de los supuestos, la mayoría de los test de verificación de la aplicación del análisis factorial se cumplen. Por tanto, se ha decidido aplicar el análisis factorial con el fin de obtener resultados provechosos.

RESULTADOS DE LA APLICACIÓN DEL ANÁLISIS FACTORIAL

En primer lugar, se ha aplicado el análisis factorial, seleccionando como método de factorización el análisis de componentes principales sin ningún tipo de rotación, sobre las siguientes bases de datos:

- Base de datos original
- Base de datos original tras la eliminación de las observaciones atípicas detectadas a nivel multivariante
- Base de datos formada por indicadores transformados mediante la familia de transformaciones Box-Cox

El objetivo de aplicar el análisis factorial sobre la base de datos una vez eliminadas las observaciones atípicas es analizar el impacto que tienen sobre el análisis y, consecuentemente, sobre los resultados obtenidos. Tras la eliminación de estas observaciones, se han obtenido unos resultados prácticamente iguales al análisis previo (datos no mostrados). Por ello, se ha considerado interesante mantener todas las observaciones detectadas como atípicas en el análisis con el fin de no perder la generalidad de los resultados.

En la sección anterior se ha comprobado que las transformaciones Box-Cox sobre los indicadores asimétricos han permitido obtener una distribución del conjunto de indicadores un poco más aproximada, aunque no lo suficiente, a la distribución normal multivariante. Sin embargo, no se ha podido conseguir una transformación única para todos los indicadores al aplicar las transformaciones Box-Cox y se ha comprobado que los resultados obtenidos en el análisis factorial con los indicadores transformados no difieren de los resultados obtenidos con la base de datos original (datos no mostrados). Además, según Hair et. al (2007), los supuestos de normalidad se pueden obviar desde el punto de vista estadístico para la aplicación del análisis factorial siendo conscientes de que su incumplimiento produce una disminución de las correlaciones observadas. En este caso las correlaciones observadas entre los indicadores transformados no aportan una mejora notable.

Finalmente y por las razones anteriores, se ha decidido aplicar el análisis factorial sobre la base de datos original.

Rotación de factores

Para mejorar la interpretación de los resultados a partir de la solución inicial se han aplicado procedimientos de rotación de factores. Mediante los factores rotados se trata que cada uno de los indicadores tenga una correlación lo más próxima posible a la unidad con uno de los factores y próxima a cero con el resto. Por tanto, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto y, de esta forma, se facilita la interpretación de los factores.

Existen dos formas básicas de realizar la rotación de factores: rotación ortogonal y rotación oblicua. En la primera los ejes se rotan de forma que quede preservada la incorrelación entre los factores. Entre los diferentes procedimientos de rotación ortogonal el método Varimax normalizado es el más aplicado y conocido, y se obtiene maximizando la suma de varianzas de las cargas factoriales al cuadrado dentro de cada factor una vez normalizadas con la normalización de Kaiser. En la rotación oblicua los ejes no son ortogonales con lo que se pierde

una propiedad deseable para los factores, pero es un método interesante si se consigue una asociación más nítida de cada indicador con el factor correspondiente. El método de rotación oblicua más usado es el denominado Oblimin con la normalización de Kaiser (Uriel, 1995).

En este estudio se han aplicado los dos tipos de rotación con el fin de mejorar la interpretación de los factores. Los resultados muestran una mejora notable de la interpretación con las dos técnicas de normalización. Sin embargo, con la rotación oblicua se obtienen factores muy poco correlacionados con lo que finalmente se aplica una rotación Varimax con un porcentaje de varianza explicado por las cinco primeras componentes del 70,55%.

Extracción del número de componentes principales

A continuación se muestra la tabla de autovalores calculados con sus correspondientes porcentajes de varianza explicada.

Componente	Autovalores iniciales		
	Total	% de la varianza	% acumulado
1	2,112	21,118	21,118
2	1,751	17,514	38,632
3	1,246	12,459	51,091
4	0,990	9,896	60,987
5	0,957	9,566	70,553
6	0,872	8,723	79,276
7	0,740	7,401	86,677
8	0,609	6,090	92,767
9	0,380	3,795	96,562
10	0,344	3,438	100,000

Tabla 6.6. Valores propios de la matriz de correlaciones y porcentaje de varianza explicada por cada uno de los factores

Criterio de Kaiser o de la raíz latente

Según este criterio se retienen los tres primeros factores, puesto que son aquellos que tienen un autovalor que excede de la unidad que es el valor medio del conjunto de todos los autovalores cuando se consideran las variables tipificadas. El cuarto y el quinto autovalor son ligeramente inferiores a la unidad (0,990 y 0,957, respectivamente), con lo que también se podrían considerar en el estudio.

Criterio de contraste de caída o test del codo de Castell

El gráfico de sedimentación muestra una curva pronunciada hasta la tercera componente. A partir de la tercera componente el paso entre componentes es casi horizontal. Por tanto, mediante este criterio se deben retener también tres factores para explicar la estructura subyacente de los datos.

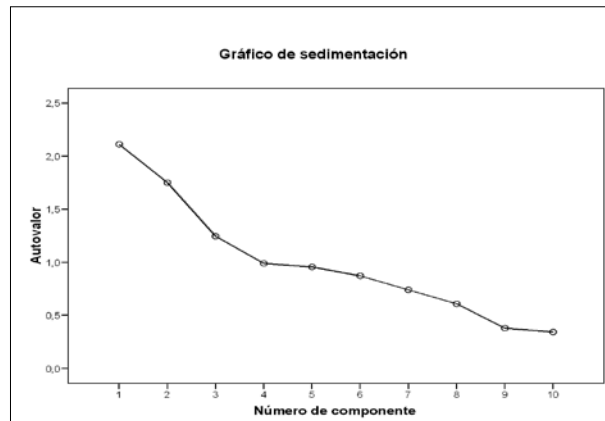


Figura 6.3. Gráfico de sedimentación

Criterio del porcentaje de varianza explicada

Se basa en retener aquellos factores hasta obtener un porcentaje acumulado específico de la varianza total extraída. En este caso si se retienen tres componentes se explica el 51,09% de la varianza, mientras que al retener una componente más el porcentaje aumenta al 60,98% siendo el 60% uno de los umbrales predefinidos más utilizados en ciencias sociales. Por tanto, mediante este criterio se seleccionan cuatro factores.

En la figura de la varianza acumulada se observa que las primeras componentes son las que mayor porcentaje de variabilidad explican de los datos. A partir de la cuarta componente el incremento en el porcentaje de varianza acumulada es cada vez más moderado.

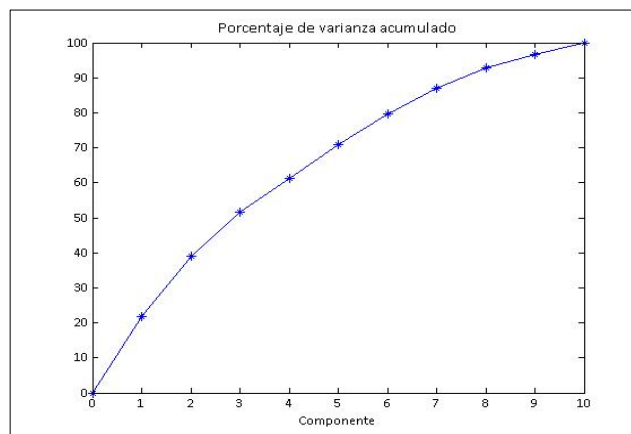


Figura 6.4. Porcentaje de varianza acumulado

Finalmente, se ha decidido trabajar con cinco factores con un porcentaje de varianza explicado del 70,55%. Aunque son los tres primeros factores los que más porcentaje de variabilidad explican de los datos, los últimos dos factores explican al menos la varianza

contenida en una variable. Se han considerado estas dos últimas componentes para el estudio puesto que se quiere obtener una clasificación de todos los indicadores en los distintos factores.

Matriz de cargas factoriales

En la tabla 6.7 se muestran las cargas factoriales de cada uno de los indicadores en los componentes retenidos:

	Componente				
	1	2	3	4	5
IAD_{FORMA}	-0,014	0,898	0,113	0,053	0,031
IAD_{GUIA}	0,101	0,142	0,609	-0,023	-0,127
IAD_{MAT}	-0,016	0,022	0,634	0,173	0,127
IAD_{DOC}	0,853	0,020	0,103	0,033	-0,024
IAD_{TUT}	0,515	0,022	0,524	0,090	0,101
IAD_{EVA}	0,855	-0,055	-0,123	-0,009	-0,027
IAD_{OTRAS}	-0,076	0,054	0,735	-0,120	-0,030
K_1	-0,013	0,902	0,084	0,039	0,037
K_2	-0,013	0,062	-0,006	-0,009	0,980
K_4	0,043	0,081	0,027	0,976	-0,011

Tabla 6.7. Método de extracción: PCA. Método de rotación: Normalización Varimax con Kaiser

En la matriz de cargas factoriales se observa que cuatro de los factores contienen variables con altas saturaciones. Estos corresponden al primer, segundo, cuarto y quinto factor. Por otra parte, el tercer factor está formado por más indicadores que el resto, pero sus cargas factoriales no son tan significativas dentro del factor.

El primer factor está formado por los indicadores IAD_{DOC} , IAD_{EVA} y IAD_{TUT} , aunque este último no contribuye tanto como los demás a la formación del eje y, además, tiene una carga factorial mayor en otra componente. Este factor representa el *Volumen de actividad docente del profesorado*, ya que los indicadores que lo forman recogen la actividad del docente correspondiente a las acciones de enseñanza, a las atenciones tutoriales llevadas a cabo y a las tareas o acciones de evaluación realizadas.

Por otro lado, el segundo factor se interpreta como los *Resultados de la evaluación del profesorado*, puesto que los indicadores que están más correlacionados con este factor son K_1 e IAD_{FORMA} , siendo el primero de ellos el indicador que más contribuye a la formación de este eje. El cálculo de las puntuaciones de los indicadores IAD_{FORMA} y K_1 depende de la nota obtenida por el profesorado en la encuesta de opinión del alumnado. Por consiguiente, la dimensión de *Resultados* depende directamente de la opinión del alumnado sobre la actuación docente del profesor.

El tercer factor mide el *Volumen de actividades adicionales al desempeño habitual de la actividad docente*. Los indicadores que contribuyen a la creación de este eje son IAD_{MAT} ,

IAD_{GUIA} , IAD_{OTRAS} y en menor medida IAD_{TUT} . El indicador IAD_{MAT} refleja la actividad del profesor relativa a la elaboración de materiales para la docencia, IAD_{GUIA} engloba los aspectos relativos a la organización docente y a la planificación de la enseñanza, mientras que IAD_{OTRAS} recoge otras actividades de carácter docente que desarrolla el profesor. Estos indicadores no están directamente relacionados con el volumen docente asignado a cada profesor, por tanto, se clasifican en otro bloque distinto.

El indicador IAD_{TUT} contribuye, debido a su definición, tanto al primer eje representado por el *Volumen de actividad docente* como al tercer eje que mide las *Actividades adicionales a la actividad docente habitual*. Por una parte, pertenece al primer eje por medir la actividad relativa a las atenciones tutoriales según el contrato del profesor. Por otra parte, también pertenece al tercer eje al tener en cuenta la participación del profesorado en el Plan de Acción Tutorial Universitario (PATU) y en el Programa de Formación Inicial Pedagógica para el Profesorado Universitario (FIPPU) que son actividades adicionales a la actividad docente habitual.

El indicador K_4 , que recoge las actividades de *Gestión de la actividad docente*, es el que forma el cuarto factor. Por último, K_2 es el único indicador que contribuye a la creación del quinto factor que representa el *Rendimiento académico* por parte de los estudiantes. Se ha forzado la obtención de estos dos ejes para saber si las variables K_2 y K_4 contribuían a la formación de un mismo eje y para poder clasificarlas dentro de un factor.

La clasificación de los indicadores obtenida mediante el análisis factorial y la clasificación que proporciona el modelo DOCENTIA se basan en fundamentos totalmente diferentes. Las dimensiones descritas en el programa DOCENTIA se han definido atendiendo a los procesos que se llevan a cabo en la universidad. En primer lugar, se planifica la docencia que se va a impartir; posteriormente, se desarrolla la enseñanza y se evalúa el aprendizaje; y finalmente, se obtienen los resultados en términos de objetivos formativos logrados por los estudiantes de modo que desde dichos resultados se inicia un nuevo ciclo de formación. Por otra parte, la clasificación obtenida mediante el análisis factorial es coherente con la estructura de correlación que existe entre los indicadores. Por esta razón, las agrupaciones de los indicadores obtenidas mediante esta técnica son diferentes a las que se proponen con el enfoque DOCENTIA.

ANÁLISIS DE CONSISTENCIA INTERNA

El coeficiente alpha de Cronbach analiza la fiabilidad de un conjunto de indicadores basándose en la correlación inter-elementos promedio. Si las varianzas de los indicadores son muy diferentes entre sí se deben normalizar los datos mediante la normalización z-score con el objetivo de obtener igualdad en las varianzas (Nardo et al., 2008).

El acuerdo general sobre el límite inferior para este coeficiente es de 0,7, aunque puede bajar a 0,6 en la investigación exploratoria (Hair et al., 2007). Al calcular el coeficiente alpha de

Cronbach sobre el conjunto de indicadores estandarizados se obtiene un valor de 0,546. En este caso, el coeficiente alpha de Cronbach es relativamente bajo lo que denota que las actividades que se definen para la medición de la evaluación docente del profesorado son bastante diferentes entre sí.

Dimensión	Alpha de Cronbach
Volumen de actividad docente	0,706
Resultados	0,791
Actividades adicionales a la actividad docente habitual	0,533

Tabla 6.8. Coeficiente alpha de Cronbach para cada constructo del análisis factorial

La consistencia interna de las primeras dos dimensiones es aceptable, mientras que el coeficiente alpha de Cronbach en la tercera dimensión es relativamente bajo (véase tabla 6.8). Este resultado se debe a que la tercera dimensión engloba actividades adicionales a la docencia habitual muy diferentes entre sí (p. ej. elaboración de materiales y recursos para la docencia, participación en los programas FIPPU y PATU, dirección de proyectos final de carrera, asistencia a congresos docentes con presentación de comunicación, etc.).

En la tabla 6.9 se muestra el coeficiente alpha de Cronbach obtenido con el conjunto de indicadores tras la eliminación de cada uno de ellos.

Indicador excluído	Coeficiente alpha de Cronbach
IAD_{FORMA}	0,496
IAD_{GUIA}	0,510
IAD_{MAT}	0,516
IAD_{DOC}	0,499
IAD_{TUT}	0,471
IAD_{EVA}	0,548
IAD_{OTRAS}	0,523
K_1	0,501
K_2	0,573
K_4	0,548

Tabla 6.9. Coeficiente alpha de Cronbach tras la eliminación de cada uno de los indicadores

Los indicadores K_2 y K_4 hacen aumentar moderadamente el coeficiente alpha de Cronbach tras su eliminación. Este resultado refleja que los dos indicadores están explicados por un constructo muy diferente al resto. Por otro lado, IAD_{FORMA} es el indicador que hace disminuir en mayor medida el coeficiente tras su eliminación lo cual indica que este indicador es importante para el estudio.

En general, el coeficiente alpha de Cronbach sigue siendo muy bajo al eliminar cada uno de los indicadores y, además, no existe ningún indicador que tras su eliminación haga aumentar de forma considerable el coeficiente. Por tanto, se concluye que todos los indicadores son, en mayor o menor medida, imprescindibles para el estudio. En este sentido, la selección del conjunto de indicadores simples realizada por la UPV es adecuada para medir la actividad docente del profesorado.

6.2.4.4. Información agrupada con respecto a las unidades de análisis

El objetivo del análisis cluster o de conglomerados aplicado a este estudio es identificar diferentes perfiles de profesorado según los indicadores medidos.

Se han realizado dos tipos de análisis cluster, un primer análisis considerando los once indicadores (excluyendo el indicador constante K_3) y un segundo análisis en el que además se ha analizado la capacidad docente asignada a cada profesor, según su tipo de contrato y las reducciones debidas a la investigación y gestión académica. Este segundo análisis se ha considerado interesante, puesto que refleja las diferencias entre la actividad docente que cada profesor debería impartir según su capacidad docente asignada y la que realmente imparte.

La técnica aplicada es el análisis de conglomerados en dos fases. Se trata de una herramienta de exploración diseñada para descubrir las agrupaciones o conglomerados naturales de un conjunto de datos que de otra manera no sería posible detectar (Pérez, 2005). Es un método de agrupación que combina iterativamente los procedimientos jerárquico y no jerárquico de conglomeración permitiendo la creación de modelos de conglomerados basados al mismo tiempo en variables categóricas y continuas, y diseñado para manejar conjuntos de datos de gran tamaño.

El algoritmo que emplea este procedimiento incluye varias atractivas funciones que lo hacen diferente de las técnicas de conglomeración tradicionales (Pérez, 2005):

- **Tratamiento de variables categóricas y continuas.** Al suponer que las variables son aproximadamente independientes, es posible aplicar una distribución normal multinomial conjunta en las variables continuas y categóricas
- **Selección automática del número de conglomerados.** Mediante la comparación de los valores de un criterio de selección del modelo para diferentes soluciones de conglomeración, el procedimiento puede determinar automáticamente el número óptimo de conglomerados. También existe la opción de seleccionar el número de conglomerados manualmente
- **Escalabilidad.** Mediante la construcción de un árbol de características de conglomerados que resume los registros, el algoritmo en dos fases puede analizar archivos de datos de gran tamaño

El algoritmo se compone de dos pasos: 1) conglomerar previamente los casos (o registros) en muchos subconglomerados pequeños (procedimiento no jerárquico) y 2) conglomerar los subconglomerados resultantes del paso de conglomeración previo en el número de casos deseado (procedimiento jerárquico).

La medida de distancia utilizada en este método de agrupación que determina cómo se calcula la similitud entre dos conglomerados puede ser la distancia log-verosimilitud o la euclídea:

- **Log-verosimilitud.** La medida de la verosimilitud realiza una distribución de probabilidad entre las variables. Para la aplicación de esta medida se supone que las variables del modelo de conglomerados son aproximadamente independientes. Además, se supone que cada variable continua sigue una distribución normal y que cada variable categórica sigue una distribución multinomial. Las comprobaciones empíricas internas indican que este procedimiento es bastante robusto frente a las violaciones tanto del supuesto de independencia como de las distribuciones

La medida de la verosimilitud se aplica cuando se dispone de variables tanto categóricas como continuas. La distancia entre dos clusters se define como el decrecimiento en verosimilitud al combinarse en un solo cluster (Chiu et al., 2001). Analíticamente, la distancia entre dos clusters i y j se define como:

$$d(i, j) = \tau_i + \tau_j - \tau_{\langle i, j \rangle} \text{ donde,}$$

$$\tau_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right)$$

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v} \text{ para } v = i, j, \langle i, j \rangle$$

siendo:

K^A : número de variables continuas del estudio

K^B : número de variables categóricas del estudio

L_k : número de categorías para la k -ésima variable categórica

N_v : número de observaciones en el cluster v

$\hat{\sigma}_k^2$: varianza estimada de la variable continua k -ésima

$\hat{\sigma}_{vk}^2$: varianza estimada de la variable continua k -ésima en el cluster v

N_{vkl} : número de observaciones en el cluster v cuya variable categórica k -ésima pertenece a la l -ésima categoría

$d(i, j)$: distancia entre los clusters i, j

$\langle i, j \rangle$: índice que representa el cluster formado por la combinación de los clusters i, j

Si $\hat{\sigma}_k^2$ se ignora en la expresión τ_v , la distancia entre los clusters i, j sería exactamente el decrecimiento en log-verosimilitud al combinarse los clusters. El término $\hat{\sigma}_{vk}^2$ se añade para resolver el problema cuando $\hat{\sigma}_{vk}^2 = 0$ (puede ocurrir cuando un cluster presenta un único caso), puesto que el logaritmo natural en este caso no está definido.

- **Euclídea.** La medida euclídea es la distancia según una "línea recta" entre dos conglomerados. Sólo se puede utilizar cuando todas las variables son continuas

El procedimiento genera criterios de información según el número de conglomerados de la solución, las frecuencias de los conglomerados para la conglomeración final y los estadísticos descriptivos por conglomerado para la agrupación final. Se puede especificar tanto el **criterio de información bayesiano (BIC)** como el **criterio de información de Akaike (AIC)**.

El criterio BIC trata de seleccionar el modelo correcto con máxima probabilidad a posteriori. Es un criterio consistente y selecciona modelos con menor número de parámetros que el criterio AIC. Por otro lado, el criterio AIC no pretende seleccionar el modelo correcto, ya que admite que este modelo verdadero puede no estar entre los estimados. Este criterio trata de obtener el modelo que proporcione mejores predicciones entre los existentes (Peña, 2002).

Una vez seleccionado el número de conglomerados del modelo, se puede analizar la importancia de cada una de las variables para la formación de cada cluster. La importancia de una variable continua k para la formación del cluster j se mide mediante el estadístico t bajo la hipótesis nula que indica que si t se distribuye como una t-Student con $N_k - 1$ grados de libertad, se considera que la variable k es significativa para la formación del cluster.

$$t = \frac{\hat{\mu}_k - \hat{\mu}_{jk}}{\hat{\sigma}_{jk} / \sqrt{N_k}}$$

siendo:

$\hat{\mu}_k$: media estimada de la variable continua k

$\hat{\mu}_{jk}$: media estimada de la variable continua k en el j -ésimo cluster

$\hat{\sigma}_{jk}$: desviación típica estimada de la variable continua k en el j -ésimo cluster

N_k : número de observaciones del k -ésimo cluster

Cuando la variable es categórica la importancia para la contribución a la formación del cluster viene dada por el estadístico χ^2 bajo la hipótesis nula de que si este estadístico sigue una distribución Chi-cuadrado con L_k grados de libertad la variable será significativa.

$$\chi^2 = \sum_{l=1}^{L_k} \left(\frac{N_{jkl} - N_{kl}}{N_{kl}} \right)^2$$

siendo:

L_k : número de categorías para la k -ésima variable categórica

N_{jkl} : número de observaciones en el cluster j cuya variable categórica k -ésima pertenece a la l -ésima categoría

N_{kl} : número de observaciones cuya variable categórica k -ésima pertenece a la l -ésima categoría

Para el estudio se ha considerado adecuado aplicar el análisis de conglomerados en dos fases con la distancia log-verosimilitud, puesto que permite trabajar con indicadores numéricos y categóricos al mismo tiempo sobre un conjunto de datos grande. Se ha elegido el criterio bayesiano BIC para la selección del número de conglomerados, puesto que es el más utilizado siempre que no se aumente la complejidad del modelo verdadero con el tamaño del conjunto de datos (Burnham y Anderson, 2002).

Análisis cluster incluyendo la variable 'capacidad docente'

En primer lugar, se muestran los resultados obtenidos del análisis de conglomerados al incluir la variable capacidad docente (CAPDOC). En la tabla 6.10 se exponen los resultados de la agrupación automática en los que aparece el criterio de selección elegido para cada posible modelo, según el número de clusters:

Número de conglomerados	Criterio bayesiano de Schwarz (BIC)	Cambio en BIC	Razón de cambios en BIC	Razón de medidas de distancia
1	23133,242			
2	20246,654	-2886,588	1,000	1,872
3	18789,036	-1457,618	0,505	1,305
4	17714,315	-1074,721	0,372	1,294
5	16924,791	-789,525	0,274	1,16
6	16269,161	-655,629	0,227	1,656
7	15945,331	-323,831	0,112	1,323
8	15744,788	-200,543	0,069	1,106
9	15580,981	-163,808	0,057	1,08
10	15442,662	-138,319	0,048	1,001
11	15304,577	-138,085	0,048	1,163
12	15211,195	-93,382	0,032	1,058
13	15132,786	-78,409	0,027	1,011
14	15057,206	-75,58	0,026	1,028
15	14988,641	-68,564	0,024	1,076

Tabla 6.10. Criterio bayesiano para la selección del número de conglomerados con los indicadores simples y la capacidad docente

Inicialmente el número de conglomerados adecuado es el que tiene un menor valor BIC. No obstante, el valor BIC continua decreciendo conforme el número de grupos se incrementa.

Para evitar este problema, se evalúa el cambio en BIC al pasar de un número de conglomerados a otro y el cambio en la medida de la distancia para determinar la mejor solución. Como consecuencia, una buena solución para la selección del número de clusters será aquella que tenga un elevado valor en la razón de cambios en BIC y un elevado valor en la razón de medidas de distancia, puesto que proporcionará una división significativa entre los diferentes grupos formados.

Aplicando estas conclusiones al estudio, se ha elegido un modelo con seis conglomerados para la identificación de los perfiles de profesorado. Este modelo presenta un valor BIC menor

que los modelos con menor número de conglomerados, pero mayor que los modelos con más conglomerados por la razón explicada anteriormente. Por tanto, la justificación de su elección viene dada por su elevado valor en la razón de cambios en BIC y en la razón de medidas de distancia frente al resto de modelos con mayor número de clusters.

La distribución de las observaciones dentro de los seis conglomerados y la importancia de cada uno de ellos queda dividida de la siguiente manera:

Conglomerado	N	% del total
1	323	12,00%
2	280	10,40%
3	323	12,00%
4	608	22,60%
5	774	28,80%
6	381	14,20%
Combinados	2689	100,00%
Total	2689	100,00%

Tabla 6.11. Distribución de las observaciones en los seis conglomerados

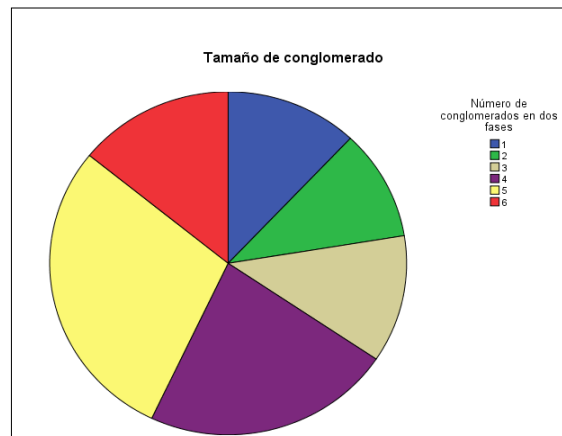


Figura 6.5. Gráfico de sectores con la distribución de las observaciones en los seis conglomerados

En los resultados obtenidos se observa que un 28,80% del total de las observaciones pertenece al conglomerado 5 que es el más grande, mientras que un 10,40% de las observaciones se clasifica en el conglomerado más pequeño que corresponde al conglomerado número 2.

No hay ningún conglomerado que contenga más del 50% de la información de las observaciones según los indicadores del análisis, ni tampoco ningún conglomerado formado por un porcentaje de observaciones muy pequeño con respecto al total. Este resultado indica que la elección del número de conglomerados es bastante adecuada mediante el criterio BIC.

La tabla de frecuencias 6.12 muestra las propiedades de los clusters según el indicador $IAD_{INFORMA}$, es decir, según si el profesor informa o no del horario de tutorías desde el inicio del curso académico.

Conglomerado	$IAD_{INFORMA}$			
	0		1	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
1	0	0,00%	323	14,50%
2	77	16,80%	203	9,10%
3	0	0,00%	323	14,50%
4	0	0,00%	608	27,30%
5	0	0,00%	774	34,70%
6	381	83,20%	0	0,00%
Combinados	458	100,00%	2231	100,00%

Tabla 6.12. Distribución de las observaciones con respecto al indicador binario en los seis conglomerados

Los clusters 1, 3, 4 y 5 están formados exclusivamente por profesores que sí informan del horario de tutorías desde el inicio del curso académico. Por otra parte, el cluster 6 está formado exclusivamente por profesores que no informan sobre el horario de tutorías.

El 27,50% de los profesores que definen el segundo cluster no informan sobre el plan de tutorías mientras que el 72,50% de los profesores de este conglomerado sí lo hacen.

El análisis de la contribución a la formación de los clusters se ha realizado para todos los indicadores pero, a modo de ejemplo, se va a mostrar la contribución de dos de ellos. En las figuras que se muestran a continuación los indicadores se sitúan en el eje de ordenadas de forma decreciente en relación a su importancia con respecto a la formación del cluster. Las líneas discontinuas verticales muestran los valores críticos que determinan la significatividad de cada indicador en la formación del cluster. Por tanto, un indicador será significativo si el estadístico t de Student excede la línea de guiones positiva o negativa.

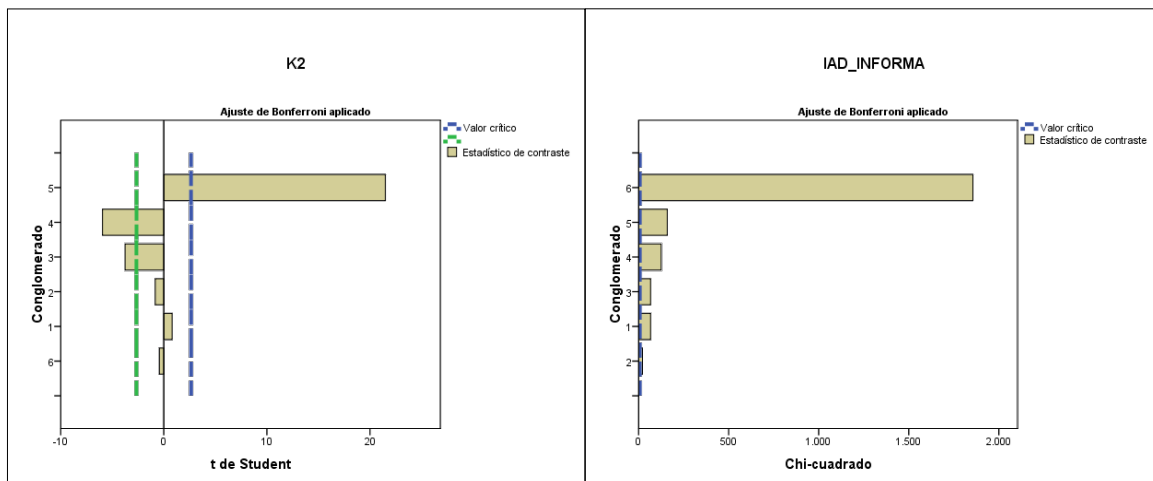


Figura 6.6. Contribución de los indicadores K_2 y $IAD_{INFORMA}$ a la formación de los conglomerados

Un valor negativo de la t indica que la variable toma un valor en el cluster inferior a su media, mientras que un valor positivo indica lo contrario. Se observa que el indicador K_2 contribuye en gran medida y de forma positiva a la formación del quinto conglomerado, mientras que contribuye en menor grado y de forma negativa a la formación de los conglomerados 3 y 4.

Para el indicador $IAD_{INFORMA}$ se ha realizado el contraste de la Chi-cuadrado obteniendo como resultado una contribución en gran medida y de forma positiva a la formación del cluster número 6. La contribución de este indicador a los demás clusters sigue siendo positiva y significativa pero en menor grado.

La tabla de centroides muestra que los clusters están bien diferenciados entre ellos con respecto a los indicadores de tipo continuo.

		Centroides						
Indicador	Estadísticos	Conglomerado						Combinados
		1	2	3	4	5	6	
CAPDOC	Media	29,19	8,52	20,88	18,96	18,89	19,29	19,36
	Desv. típica	5,39	2,01	5,51	4,22	3,78	5,72	6,67
IAD_{FORMA}	Media	14,54	12,00	4,93	14,84	14,76	12,09	12,90
	Desv. típica	1,71	5,36	4,25	0,99	1,27	5,17	4,46
IAD_{GUIA}	Media	8,36	3,05	4,73	7,24	5,82	5,24	5,95
	Desv. típica	6,26	2,77	4,27	4,52	4,38	4,50	4,78
IAD_{MAT}	Media	8,14	5,64	7,22	12,00	5,75	6,82	7,77
	Desv. típica	4,13	1,99	3,54	3,83	1,59	3,20	3,92
IAD_{DOC}	Media	25,42	10,30	16,99	15,40	15,79	15,92	16,45
	Desv. típica	5,68	4,66	5,78	4,67	4,76	6,37	6,43
IAD_{TUT}	Media	8,87	5,21	8,76	8,90	8,72	8,71	8,41
	Desv. típica	0,48	0,85	0,43	0,58	0,39	0,41	1,21
IAD_{EVA}	Media	10,00	2,93	6,74	4,89	5,17	5,56	5,69
	Desv. típica	4,36	1,60	3,46	2,32	2,33	3,18	3,40
IAD_{OTRAS}	Media	6,48	1,68	4,01	9,23	3,69	5,10	5,31
	Desv. típica	5,98	2,35	4,73	7,35	3,04	5,77	5,75
K_1	Media	1,26	1,21	1,07	1,28	1,26	1,21	1,23
	Desv. típica	0,09	0,13	0,07	0,09	0,09	0,12	0,12
K_2	Media	1,45	1,44	1,42	1,41	1,48	1,44	1,44
	Desv. típica	0,09	0,11	0,12	0,12	0,04	0,10	0,10
K_4	Media	1,48	1,45	1,45	1,48	1,47	1,46	1,47
	Desv. típica	0,05	0,09	0,08	0,04	0,06	0,07	0,07

Tabla 6.13. Tabla de centroides

Por ejemplo, el primer cluster se diferencia del resto por estar formado por aquellos profesores que tienen asignada mucha carga docente CAPDOC y, por tanto, dedican muchas horas al desarrollo de la actividad docente, IAD_{DOC} , y a las tareas de evaluación IAD_{EVA} . También forma parte de este cluster el profesorado con una alta puntuación en la variable IAD_{GUIA} por la realización de la guía docente de las asignaturas.

Para interpretar mejor estos resultados se han obtenido intervalos de confianza al 95% para las medias de los indicadores cuantitativos y para el porcentaje del indicador $IAD_{INFORMA}$ dentro de cada conglomerado. Sobre estas figuras se han marcado en rojo los intervalos de confianza para las medias de aquellos conglomerados en los que el indicador contribuye significativamente a su formación, según los gráficos de contribución explicados anteriormente. Los estadísticos marcados en rojo son los que hay que tener en cuenta para la búsqueda de los perfiles de profesorado.

Las figuras 6.7-6.18 muestran los resultados agrupados según las variables latentes obtenidas mediante la técnica de análisis de componentes principales, con el objetivo de encontrar tendencias en los grupos de profesores de forma más general.

El indicador CAPDOC es la carga docente asignada a cada profesor y está muy correlacionado con los indicadores correspondientes al volumen de actividad docente, por tanto, se ha considerado en este grupo para la interpretación de los distintos perfiles.

	IAD_{FORMA}	IAD_{GUIA}	$IAD_{INFORMA}$	IAD_{MAT}	IAD_{DOC}	IAD_{TUT}	IAD_{EVA}	IAD_{OTR}	K_1	K_2	K_4
CAPDOC	0,005	0,237	0,126	0,084	0,59	0,535	0,541	0,119	-0,011	-0,014	0,071

Tabla 6.14. Correlación del indicador CAPDOC con el resto de indicadores simples

Una vez definido el modelo con seis conglomerados se ha asociado a cada uno de ellos, en el caso en que ha sido posible la identificación, una categoría de profesorado. En la tabla 6.15 se clasifican los profesores con distintas categorías laborales en los clusters definidos según el porcentaje obtenido en cada uno de ellos.

Categoría		Número de conglomerados en dos fases						Total
		1	2	3	4	5	6	
ASO	Recuento	5	204	82	53	198	120	662
	% de CATEGORIA	0,80%	30,80%	12,40%	8,00%	29,90%	18,10%	100,00%
	% de Número de conglomerados en dos fases	1,50%	72,90%	25,40%	8,70%	25,60%	31,50%	24,60%
AY	Recuento	0	75	0	5	0	0	80
	% de CATEGORIA	0,00%	93,80%	0,00%	6,20%	0,00%	0,00%	100,00%
	% de Número de conglomerados en dos fases	0,00%	26,80%	0,00%	0,80%	0,00%	0,00%	3,00%
AYD	Recuento	2	0	2	6	7	1	18
	% de CATEGORIA	11,10%	0,00%	11,10%	33,30%	38,90%	5,60%	100,00%
	% de Número de conglomerados en dos fases	0,60%	0,00%	0,60%	1,00%	0,90%	0,30%	0,70%
CEU	Recuento	5	0	7	26	18	5	61
	% de CATEGORIA	8,20%	0,00%	11,50%	42,60%	29,50%	8,20%	100,00%
	% de Número de conglomerados en dos fases	1,50%	0,00%	2,20%	4,30%	2,30%	1,30%	2,30%
COD	Recuento	25	0	14	84	65	20	208
	% de CATEGORIA	12,00%	0,00%	6,70%	40,40%	31,20%	9,60%	100,00%
	% de Número de conglomerados en dos fases	7,70%	0,00%	4,30%	13,80%	8,40%	5,20%	7,70%
COL	Recuento	72	0	15	46	55	18	206
	% de CATEGORIA	35,00%	0,00%	7,30%	22,30%	26,70%	8,70%	100,00%
	% de Número de conglomerados en dos fases	22,30%	0,00%	4,60%	7,60%	7,10%	4,70%	7,70%
CU	Recuento	5	1	39	73	79	76	273
	% de CATEGORIA	1,80%	0,40%	14,30%	26,70%	28,90%	27,80%	100,00%
	% de Número de conglomerados en dos fases	1,50%	0,40%	12,10%	12,00%	10,20%	19,90%	10,20%
TEU	Recuento	190	0	81	107	138	72	588
	% de CATEGORIA	32,30%	0,00%	13,80%	18,20%	23,50%	12,20%	100,00%
	% de Número de conglomerados en dos fases	58,80%	0,00%	25,10%	17,60%	17,80%	18,90%	21,90%
TU	Recuento	19	0	83	208	214	69	593
	% de CATEGORIA	3,20%	0,00%	14,00%	35,10%	36,10%	11,60%	100,00%
	% de Número de conglomerados en dos fases	5,90%	0,00%	25,70%	34,20%	27,60%	18,10%	22,10%
Total	Recuento	323	280	323	608	774	381	2689
	% de CATEGORIA	12,00%	10,40%	12,00%	22,60%	28,80%	14,20%	100,00%
	% de Número de conglomerados en dos fases	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Tabla 6.15. Porcentaje de profesores según distintas categorías asociados cada uno de los seis clusters

ASO: Asociado

AY: Ayudante

AYD: Ayudante Doctor

CEU: Catedrático de Escuela Universitaria

COD: Colaborador Doctor

COL: Colaborador

CU: Catedrático de Universidad

TEU: Titular de Escuela Universitaria

TU: Titular de Universidad

VOLUMEN DE ACTIVIDAD DOCENTE

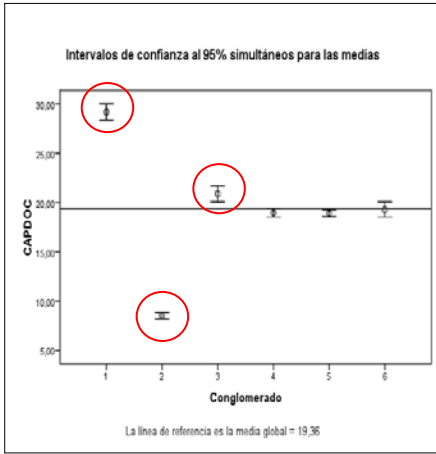


Figura 6.7. Contribución de CAPDOC a los clusters

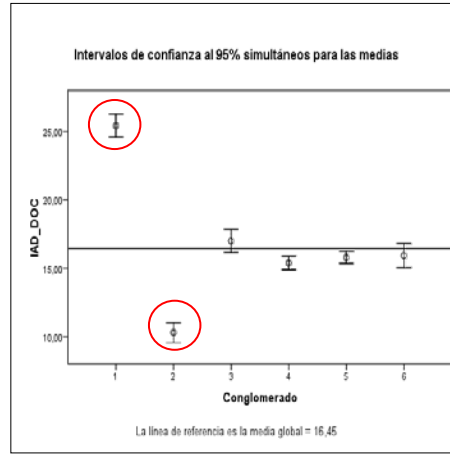


Figura 6.8. Contribución de IAD_{DOC} a los clusters

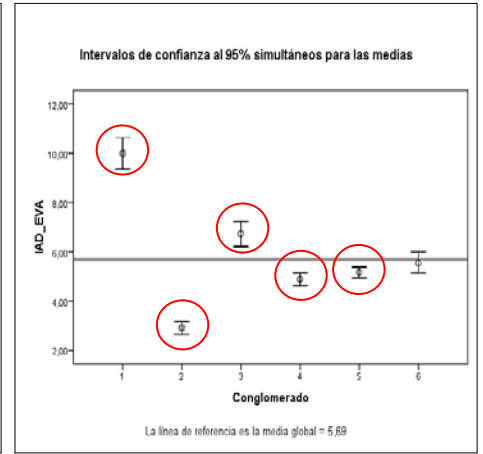


Figura 6.9. Contribución de IAD_{EVA} a los clusters

GESTIÓN

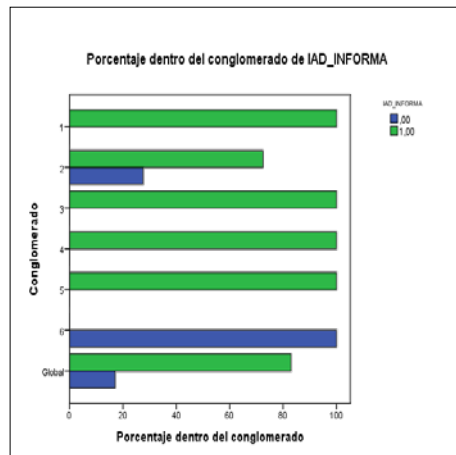


Figura 6.10. Contribución de IAD_{INFORMA} a los clusters

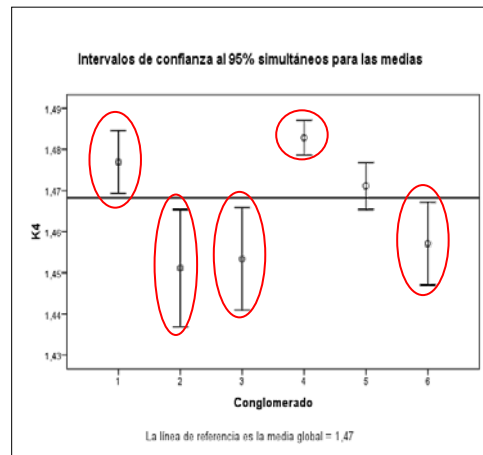


Figura 6.11. Contribución de K₄ a los clusters

RESULTADOS

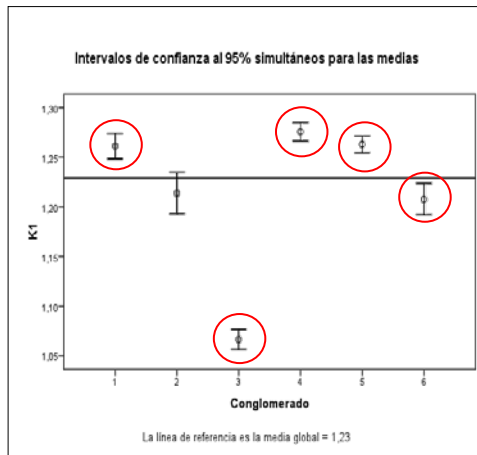


Figura 6.12. Contribución de K₁ a los clusters

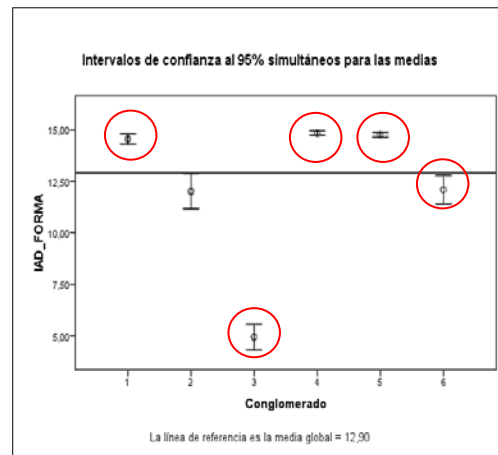


Figura 6.13. Contribución de IAD_{FORMA} a los clusters

ACTIVIDADES ADICIONALES

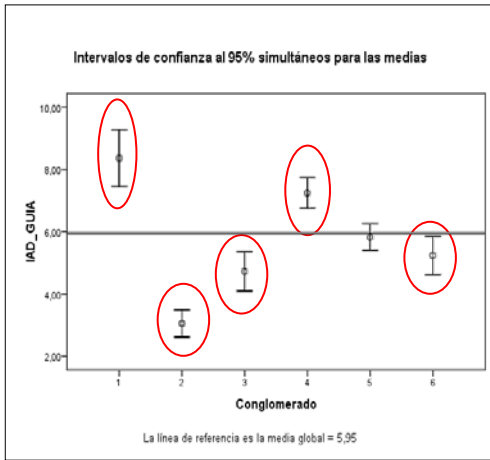


Figura 6.14. Contribución de IAD_{GUIA} a los cluster

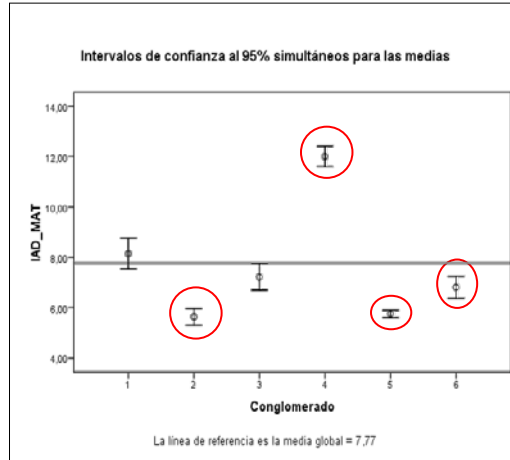


Figura 6.15. Contribución de IAD_{MAT} a los clusters

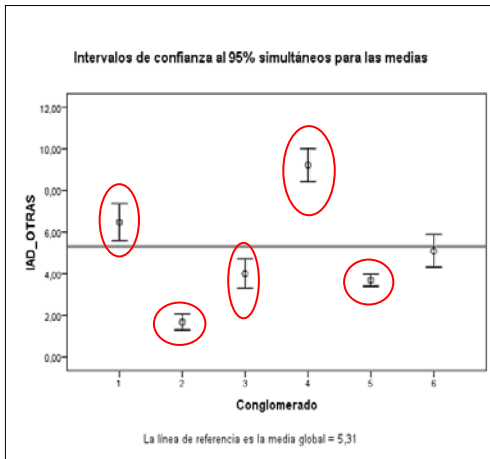


Figura 6.16. Contribución de IAD_{OTRAS} a los clusters

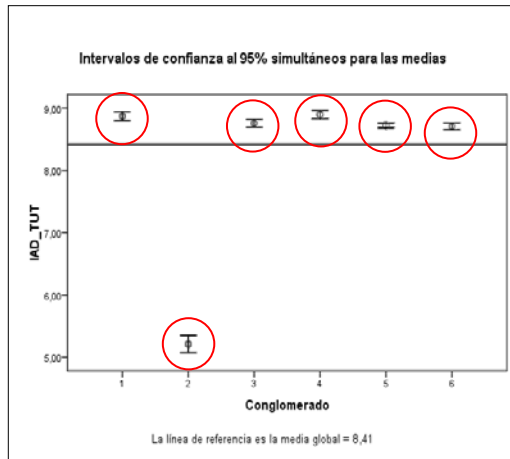


Figura 6.17. Contribución de IAD_{TUT} a los clusters

RENDIMIENTO

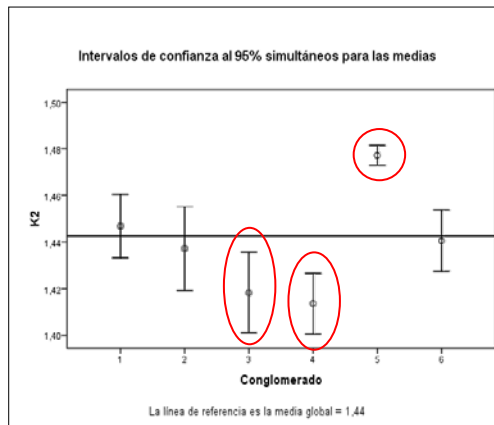


Figura 6.18. Contribución de K₂ a los clusters

El primer cluster define un perfil de profesorado con un volumen de actividad docente y unos resultados elevados. Se trata de profesores que suelen realizar las guías docentes de sus asignaturas y que realizan algunas actividades de carácter docente que no están directamente relacionadas con la docencia impartida. También realizan mucha actividad de gestión relacionada con su actividad docente. Este perfil de profesorado se caracteriza por estar formado en gran medida por profesores Titular de Escuela Universitaria (el 58,80%).

El segundo cluster está formado por profesores que presentan unos niveles muy bajos de actividad docente y de gestión relacionada con la docencia. Además, los profesores de este cluster se caracterizan por su poca participación en actividades docentes que no están relacionadas con las asignaturas impartidas. Ello se debe a que son profesores que llevan muy poca carga docente y que, seguramente, realicen otras actividades en otras organizaciones que no se recogen en estos indicadores. Este segundo perfil está muy bien caracterizado por la categoría profesional del profesorado, ya que en su mayoría está formado por profesores Asociados (72,90%) y Ayudantes (26,90%).

El tercer cluster define un perfil de profesorado con un volumen de actividad docente medio, pero con unos resultados y un rendimiento académico muy bajo. Además, se trata de profesores que no desarrollan un gran volumen de actividades docentes adicionales a la docencia habitual impartida.

El cuarto perfil, correspondiente al cuarto cluster, define un grupo de profesorado que imparte e invierte una cantidad media de horas en docencia y evaluación, dedicando una cantidad considerable de horas a las atenciones tutoriales. Además, obtienen unos resultados muy buenos y realizan actividades de gestión relacionada con la actividad docente. Se trata de profesores que dedican tiempo a la participación de actividades adicionales al desempeño de la docencia habitual. Sin embargo, el rendimiento académico de los alumnos es relativamente bajo.

El quinto perfil es similar al anterior con la diferencia de tratarse de profesores que no suelen participar en actividades adicionales a la docencia habitual y que obtienen un rendimiento académico muy elevado por parte de sus alumnos.

El último perfil define a un grupo de profesorado que no informa de forma pública del horario de tutorías desde el inicio del curso académico y que tampoco cumple con los plazos de entrega de actas. Además, no suele participar en actividades adicionales a la docencia habitual y los resultados obtenidos en la encuesta son muy bajos.

Los cuatro últimos perfiles no se asocian de forma bien definida con ninguna categoría de profesorado.

Análisis cluster excluyendo la variable ‘capacidad docente’

Este segundo análisis se ha considerado interesante para ver cómo se definen los distintos perfiles de profesorado utilizando el conjunto de indicadores simples definidos por la UPV sin tener en cuenta la carga docente.

En la tabla 6.16 se muestran los resultados de la agrupación automática en los que aparece para cada posible modelo, según el número de clusters, el criterio de selección elegido:

Número de conglomerados	Criterio bayesiano de Schwarz (BIC)	Cambio en BIC(a)	Razón de cambios en BIC(b)	Razón de medidas de distancia(c)
1	21254,076			
2	18835,499	-2418,577	1	1,49
3	17266,656	-1568,843	0,649	1,367
4	16163,709	-1102,947	0,456	1,927
5	15671,197	-492,513	0,204	1,192
6	15284,908	-386,288	0,16	1,013
7	14905,833	-379,075	0,157	1,086
8	14569,984	-335,85	0,139	1,07
9	14267,026	-302,958	0,125	1,002
10	13965,182	-301,844	0,125	1,367
11	13788,902	-176,28	0,073	1,107
12	13645,65	-143,252	0,059	1,011
13	13505,856	-139,794	0,058	1,385
14	13451,041	-54,815	0,023	1,19
15	13431,383	-19,658	0,008	1,057

Tabla 6.16. Criterio bayesiano para la selección del número de conglomerados con todos los indicadores simples

Aplicando los criterios seguidos en el análisis anterior, se ha seleccionado un modelo con cuatro conglomerados.

La distribución de las observaciones en los cuatro conglomerados y la importancia de cada uno de ellos se muestra en la tabla 6.17 y el gráfico de sectores en la figura 6.19:

Conglomerado	N	% del total
1	1093	40,60%
2	378	14,10%
3	760	28,30%
4	458	17,00%
Combinados	2689	100,00%
Total	2689	100,00%

Tabla 6.17. Distribución de las observaciones en los cuatro conglomerados

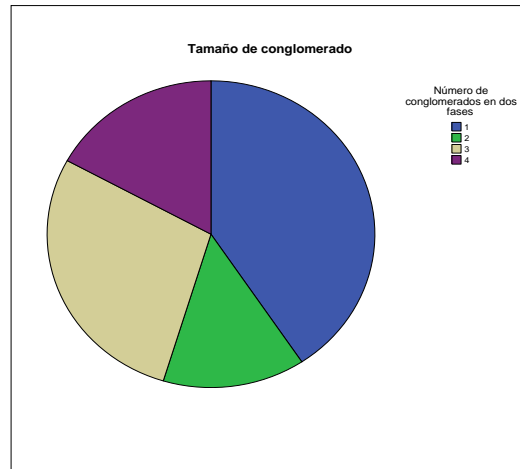


Figura 6.19. Gráfico de sectores con la distribución de las observaciones en los cuatro conglomerados

El conglomerado que está formado por la mayoría de las observaciones (40,60%) respecto a los indicadores de tipo continuo es el primero. El tercer conglomerado presenta el 28,30% de las observaciones, mientras que el resto de observaciones se reparte de la misma manera entre los otros dos clusters.

Respecto al indicador $IAD_{INFORMA}$, el 100% de los profesores que no informan de manera pública sobre los horarios de tutoría pertenecen al conglomerado 4. El resto de profesores que cumplen la característica que define al indicador se reparten en los tres primeros conglomerados, siendo el grupo 1 el que contiene el 49% de estos.

Conglomerado	$IAD_{INFORMA}$			
	0		1	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
1	0	0,00%	1093	49,00%
2	0	0,00%	378	16,90%
3	0	0,00%	760	34,10%
4	458	100,00%	0	0,00%
Combinados	458	100,00%	2231	100,00%

Tabla 6.18. Distribución de las observaciones con respecto al indicador binario en los cuatro conglomerados

Las figuras 6.20-6.30 representan los intervalos de confianza al 95% para las medias de los indicadores cuantitativos y el porcentaje del indicador $IAD_{INFORMA}$ para cada uno de los conglomerados. Se han marcado con un círculo en rojo los intervalos de confianza para aquellos conglomerados en los que el indicador evaluado contribuye a su formación según la significación del estadístico t de Student.

En la tabla 6.19 se han clasificado los profesores con distinta categoría laboral en los clusters definidos según el porcentaje obtenido en cada uno de ellos.

Categoría		Número de conglomerados en dos fases				
		1	2	3	4	Total
ASO	Recuento	88	124	264	186	662
	% de CATEGORIA	13,30%	18,70%	39,90%	28,10%	100,00%
	% de Número de conglomerados en dos fases	8,10%	32,80%	34,70%	40,60%	24,60%
AY	Recuento	4	5	61	10	80
	% de CATEGORIA	5,00%	6,30%	76,30%	12,50%	100,00%
	% de Número de conglomerados en dos fases	0,40%	1,30%	8,00%	2,20%	3,00%
AYD	Recuento	11	2	4	1	18
	% de CATEGORIA	61,10%	11,10%	22,20%	5,60%	100,00%
	% de Número de conglomerados en dos fases	1,00%	0,50%	0,50%	0,20%	0,70%
CEU	Recuento	36	6	14	5	61
	% de CATEGORIA	59,00%	9,80%	23,00%	8,20%	100,00%
	% de Número de conglomerados en dos fases	3,30%	1,60%	1,80%	1,10%	2,30%
COD	Recuento	139	13	36	20	208
	% de CATEGORIA	66,80%	6,30%	17,30%	9,60%	100,00%
	% de Número de conglomerados en dos fases	12,70%	3,40%	4,70%	4,40%	7,70%
COL	Recuento	135	18	35	18	206
	% de CATEGORIA	65,50%	8,70%	17,00%	8,70%	100,00%
	% de Número de conglomerados en dos fases	12,40%	4,80%	4,60%	3,90%	7,70%
CU	Recuento	82	38	76	77	273
	% de CATEGORIA	30,00%	13,90%	27,80%	28,20%	100,00%
	% de Número de conglomerados en dos fases	7,50%	10,10%	10,00%	16,80%	10,20%
TEU	Recuento	324	87	105	72	588
	% de CATEGORIA	55,10%	14,80%	17,90%	12,20%	100,00%
	% de Número de conglomerados en dos fases	29,60%	23,00%	13,80%	15,70%	21,90%
TU	Recuento	274	85	165	69	593
	% de CATEGORIA	46,20%	14,30%	27,80%	11,60%	100,00%
	% de Número de conglomerados en dos fases	25,10%	22,50%	21,70%	15,10%	22,10%
Total	Recuento	1093	378	760	458	2689
	% de CATEGORIA	40,60%	14,10%	28,30%	17,00%	100,00%
	% de Número de conglomerados en dos fases	100,00%	100,00%	100,00%	100,00%	100,00%

Tabla 6.19. Porcentaje de profesores según distintas categorías en cada uno de los cuatro clusters

ASO: Asociado

AY: Ayudante

AYD: Ayudante Doctor

CEU: Catedrático de Escuela Universitaria

COD: Colaborador Doctor

COL: Colaborador

CU: Catedrático de Universidad

TEU: Titular de Escuela Universitaria

TU: Titular de Universidad

VOLUMEN DE ACTIVIDAD DOCENTE

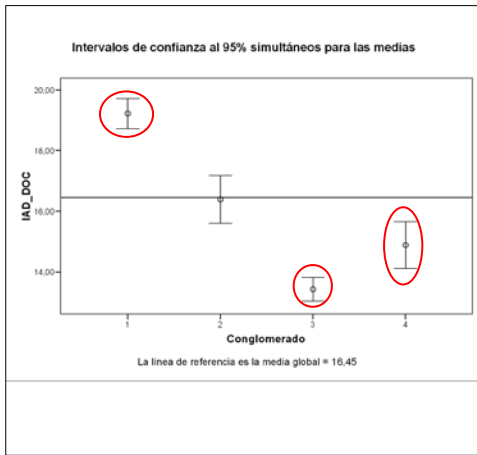


Figura 6.20. Contribución de IAD_{DOC} a los clusters

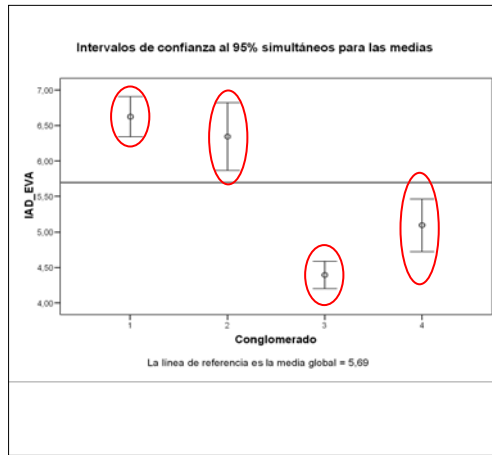


Figura 6.21. Contribución de IAD_{EVA} a los clusters

GESTIÓN

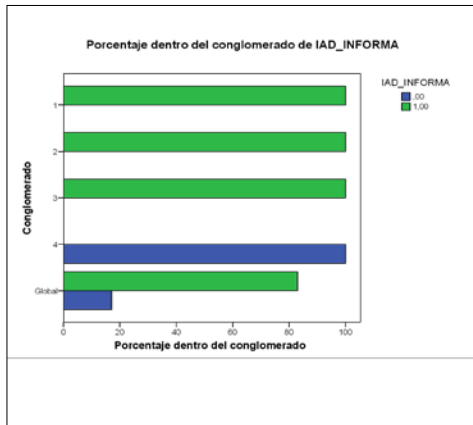


Figura 6.22. Contribución de IAD_{INFORMA} a los clusters

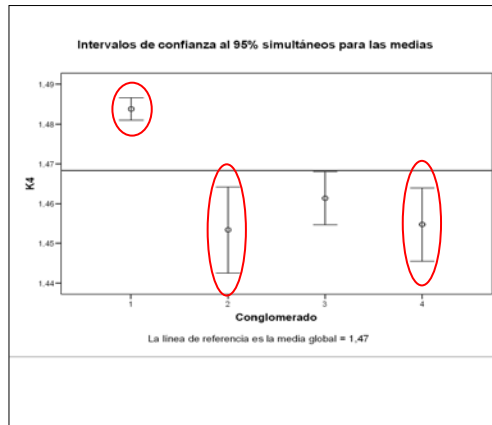


Figura 6.23. Contribución de K₄ a los clusters

RESULTADOS

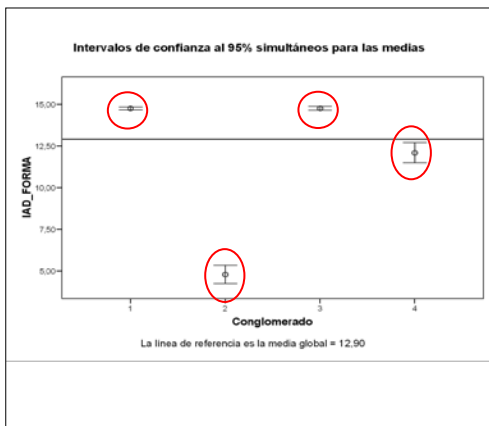


Figura 6.24. Contribución de IAD_{FORMA} a los clusters

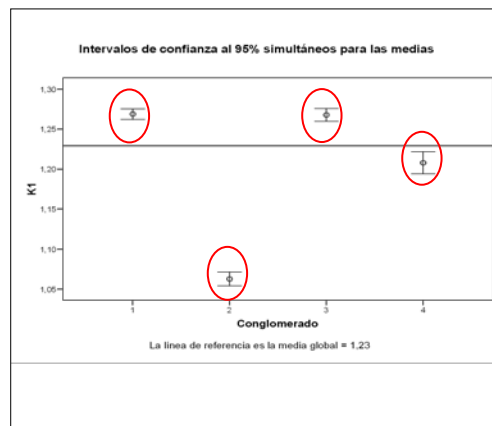


Figura 6.25. Contribución de K₁ a los clusters

ACTIVIDADES ADICIONALES

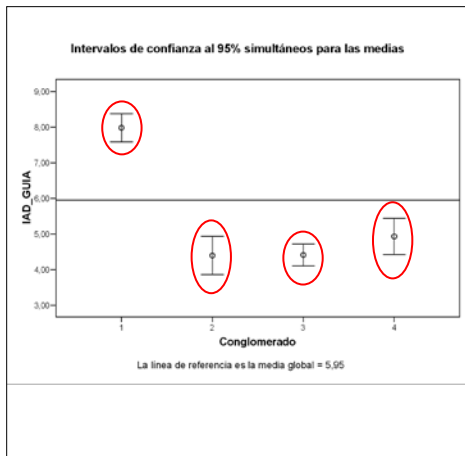


Figura 6.26. Contribución de IAD_{GUIA} a los clusters

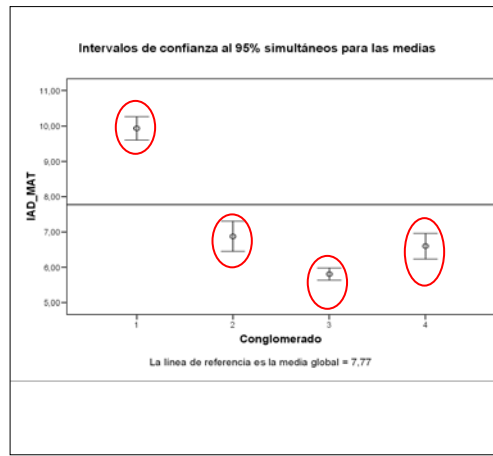


Figura 6.27. Contribución de IAD_{MAT} a los clusters

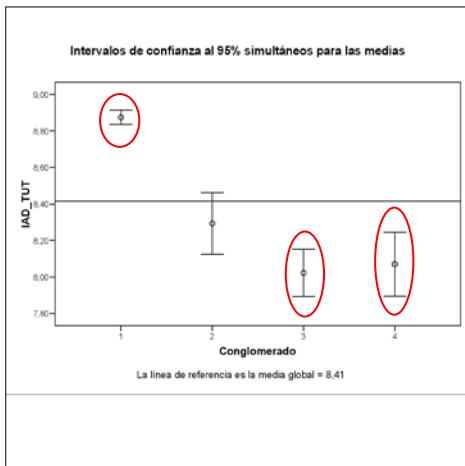


Figura 6.28. Contribución de IAD_{TUT} a los clusters

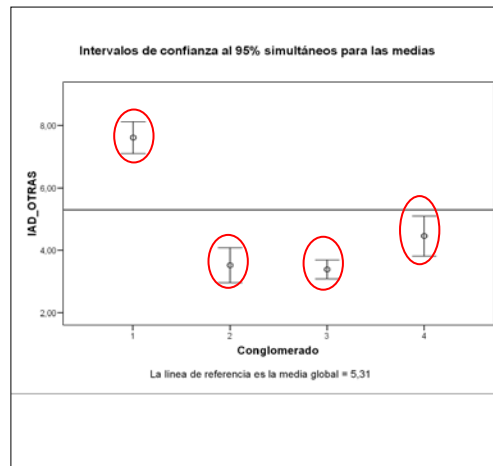


Figura 6.29. Contribución de IAD_{OTRAS} a los clusters

RENDIMIENTO

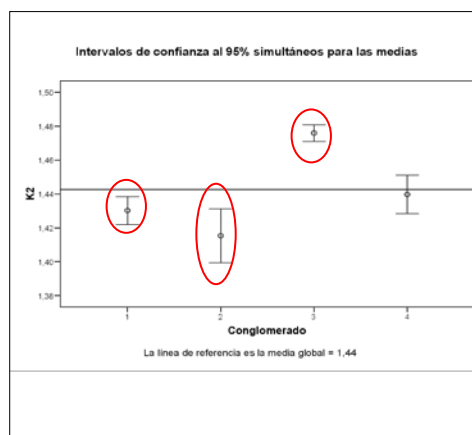


Figura 6.30. Contribución de K_2 a los clusters

En este caso, se han podido definir solamente cuatro perfiles de profesorado. El primer conglomerado se identifica con el primer conglomerado del análisis previo. El segundo conglomerado define un perfil de profesorado con alta puntuación en la variable IAD_{EVA} , lo cual significa que son profesores que tienen muchos alumnos matriculados en sus clases. Se trata de profesores que suelen informar del horario de tutorías desde el inicio del curso académico pero, sin embargo, no cumplen con los plazos de entrega de actas y tampoco dedican más tiempo a otras actividades adiciones a la docencia habitual. Además, los resultados obtenidos según la nota de la encuesta y el rendimiento académico de los alumnos son muy bajos. El tercer cluster se identifica con el segundo conglomerado del análisis previo. Y, finalmente, el cuarto cluster define un perfil de profesorado similar al sexto y último cluster del análisis anterior.

En este segundo análisis no se ha podido asociar ninguna categoría laboral de profesorado de forma bien definida a ninguno de los perfiles identificados.

6.2.5. Normalización de los datos

La normalización de los datos es una etapa importante que se debe tener en consideración antes de aplicar cualquier tipo de agregación con el fin de poder tratar y comparar los indicadores simples entre sí.

Según el tipo de normalización conveniente en cada caso se pueden ajustar los datos para unificarlos en una misma unidad de medida, en un mismo rango de variación o ajustarlos de tal forma que sigan distribuciones no tan asimétricas y disminuya la presencia de valores atípicos. En este estudio se han aplicado las técnicas de normalización que se exponen a continuación, empleando la siguiente notación:

x_{qc}^t : puntuación del indicador simple q para la unidad de análisis c en el momento de tiempo t para $q = 1, \dots, Q$ y $c = 1, \dots, M$, donde Q es el número de indicadores simples y M el número de unidades de análisis/observaciones. En el caso de estudio las unidades de análisis se corresponden con los profesores y t es el año académico en el que se está midiendo su actividad docente.

Las características y propiedades de cada una de las técnicas de normalización aplicadas se describen en la sección 3.4.5 del capítulo 3.

Estandarización (ó z-score)

$$I_{qc}^t = \frac{x_{qc}^t - x_{qc=\bar{c}}^t}{\sigma_{qc=\bar{c}}^t}$$

Se ha considerado adecuado aplicar la técnica de estandarización, ya que convierte los indicadores en una misma escala adimensional con media cero y desviación típica la unidad, de modo que se mantienen las distancias relativas. De esta forma, se pueden comparar los indicadores con una misma desviación típica.

Según Freudenberg (2003) se trata de uno de los métodos más usados en la construcción de ICs. Este tipo de normalización ya se ha aplicado anteriormente para la detección de casos atípicos a nivel univariante.

Re-escalamiento (ó min-max)

$$I_{qc}^t = \frac{x_{qc}^t - \min_c(x_q^t)}{\max_c(x_q^t) - \min_c(x_q^t)}$$

La normalización min-max transforma los indicadores a un mismo rango de variación [0,1]. Por tanto, transforma los indicadores en una misma escala adimensional preservando las distancias relativas.

En el estudio se ha aplicado también este tipo de normalización para conseguir que todos los indicadores se muevan en un mismo rango. Tal y como se ha visto en la etapa de exploración de los datos, los indicadores presentan rangos de variación muy diferentes entre ellos. Por tanto, es conveniente realizar algún tipo de normalización sobre los datos. Si no fuese así el indicador IAD_{DOC} que tiene un rango de [5,5;54] tendría una contribución mucho mayor en la construcción del IC que el indicador K_1 con rango [0,1]. Este es un claro ejemplo de la heterogeneidad que presentan los datos originales.

En cuanto a los demás tipos de normalización que se presentan en el capítulo 3 no se ha considerado conveniente su aplicación sobre el conjunto de datos por las tres razones siguientes:

- Se pierde mucha información a nivel absoluto (por ejemplo en el método *ranking*)
- Se deben tomar decisiones subjetivas para la asignación de un umbral o una unidad referencial con la que compararse (por ejemplo en los métodos: distancia a una unidad referencial, categorización de escalas, categorización de valores por encima o por debajo de la media)
- En este estudio se pretende obtener una puntuación de cada profesor para un año académico en particular, independientemente de su posterior comparación con el año anterior (ejemplos de métodos de evolución temporal: indicadores cíclicos, porcentaje de diferencias anuales en años consecutivos)

6.2.6. Ponderación y agregación de los indicadores simples

Las técnicas de ponderación y de agregación de los indicadores simples se han realizado teniendo en cuenta dos aspectos:

- Considerando todo el conjunto de indicadores simples como una única dimensión para la etapa de ponderación y agregación en un único valor.

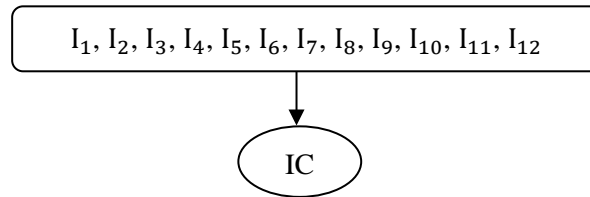


Figura 6.31. Ponderación y agregación aplicada al conjunto de indicadores simples para la obtención del IC

- Agrupando el conjunto de indicadores según las dimensiones del marco conceptual teórico y, por tanto, según si pertenecen a las dimensiones **Planificación** de la docencia, **Desarrollo** de la enseñanza o **Resultados**. En este caso se trabaja con dos niveles:

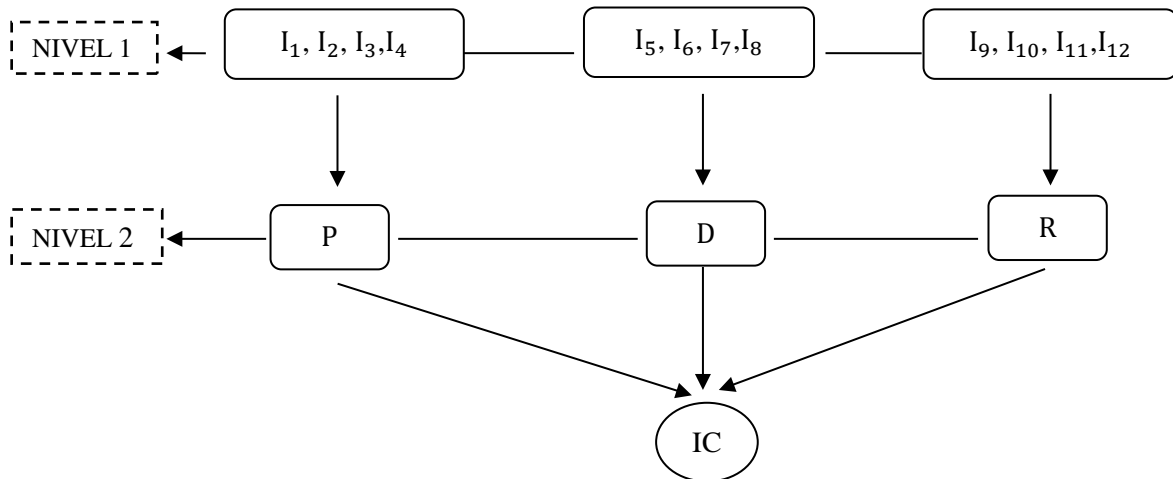


Figura 6.32. Ponderación y agregación aplicada a distintos niveles para la obtención del IC

En el nivel 1 se han aplicado las mismas técnicas de ponderación y de agregación en cada uno de los tres grupos de indicadores por separado. En el nivel 2 las tres dimensiones están representadas por un único valor obtenido en el nivel 1. En este segundo nivel se han aplicado técnicas de ponderación y agregación a las tres dimensiones con el fin de obtener una puntuación global para cada una de las observaciones.

TÉCNICAS DE PONDERACIÓN

Adicionalmente al conjunto de pesos asignado en la construcción del IAD definido por la UPV, se ha considerado interesante aplicar dos métodos de ponderación alternativos: la asignación igualitaria de pesos y la asignación presupuestaria como técnica participativa.

Para la aplicación de la técnica de asignación presupuestaria se ha solicitado la participación de 13 expertos pertenecientes a la Comisión de Evaluación Docente (CED) que es el órgano responsable de la evaluación de la docencia en la UPV. La técnica ha consistido en la distribución, por parte de cada experto, de 100 puntos entre cada uno de los indicadores simples, basándose en la importancia que cada uno de ellos debe tener sobre el IAD. Una vez obtenido un vector de pesos para cada indicador se ha abierto un debate entre los expertos con el fin de llegar a un consenso final sobre el peso asignado a cada indicador en base a los objetivos estratégicos planteados por la UPV. Así es como se ha definido el conjunto de pesos objetivo⁵.

La definición del conjunto de pesos objetivo ha sido de gran utilidad para definir las pautas a tener en consideración en la construcción de un nuevo índice en el caso en que la UPV se replantease reorientar hacia otra dirección el indicador de actividad docente vigente en la actualidad, o quisiese desarrollar otro indicador compuesto en la gestión universitaria.

Las dos técnicas de ponderación se han aplicado según se describe a continuación:

- Ponderación de los indicadores simples, con $\sum_i w_i = 1$, mediante la asignación de pesos iguales a todos ellos (EW, por sus siglas en inglés)
- Asignación de pesos iguales intra-dimensión (nivel 1) e inter-dimensiones (nivel 2). La suma de los pesos debe ser la unidad tanto dentro de cada dimensión como entre dimensiones
- Asignación presupuestaria (BA, por sus siglas en inglés), con $\sum_i w_i = 1$, a cada uno de los indicadores simples mediante la aplicación de técnicas de consenso del grupo de expertos de la CED
- Asignación presupuestaria dentro de cada dimensión (nivel 1) y entre las dimensiones (nivel 2). La suma de los pesos debe ser la unidad tanto dentro de cada dimensión como entre dimensiones

No se ha aplicado la técnica de ponderación basada en el análisis factorial puesto que, tal y como se ha visto en la etapa de exploración de los datos, no existe una fuerte correlación entre éstos. Además, la correlación existente no muestra las relaciones reales que existen entre los indicadores simples según el marco conceptual teórico.

⁵ A lo largo de este capítulo aparece el concepto “*pesos objetivo*” refiriéndose al conjunto de pesos definido, por consenso, por el grupo de expertos de la CED de la UPV

TÉCNICAS DE AGREGACIÓN

En cuanto a las técnicas de agregación se ha aplicado la agregación lineal y geométrica según se describe a continuación:

- Agregación lineal al considerar el conjunto de indicadores simples como una única dimensión
- Agregación geométrica al considerar el conjunto de indicadores simples como una única dimensión
- Agregación lineal en ambos niveles, tanto en el nivel de indicadores simples intra-dimensión (nivel 1) como entre dimensiones (nivel 2)
- Agregación geométrica en ambos niveles
- Agregación lineal en el primer nivel y agregación geométrica entre dimensiones (nivel 2)

La agregación lineal es una de las técnicas de agregación más usada en la construcción de ICs (Nardo et al., 2008). La aplicación de esta técnica asume una compensación perfecta entre las puntuaciones de los diferentes indicadores simples evaluados.

La agregación geométrica da cuenta de la compensación imperfecta entre los indicadores o las dimensiones. De esta manera, se aborda una de las críticas más serias a la fórmula de agregación lineal que permite una perfecta compensación entre indicadores o dimensiones (Munda, 2008; Munda y Nardo, 2009; Billaut, Bouyssou y Vincke, 2010). Algunos de los ICs desarrollados anualmente, como por ejemplo el Índice de Desarrollo Humano, han incluido la agregación geométrica en su formulación para evitar la perfecta compensación entre los indicadores (Programa de las Naciones Unidas para el Desarrollo, 2010).

Aun así, en ambas técnicas subyace un razonamiento compensatorio entre los indicadores, lo que supone la existencia de vínculos entre ellos que hace que no se puedan tratar los pesos asignados como una medida de la importancia de cada uno de éstos sobre el IC (Munda, 2008).

Alternativamente a estas dos técnicas, las técnicas de agregación multi-criterio no compensatorias no reflejan la posibilidad de compensación entre los indicadores. Además, los pesos asignados se pueden tratar como medidas de importancia de cada uno de los indicadores simples (Munda, 2008). Sin embargo, no se han aplicado en el estudio puesto que son más complicadas de interpretar y de ser aceptadas por el público en general, ya que su uso no es tan común. Este tipo de técnicas constituyen una línea de investigación futura muy interesante en la metodología de construcción de indicadores compuestos.

Finalmente, las técnicas de agregación mediante la suma de rankings y el conteo de los indicadores que superan una referencia dada tampoco se han aplicado, ya que tienen el inconveniente de perder mucha información al considerar sólo la posición de cada una de las observaciones para la agregación.

6.2.7. Escenarios de trabajo

En la tabla 6.20 se describen los escenarios de trabajo que se han propuesto en el estudio según el tipo de normalización, ponderación y agregación aplicado en la construcción del IC:

Escenario	Normalización	Ponderación en el nivel 1	Ponderación en el nivel 2	Agregación en el nivel 1	Agregación en el nivel 2
EUPV	No	BA	BA	Lineal	Mixta
E1	Z-score	EW	EW	Lineal	Lineal
E2	Min-Max	EW	EW	Lineal	Lineal
E3	Min-Max	BA	BA	Geométrica	Geométrica
E4	Z-score		EW ²		Lineal ²
E5	Min-Max		EW ²		Lineal ²
E6	Min-Max		EW ²		Geométrica ²
E7	Z-score		BA ²		Lineal ²
E8	Min-Max		BA ²		Lineal ²
E9	Min-Max		BA ²		Geométrica ²
E10	Min-Max	EW	EW	Lineal	Geométrica
E11	Min-Max	BA	BA	Lineal	Geométrica
E12	Min-Max	BA	BA	Lineal	Mixta
E13	Z-score	BA	BA	Lineal	Mixta
E14	Min-Max	EW	EW	Geométrica	Geométrica
E15 ¹	Z-score		BA ³		Lineal

Tabla 6.20. Escenarios de trabajo definidos

¹ Este escenario se ha añadido una vez calculado el análisis de sensibilidad de E7 para ajustar los pesos objetivo propuestos por un grupo de expertos a la importancia relativa obtenida de cada indicador sobre el IC (véase sección 6.2.8)

² En este caso no se consideran los indicadores agrupados en dimensiones

³ Los pesos se ajustan una vez realizado el análisis de sensibilidad de E7

Las casillas sombreadas en gris siguen el mismo escenario de trabajo propuesto por la UPV

Los escenarios E3, E6, E9, E10, E11 y E14 se han construido aplicando un tipo de normalización min-max combinado con una agregación geométrica en alguno de los niveles. Para evitar que los ceros que produce la normalización min-max anulen completamente el indicador compuesto una vez realizada la agregación geométrica se ha aplicado una transformación lineal de modo que la normalización min-max se ha considerado en una escala [1,10].

Por otra parte, en este estudio no se han considerado aquellos escenarios en los que se combina una normalización z-score con una agregación geométrica, puesto que se pueden encontrar cálculos inconsistentes para los valores negativos de la normalización. Una posible solución a este problema es realizar una transformación lineal, sin embargo, la presencia de valores extremos requiere la aplicación de algún tipo de transformación logarítmica. Para no alterar la interpretación de los indicadores no se han propuesto este tipo de escenarios en este estudio.

Representando por i cada una de las observaciones del estudio, con $i = 1, \dots, 2.689$, las ecuaciones correspondientes a cada uno de los escenarios que se han usado en la formación de cada IC son las siguientes:

EUPV:

$$IC_i = (0,3 * IAD_{FORMA} + 0,3 * IAD_{GUIA} + 0,3 * IAD_{INFORMA} + 0,3 * IAD_{MAT} + 0,7 * IAD_{DOC} + 0,7 * IAD_{TUT} + 0,7 * IAD_{EVA} + 0,7 * IAD_{OTRAS}) * (0,6 * K_1 + 0,3 * K_2 + 0,05 * K_3 + 0,05 * K_4)$$

E1, E2:

$$IC_i = \frac{1}{3} * (\frac{1}{4} * IAD_{FORMA} + \frac{1}{4} * IAD_{GUIA} + \frac{1}{4} * IAD_{INFORMA} + \frac{1}{4} * IAD_{MAT}) + \frac{1}{3} * (\frac{1}{4} * IAD_{DOC} + \frac{1}{4} * IAD_{TUT} + \frac{1}{4} * IAD_{EVA} + \frac{1}{4} * IAD_{OTRAS}) + \frac{1}{3} * (\frac{1}{3} * K_1 + \frac{1}{3} * K_2 + \frac{1}{3} * K_4)$$

E3:

$$IC_i = (IAD_{FORMA}^{0,343} * IAD_{GUIA}^{0,228} * IAD_{INFORMA}^{0,143} * IAD_{MAT}^{0,286})^{0,35} * (IAD_{DOC}^{0,375} * IAD_{TUT}^{0,175} * IAD_{EVA}^{0,2} * IAD_{OTRAS}^{0,25})^{0,4} * (K_1^{0,76} * K_2^{0,2} * K_4^{0,04})^{0,25}$$

E4, E5:

$$IC_i = \frac{1}{11} * IAD_{FORMA} + \frac{1}{11} * IAD_{GUIA} + \frac{1}{11} * IAD_{INFORMA} + \frac{1}{11} * IAD_{MAT} + \frac{1}{11} * IAD_{DOC} + \frac{1}{11} * IAD_{TUT} + \frac{1}{11} * IAD_{EVA} + \frac{1}{11} * IAD_{OTRAS} + \frac{1}{11} * K_1 + \frac{1}{11} * K_2 + \frac{1}{11} * K_4$$

E6:

$$IC_i = IAD_{FORMA}^{\frac{1}{11}} * IAD_{GUIA}^{\frac{1}{11}} * IAD_{INFORMA}^{\frac{1}{11}} * IAD_{MAT}^{\frac{1}{11}} * IAD_{DOC}^{\frac{1}{11}} * IAD_{TUT}^{\frac{1}{11}} * IAD_{EVA}^{\frac{1}{11}} * IAD_{OTRAS}^{\frac{1}{11}} * K_1^{\frac{1}{11}} * K_2^{\frac{1}{11}} * K_4^{\frac{1}{11}}$$

E7, E8:

$$IC_i = \frac{12}{100} * IAD_{FORMA} + \frac{8}{100} * IAD_{GUIA} + \frac{5}{100} * IAD_{INFORMA} + \frac{10}{100} * IAD_{MAT} + \frac{15}{100} * IAD_{DOC} + \frac{7}{100} * IAD_{TUT} + \frac{8}{100} * IAD_{EVA} + \frac{10}{100} * IAD_{OTRAS} + \frac{19}{100} * K_1 + \frac{5}{100} * K_2 + \frac{1}{100} * K_4$$

E9:

$$IC_i = IAD_{FORMA}^{\frac{12}{100}} * IAD_{GUIA}^{\frac{8}{100}} * IAD_{INFORMA}^{\frac{5}{100}} * IAD_{MAT}^{\frac{10}{100}} * IAD_{DOC}^{\frac{15}{100}} * IAD_{TUT}^{\frac{7}{100}} * IAD_{EVA}^{\frac{8}{100}} * IAD_{OTRAS}^{\frac{10}{100}} * K_1^{\frac{19}{100}} * K_2^{\frac{5}{100}} * K_4^{\frac{1}{100}}$$

E10:

$$IC_i = (\frac{1}{4} * IAD_{FORMA} + \frac{1}{4} * IAD_{GUIA} + \frac{1}{4} * IAD_{INFORMA} + \frac{1}{4} * IAD_{MAT})^{\frac{1}{3}} * (\frac{1}{4} * IAD_{DOC} + \frac{1}{4} * IAD_{TUT} + \frac{1}{4} * IAD_{EVA} + \frac{1}{4} * IAD_{OTRAS})^{\frac{1}{3}} * (\frac{1}{3} * K_1 + \frac{1}{3} * K_2 + \frac{1}{3} * K_4)^{\frac{1}{3}}$$

E11:

$$IC_i = (0,343 * IAD_{FORMA} + 0,228 * IAD_{GUIA} + 0,143 * IAD_{INFORMA} + 0,286 * IAD_{MAT})^{0,35} * (0,375 * IAD_{DOC} + 0,175 * IAD_{TUT} + 0,2 * IAD_{EVA} + 0,25 * IAD_{OTRAS})^{0,4} * (0,76 * K_1 + 0,2 * K_2 + 0,04 * K_4)^{0,25}$$

E12, E13:

$$IC_i = (0,3 * IAD_{FORMA} + 0,3 * IAD_{GUIA} + 0,3 * IAD_{INFORMA} + 0,3 * IAD_{MAT} + 0,7 * IAD_{DOC} + 0,7 * IAD_{TUT} + 0,7 * IAD_{EVA} + 0,7 * IAD_{OTRAS}) * (0,65 * K_1 + 0,3 * K_2 + 0,05 * K_4)$$

E14:

$$IC_i = (IAD_{FORMA}^{\frac{1}{4}} * IAD_{GUIA}^{\frac{1}{4}} * IAD_{INFORMA}^{\frac{1}{4}} * IAD_{MAT}^{\frac{1}{4}})^{\frac{1}{3}} * (IAD_{DOC}^{\frac{1}{4}} * IAD_{TUT}^{\frac{1}{4}} * IAD_{EVA}^{\frac{1}{4}} * IAD_{OTRAS}^{\frac{1}{4}})^{\frac{1}{3}} * (K_1^{\frac{1}{3}} * K_2^{\frac{1}{3}} * K_4^{\frac{1}{3}})^{\frac{1}{3}}$$

E15:

$$IC_i = \frac{6}{100} * IAD_{FORMA} + \frac{7}{100} * IAD_{GUIA} + \frac{7}{100} * IAD_{INFORMA} + \frac{12}{100} * IAD_{MAT} + \frac{16}{100} * IAD_{DOC} + \frac{1}{100} * IAD_{TUT} + \frac{10}{100} * IAD_{EVA} + \frac{10}{100} * IAD_{OTRAS} + \frac{18}{100} * K_1 + \frac{12}{100} * K_2 + \frac{1}{100} * K_4$$

En los escenarios en los que se ha realizado una normalización previa a los datos se ha prescindido del indicador K_3 . La eliminación de este indicador no altera el orden final de las posiciones del profesorado, puesto que se trata de un indicador constante para todas las observaciones. El peso asignado a este indicador se ha añadido al indicador que más pesa dentro de su misma dimensión, el indicador K_1 , para que la suma de los pesos asignados sea la unidad. Esta decisión se ha tomado basándose en la opinión del grupo de expertos.

6.2.8. Análisis de Sensibilidad Global

Los ICs se construyen combinando los indicadores simples de entrada mediante un modelo matemático en el que se aplica un tratamiento previo sobre el conjunto de indicadores, una asignación de pesos a cada uno de ellos y una posterior agregación para la obtención del IC final.

En la fase de ponderación se asigna un peso nominal⁶ a cada indicador simple. Este peso nominal suele tratarse como una medida de importancia del indicador simple sobre el IC (Freudenberg, 2003). La asignación presupuestaria (BA, por sus siglas en inglés) como técnica de ponderación es un claro ejemplo donde se evidencia la interpretación de los pesos nominales, por parte de los expertos, como medidas de importancia relativa de cada indicador. No obstante, en muchos casos puede ocurrir que la importancia que se le asigna en forma de peso nominal a

⁶ A lo largo de este capítulo aparece el concepto “pesos nominales” refiriéndose al conjunto de pesos asignados a los indicadores simples en la fase de ponderación de la construcción del IC

cada indicador no sea la misma que la que finalmente se refleja en el IC (Paruolo, Saltelli y Saisana, 2013).

Esto ocurre, por ejemplo, cuando se utiliza la técnica más común de agregación lineal aditiva para la construcción del IC:

$$Y_c = IC_c = \sum_{q=1}^Q w_q I_{qc}$$

con $\sum_{q=1}^Q w_q = 1$; $0 \leq w_q \leq 1$, y I_{qc} el valor del indicador q para la observación c , para $q = 1, \dots, Q$ y $c = 1, \dots, M$. En este caso, el ratio entre dos pesos nominales, w_i/w_j , es una medida de *trade-off* o factor de escala que informa del efecto de sustitución entre las dos variables individuales, es decir, cuánto debe aumentar I_{ic} para compensar o equilibrar una disminución de una unidad en I_{jc} (Boyssou et al., 2006; Decancq y Lugo, 2013). En este ejemplo el efecto real de los indicadores simples sobre el IC no va a coincidir exactamente con el peso nominal que se le asigna a cada uno de ellos en el proceso de construcción.

El efecto real o principal, también conocido como peso efectivo, de cada uno de los indicadores simples sobre el IC va a depender del tratamiento previo de los indicadores (transformación de los datos, normalización, imputación, agregación...) y de la estructura de covarianza y correlación de los indicadores simples (Paruolo, Saltelli y Saisana, 2013).

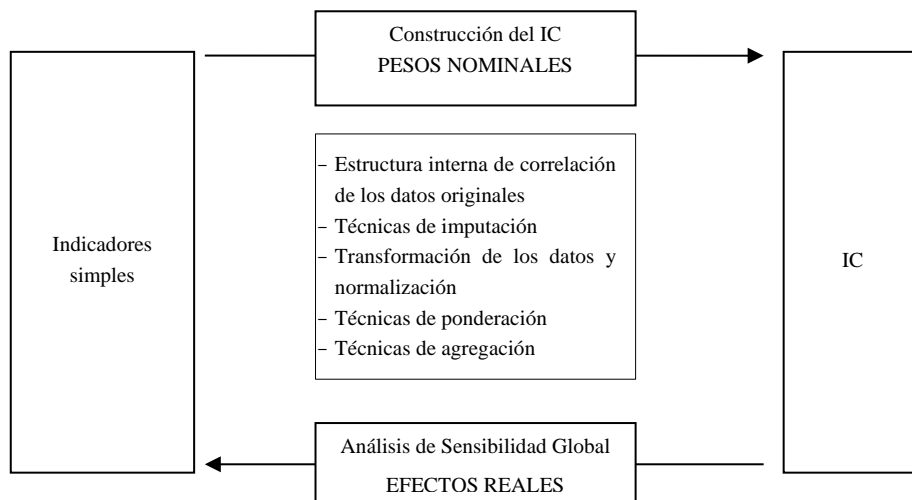


Figura 6.33. Fuentes de variación que afectan a la construcción del IC

A modo de ejemplo en el que se refleja lo dicho anteriormente se va a ilustrar un caso descrito por Paruolo, Saltelli y Saisana (2013). Supóngase que el decano académico de una universidad está interesado en evaluar el rendimiento del personal académico de su facultad, asignando la misma importancia a los indicadores simples, x_1 : número de publicaciones; x_2 : número de horas de docencia impartida y x_3 : número de horas de gestión administrativa. Sea

$y = \frac{x_1+x_2+x_3}{3}$ el indicador compuesto definido y sea $R_i^2 := \text{corr}^2(y, x_i)$ una medida de asociación entre el IC y los indicadores simples x_1, x_2 y x_3 .

Considérense dos situaciones distintas en las que se refleja la influencia de la varianza y de la correlación de los indicadores simples sobre el rendimiento del personal académico. En ambas situaciones supóngase que los tres indicadores siguen una distribución normal conjunta de media cero:

1. En primer lugar, sea $\sigma_1^2=7$; $\sigma_2^2=1$ y $\sigma_3^2=1$. Supóngase que las variables x_i están incorreladas. Además, la varianza $\sigma_1^2=7$ se ha elegido de tal forma que $\sigma_y^2=1$. Por tanto, se tiene:

$$R_1^2 = \frac{7}{9} \approx 0,778, \quad R_2^2 = R_3^2 = \frac{1}{63} \approx 0,016$$

Lo que implica que la importancia (medida mediante el coeficiente R_i^2) de las variables x_2 y x_3 relativa a x_1 es $1/49 \approx 0,020$. Este primer ejemplo muestra cómo las varianzas pueden afectar al cálculo de las medidas de importancia. Por tanto, el decano debería tener en consideración la varianza de los indicadores previamente al cómputo del IC. Una forma de corregir este problema es aplicar alguna técnica de normalización que iguale las varianzas de los indicadores. Mediante la normalización z-score (o estandarización) se corrige el problema anterior, ya que se elimina la heterogeneidad entre los indicadores simples, aunque todavía se mantiene la estructura de correlación existente entre los indicadores.

2. Considérese la segunda situación donde los indicadores x_1, x_2 y x_3 están estandarizados y, por tanto, tienen la misma varianza igual a la unidad. Supóngase que las correlaciones entre los indicadores son cero, $\rho_{i,j} := \text{corr}(x_i, x_j) = 0$, excepto para los indicadores x_2 y x_3 , $\rho_{2,3} = \rho_{3,2} > 0$. Por tanto se tiene,

$$R_1^2 = \frac{1}{3 + 2\rho_{2,3}}; \quad R_2^2 = R_3^2 = \frac{(1 + \rho_{2,3})^2}{3 + 2\rho_{2,3}}; \quad \frac{R_1^2}{R_2^2} = \frac{1}{(1 + \rho_{2,3})^2}$$

donde la importancia de los indicadores x_2 y x_3 es la misma; propiedad general de los indicadores estandarizados. Nótese que la importancia del indicador x_2 y x_3 es más elevada que la importancia del indicador x_1 , debido a que $\rho_{3,2} > 0$. Sea, por ejemplo, $\rho_{3,2}=0,7$:

$$R_1^2 = \frac{5}{22} \approx 0,227, \quad R_2^2 = R_3^2 = \frac{289}{440} \approx 0,657, \quad \frac{R_1^2}{R_2^2} = \frac{100}{289} \approx 0,346$$

De esta forma, aunque el modelo de construcción del IC sugiera que los tres indicadores tienen la misma importancia, en realidad los indicadores de docencia y administración se

valoran más que las publicaciones y, por tanto, tienen una mayor influencia en el rendimiento del personal docente debido a la presencia de correlación entre ellos.

Este ejemplo describe las dos situaciones que se pueden presentar: la presencia de heterogeneidad de varianzas y la presencia de correlación en los indicadores simples. El problema de heterogeneidad de varianzas en los indicadores simples suele aparecer con frecuencia en la construcción de numerosos ICs, ya que muchos de ellos están contruidos mediante la normalización min-max que no asigna la misma varianza a los indicadores, sino que simplemente reduce a un rango de variación común. Por otra parte, si se corrige este inconveniente aplicando una normalización z-score, el problema puede generarse por la presencia de correlación entre los indicadores. Este último caso también suele estar presente en la mayoría de ICs, puesto que es muy común la existencia de correlación entre los indicadores simples.

Como consecuencia, no se puede asegurar que la asignación igualitaria de pesos al conjunto de indicadores simples sea una condición suficiente para que todos los indicadores tengan un mismo peso efectivo o efecto principal sobre el IC. De forma análoga, la asignación de pesos diferentes no es garantía suficiente para conseguir un efecto principal distinto de los indicadores sobre el IC. Esto último ocurre cuando todos los indicadores están perfectamente correlacionados. Independientemente del peso que se le asigne a cada uno de ellos, el efecto principal de cada indicador sobre el IC es el mismo, puesto que cada uno de ellos ordena de la misma forma las observaciones y el coeficiente de correlación de Spearman entre la ordenación del IC y las ordenaciones de cada uno de los indicadores simples es el mismo ($\rho = 1$).

Se concluye, por tanto, que el peso nominal asignado a cada indicador simple en el proceso de construcción del IC no se puede tratar como una medida de la importancia de cada indicador simple sobre el índice final cuando se emplean técnicas de agregación compensatorias. Es por eso que se plantea la necesidad de definir medidas de importancia que reflejen el efecto principal de cada indicador simple en la construcción del IC.

Tal y como se expone en el capítulo 3, los métodos GSA basados en el cálculo de varianzas presentan algunas ventajas sobre el resto de métodos GSA: se trata de métodos que se pueden aplicar independientemente de la forma funcional del modelo; permiten distinguir, mediante el uso de coeficientes de sensibilidad, los principales factores que afectan a la sensibilidad del IC; permiten distinguir los efectos principales de los efectos de orden superior; recogen la influencia de todo el rango de variación de cada factor input; permiten tratar a los factores input de forma agrupada y son fáciles de interpretar. Sin embargo, requieren un código de implementación elaborado, son computacionalmente costosos y asumen que toda la incertidumbre del factor es capturada por la varianza.

Los métodos basados en el cálculo de varianzas se centran, principalmente, en la estimación de los coeficientes de sensibilidad de cada factor input X_i sobre el factor output Y :

$$S_i = \frac{V(f_i(X_i))}{V(Y)} = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)}$$

donde $f_i(X_i)$ es el término aditivo de la descomposición de la función $Y = f(X_1, X_2, \dots, X_Q)$ mediante la representación HDMR (véase capítulo 3). Estos coeficientes de sensibilidad son los efectos principales o pesos efectivos y cuantifican la importancia de un factor input X_i sobre el factor output Y .

En este estudio se han considerado como factores input del modelo el conjunto de indicadores simples, mientras que el factor output se ha identificado con el IC. De esta forma, el modelo se ha formulado como $Y = IC = f(X_1, X_2, \dots, X_Q)$, donde X_i son los indicadores simples, $i = 1, \dots, Q$. El resto de fuentes de incertidumbre que aparecen en el esquema de construcción de un IC (transformación, normalización, agregación) se han fijado en los escenarios definidos en la tabla 6.20.

El objetivo del GSA aplicado a este estudio es estimar los coeficientes de sensibilidad de primer orden, o efectos principales, para cada indicador simple que reflejen el efecto de éste en el proceso de construcción del IC.

El efecto principal de cada indicador simple, S_i , se considera una medida de la importancia del indicador simple sobre el IC por las siguientes razones (Paruolo, Saltelli y Saisana, 2013):

- Proporciona una definición precisa del concepto de “importancia” de un indicador simple, ya que mide la reducción esperada en la varianza de un IC que se produce al fijar el indicador simple X_i , $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$
- Se trata de un modelo libre de supuestos. Se puede usar incluso cuando la técnica de agregación utilizada para la construcción del IC no es lineal. Por tanto, a diferencia del coeficiente de correlación de Pearson o Spearman no se ve limitado por los supuestos de linealidad.

La estimación de los coeficientes de sensibilidad es un tema de investigación reciente. En la tabla 6.21 se enumeran distintos métodos clásicos de GSA, basados en el cálculo de varianzas, que se pueden aplicar. Saltelli et al. (2008) describen, de forma detallada, el procedimiento analítico de cada uno de estos métodos. Además, algunos de ellos están implementados en el software libre SIMLAB (Simlab, 2011):

MÉTODO	REFERENCIA
FAST (Fourier Amplitude Sensitivity Test)	(Cukier et al., 1973)
El método de Sobol'	(Sobol', 1993)
El método de Jansen	(Jansen, 1994)
Cálculo del Índice de Sensibilidad de orden Total	(Homma y Saltelli, 1996)
Extensión del método FAST (EFAST)	(Saltelli, Tarantola y Chan, 1999)
Una mejora del método de Sobol'	(Saltelli, 2002)
Diseño Random Balance	(Tarantola, Gatelli y Mara, 2006)
Una mejora del método de Sobol' para índices pequeños	(Sobol' et al., 2007)
Buenas prácticas del cálculo de los índices de primer orden y de orden total	(Saltelli et al., 2010)

Tabla 6.21. Métodos GSA basados en el cálculo de varianzas

Las principales limitaciones de la estimación de los coeficientes S_i mediante la aplicación de los métodos anteriores son las siguientes (Saltelli et al., 2008):

- Requieren independencia en los factores input
- Requieren un código de implementación elaborado
- Son más costosos de calcular computacionalmente
- Requieren la generación de un gran número de simulaciones de los factores input y, por tanto, de una gran cantidad de evaluaciones. Es decir, asignan la incertidumbre del factor output mediante el enfoque de muestreo de funciones de densidad de probabilidad (PDFs) asociadas a los factores input

Los métodos más recientes en la literatura de GSA elaborados por Ratto, Pagano y Young (2007) y Paruolo, Saltelli y Saisana (2013) permiten realizar GSA basado en el cálculo de varianzas mediante la estimación de un meta-modelo que aproxima analíticamente el mapeo entre $IC = f(X_1, X_2, \dots, X_Q)$ y cada uno de los indicadores X_i para el cálculo de los efectos principales o pesos efectivos S_i .

Las técnicas de meta-modelado presentan las siguientes ventajas respecto a las técnicas clásicas de GSA basadas en el cálculo de varianzas (Ratto, Pagano y Young, 2007; Saltelli et al., 2008; Paruolo, Saltelli y Saisana, 2013):

- El coste computacional es menor. Se trata de técnicas que convergen mucho más rápido y, por tanto, son más eficientes
- Se pueden aplicar independientemente del grado de correlación de los factores input
- Se pueden aplicar cuando se dispone de factores input observables, sin necesidad de generar simulaciones de sus PDFs
- Se trata de métodos de GSA que pueden ser “no invasivos”⁷ si se utiliza la información observable de los factores input, por lo que su aplicación no modifica el valor final del

⁷ Un método de análisis de sensibilidad se define como “invasivo” cuando las fuentes de incertidumbre input se tienen que modelar explícitamente y se generan muestras de cada intervalo de incertidumbre que afectan a la construcción del factor output

factor output. Esto no ocurre con otros métodos de análisis de sensibilidad global donde se generan simulaciones de los factores input (Saisana, Saltelli y Tarantola, 2005)

Por tanto, su uso es adecuado en el campo de los ICs donde los indicadores suelen estar correlacionados. En este caso $\sum_{i=1}^Q S_i$ suele exceder de la unidad. Esto último ocurre ya que al fijar un indicador simple también se reduce el grado de variación de las variables correlacionadas y, por tanto, se obtiene una reducción mayor en la varianza $V_{X_i}(E_{X_{-i}}(Y|X_i))$ debida a los efectos combinados de los indicadores (Saltelli et al., 2008). Ante esta situación se deben normalizar los coeficientes S_i de tal forma que la suma sea la unidad.

La medida de importancia S_i se puede comparar con los pesos nominales w_i asignados originalmente a cada indicador o con el conjunto de pesos objetivo que representa la importancia deseable de cada indicador sobre el IC (Paruolo, Saltelli y Saisana, 2013). De esta forma, se dispone de una herramienta válida para evaluar el impacto de la metodología de construcción de cada escenario y para medir la calidad del IC evaluado.

En este estudio los coeficientes de sensibilidad se ha estimado usando una técnica de meta-modelo no paramétrica llamada “*State Dependent Parameter Metamodelling*” (SDP) muy parecida al método de cálculo de los “*smoothing splines*” y al método de regresión de Kernel, pero basada en el filtro de Kalman y en técnicas recursivas de *smoothing* de la forma *Fix Interval Smoothing* (FIS) (Ratto, Pagano y Young, 2007; Saltelli et al., 2008; Saltelli et al., 2010; Young, 2011).

El objetivo principal de la técnica SDP para el cálculo de los efectos principales es estimar la curva $E_{X_{-i}}(Y|X_i)$. Un ejemplo de esta estimación viene dado por la curva roja de la figura siguiente:

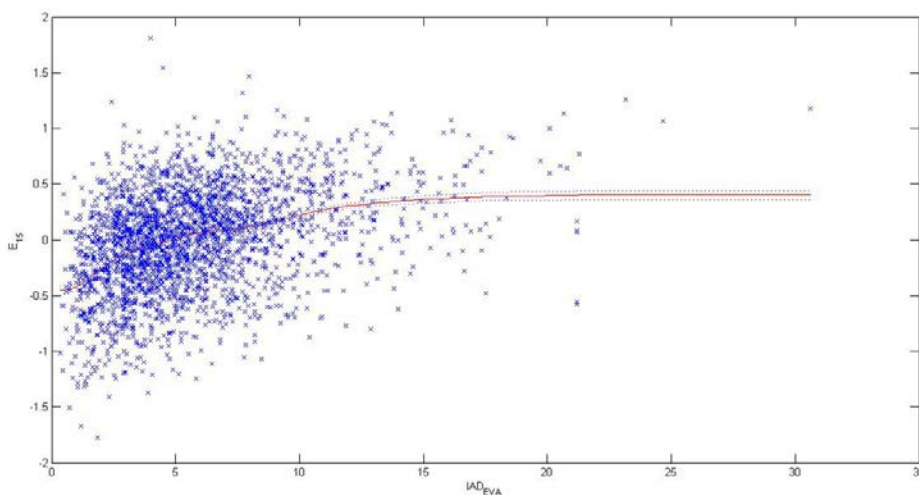


Figura 6.34. Estimación no lineal de IAD_{EVA} frente a E15

A modo de ejemplo, se ha estimado la función no lineal $E_{\mathbf{X}_{-i}}(Y|X_i)$ entre el indicador IAD_{EVA} y el IC del escenario E15. Para ello, en primer lugar se ha aplicado el filtro de Kalman recursivo con el objetivo de realizar una primera estimación de la curva $E_{\mathbf{X}_{-i}}(Y|X_i)$. El filtro de Kalman es un método de estimación cuyos parámetros se corrigen en cada iteración dependiendo del error de predicción que se haya cometido en la iteración anterior. Se trata de una técnica de estimación que ha ganado aceptación en el análisis de series temporales y permite identificar la señal de una observación que está contaminada por ruido. Una vez realizada una primera estimación de la curva, se ha aplicado una técnica de suavizado o *smoothing* llamada *Fix Interval Smoothing* (FIS). El objetivo de la técnica de *smoothing* es disminuir la variabilidad de las estimaciones obtenidas mediante el filtro de Kalman. En este caso, las estimaciones se realizan teniendo en cuenta todo el conjunto de observaciones finito N (conjunto de profesores) para cada indicador X_i . El procedimiento utilizado se llama *smoother* que en términos generales se basa en obtener un promedio local asignando ponderaciones a las observaciones. A continuación se expone de forma más detallada la metodología utilizada.

Metodología State Dependent Parameter - SDP⁸

El objetivo principal para el cálculo de los efectos principales S_i es estimar los términos de la descomposición HDMR mediante la aplicación del filtro de Kalman y del algoritmo recursivo de suavizado conocido como *Fix Interval Smoothing*, FIS, que calcula los parámetros de la formulación del mapeo input-output mediante un modelo de parámetro estado-dependiente (SDP). La metodología utilizada recoge los fundamentos básicos del procesamiento de señales y del análisis de series temporales.

Considérense los términos de primer orden de la descomposición HDMR (véase capítulo 3) del modelo $Y = f(\mathbf{X})$,

$$Y_t - f_0 = f_1(X_{1,t}) + f_2(X_{2,t}) + \dots + f_Q(X_{Q,t}) + o(\mathbf{X}\mathbf{X}')$$

La representación anterior puede expresarse como un modelo *State Dependent Regression*, SDR (Young P., 2001; Ratto, Pagano y Young, 2007),

$$Y_t - f_0 = \mathbf{X}_t^T \mathbf{p}_t + e_t = p_{1,t}X_{1,t} + p_{2,t}X_{2,t} + \dots + p_{Q,t}X_{Q,t} + e_t, \quad e_t = N(0, \sigma^2), \quad (6.1)$$

donde se asume que todos los términos de orden mayor que uno se pueden aproximar por un ruido blanco Gaussiano con media cero y varianza σ^2 . Es decir, la descomposición HDMR truncada se puede ver como un sistema no lineal estocástico. Esta última afirmación se puede justificar con la aplicación del teorema central del límite, ya que los términos de mayor orden se

⁸ En este apartado las letras en negrita denotan vectores o matrices y el símbolo " $\hat{}$ " denota las estimaciones de los parámetros

pueden considerar como suma de un gran número de variables aleatorias independientes con media cero y distribución de probabilidad arbitraria. Se adopta la notación de series temporales donde t indica la expansión de los valores de $X_{i,t}$ con $t = 1, \dots, N$.

Cada parámetro de estado-dependiente (*State Dependent Parameter* - SDP), $p_{i,t}$, es función del correspondiente factor input $X_{i,t}$, $f_i(X_{i,t}) = p_{i,t}X_{i,t}$, $i = 1, \dots, Q$. Por tanto, estimar los parámetros SDP es equivalente a proporcionar un mapeo de los términos de primer orden de la descomposición HDMR. Según la equivalencia anterior $f_i(X_{i,t}) = p_{i,t}X_{i,t}$, la estimación de los parámetros SDP puede ser inconsistente si alguno de los factores $X_{i,t}$ toma el valor cero. Esta limitación se puede resolver introduciendo una transformación sobre X_i , de tal forma que se reformula el modelo introduciendo regresores constantes igual a la unidad:

$$Y_t - f_0 = \mathbf{1}_t^T \mathbf{p}_t + e_t = p_{1,t} + p_{2,t} + \dots + p_{Q,t} + e_t, \quad e_t = N(0, \sigma^2). \quad (6.2)$$

Para estimar los parámetros $p_{i,t}$ es necesario caracterizar su variabilidad de forma estocástica. Para ello, se asume que la evolución de cada SDP puede caracterizarse por una secuencia no estacionaria conocida como *Generalized Random Walk* (GRW). En este contexto, el procedimiento *Integrated Random Walk* (IRW) proporciona el mejor resultado ya que asegura que el SDP estimado tenga las propiedades de suavizado de un *spline* cúbico. La caracterización del IRW sobre los parámetros $p_{i,t}$ se formula como sigue:

$$\begin{aligned} \text{Ecuación observada: } Y_t &= \mathbf{Z}_t \mathbf{p}_t + e_t \\ \text{Ecuación de estado: } \mathbf{p}_t &= \mathbf{p}_{t-1} + \mathbf{d}_{t-1}, \\ \mathbf{d}_t &= \mathbf{d}_{t-1} + \boldsymbol{\eta}_{t-1} \end{aligned}$$

donde \mathbf{Z}_t es \mathbf{X}_t en (6.1) o $\mathbf{1}_t$ en (6.2); mientras que e_t y $\boldsymbol{\eta}_t$, $i = 1, 2, \dots, Q$, tienen media cero y varianza σ^2 y \mathbf{Q}_a , respectivamente, siendo $E\{\boldsymbol{\eta}_t, \boldsymbol{\eta}_j\} = \mathbf{Q}_a \delta_{t,j}$ con δ la función delta de Kronecker definida como:

$$\delta_{t,j} := \begin{cases} 0 & \text{si } t \neq j \\ 1 & \text{si } t = j \end{cases}$$

Los parámetros SDP se pueden estimar aplicando el algoritmo de Kalman (KF) y el algoritmo recursivo de suavizado *Fix Interval Smoothing* (FIS). El procedimiento basado en ambos algoritmos se denota con el nombre KALMSMO, véase el cuarto capítulo de Young (2011).

El algoritmo KALMSMO se basa en la estimación por mínimos cuadrados temporal de los parámetros $p_{i,t}$. Como en este estudio no tiene sentido aplicar la temporalidad en el algoritmo los datos se ordenan para que los parámetros puedan estimarse bajo la hipótesis de que varían según cierto criterio que satisface las suposiciones del modelo IRW en un espacio ordenado.

Para ello, la estimación de cada parámetro $p_{i,t}$ se realiza ordenando los datos según su correspondiente factor $X_{i,t}$.

A continuación, se formula la ecuación de estado definida en el IRW de forma matricial para cada parámetro:

$$\begin{pmatrix} p_{i,t} \\ d_{i,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_{i,t-1} \\ d_{i,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \eta_{i,t-1} \end{pmatrix}$$

Con el fin de usar la misma notación que la empleada en la descripción del algoritmo KALMSMO por Young (2011) se define la ecuación observada y la ecuación de estado como sigue:

$$\text{Ecuación observada: } Y_t = \mathbf{h}_t^T \mathbf{x}_t + e_t$$

$$\text{Ecuación de estado: } \mathbf{x}_t = \mathbf{A} \mathbf{x}_{t-1} + \mathbf{D} \boldsymbol{\eta}_{t-1}$$

donde:

$$\mathbf{h}_t^T := [(X_{1,t} 0) \dots (X_{Q,t} 0)]_{1 \times 2Q}$$

$$\mathbf{x}_t := [(p_{1,t} d_{1,t}) \dots (p_{Q,t} d_{Q,t})]_{1 \times 2Q}^T$$

$$\mathbf{A} := \begin{pmatrix} \mathbf{A}_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{pmatrix}_{2Q \times 2Q}$$

$$\mathbf{D} := \begin{pmatrix} \mathbf{D}_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \mathbf{D}_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}_{2Q \times 2Q}$$

$$\boldsymbol{\eta}_{t-1} := [(0 \ \eta_{1,t-1}) \dots (0 \ \eta_{Q,t-1})]_{1 \times 2Q}^T$$

Por tanto, el algoritmo KALMSMO compuesto por una etapa de predicción y de corrección (filtro de Kalman) y una etapa de suavizado (algoritmo FIS) se describe a continuación:

Predicción:

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= \mathbf{A} \hat{\mathbf{x}}_{t-1} \\ \mathbf{P}_{t|t-1} &= \mathbf{A} \mathbf{P}_{t-1} \mathbf{A}^T + \mathbf{D} \mathbf{Q}_{nvr} \mathbf{D}^T \end{aligned}$$

Corrección:

$$\begin{aligned} \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{g}_t \{Y_t - \mathbf{h}_t^T \hat{\mathbf{x}}_{t|t-1}\} \\ \mathbf{g}_t &= \mathbf{P}_{t|t-1} \mathbf{h}_t [1 + \mathbf{h}_t^T \mathbf{P}_{t|t-1} \mathbf{h}_t]^{-1} \\ \mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{g}_t \mathbf{h}_t^T \mathbf{P}_{t|t-1} \\ \mathbf{P}_t^* &= \hat{\sigma}^2 \mathbf{P}_t \end{aligned}$$

Suavizado:

$$\begin{aligned} \hat{x}_{t|N} &= \hat{x}_t - \mathbf{P}_t^* \mathbf{A}^T \boldsymbol{\lambda}_t \\ \boldsymbol{\lambda}_{t-1} &= [\mathbf{I}_p - \mathbf{P}_t^* \frac{\mathbf{h}_t \mathbf{h}_t^T}{\hat{\sigma}^2}]^T \{ \mathbf{A}^T \boldsymbol{\lambda}_t - \frac{\mathbf{h}_t}{\hat{\sigma}^2} [Y_t - \mathbf{h}_t^T \mathbf{A} \hat{x}_{t-1}] \} \\ &\text{con } \boldsymbol{\lambda}_N = 0 \text{ y } p = 2N \\ \mathbf{P}_{t|N}^* &= \mathbf{P}_t^* + \mathbf{P}_t^* \mathbf{A} \mathbf{P}_{t+1|N}^{*-1} [\mathbf{P}_{t+1|N}^* - \mathbf{P}_{t+1|t}^*] \mathbf{P}_{t+1|t}^{*-1} \mathbf{A} \mathbf{P}_t^* \end{aligned}$$

En cada una de las iteraciones es necesario optimizar los hiper-parámetros asociados a la ecuación de estado. Estos son las varianzas σ^2 y $\sigma_{\eta_{i,t}}^2$. Con una pequeña reformulación de los algoritmos KF y FIS, el modelo se puede caracterizar por un único hiper-parámetro de ruido blanco $NVR_i = \sigma_{\eta_{i,t}}^2 / \sigma^2$. Los valores de NVR_i son desconocidos *a priori* y se pueden estimar mediante la optimización de máxima verosimilitud usando la descomposición de los errores de predicción (pág. 81(Young, 2011)).

Sea $\hat{p}_{i,t|N} = \hat{f}_i(X_{i,t|N})$ la estimación del parámetro $p_{i,t}$. Una vez estimados los términos de la descomposición HDMR \hat{f}_i , el cálculo de los índices $S_i = V_{X_i}[E_{\mathbf{X}_{\sim i}}(Y|X_i)]/V(Y)$ es directa. Dado el tamaño de la muestra N , el mejor estimador definido para S_i es (Doksum y Samarov, 1995):

$$\hat{s}_i = \frac{N^{-1} \sum_{s=1}^N (\hat{f}_i(X_{i,s}) - \bar{f})^2}{\sigma_Y^2}$$

donde: $\bar{f} = N^{-1} \sum \hat{f}_i(X_{i,s})$ y $\sigma_Y^2 = N^{-1} \sum (Y_s - \bar{Y})^2$.

Una aplicación práctica del cálculo de los efectos principales, S_i , estimados mediante un procedimiento no paramétrico de regresión local-lineal por kernel muy parecido a la técnica SDP aplicada en este estudio se puede encontrar en (Paruolo, Saltelli y Saisana, 2013). En dicho estudio se han analizado dos de los rankings más conocidos de IES, el *Academic Ranking of World Universities* y el *Times Higher Education Supplement*, y el Índice de Desarrollo Humano, haciendo una comparativa entre los pesos asignados por sus diseñadores en el desarrollo del IC y los efectos principales S_i .

Resultados del Análisis de Sensibilidad Global aplicado a los escenarios de trabajo

El cálculo de los efectos principales S_i se ha realizado con el paquete de computación SS-ANOVA-R de la técnica SDP en el software MATLAB 7.12.0.635 (R2011a). Este código ha sido elaborado conjuntamente por la Universidad de Lancaster y el Centro Común de Investigación - *Joint Research Centre* (Institute for the Protection and Security of Citizen, 2013).

Se ha considerado adecuada la aplicación de la técnica SDP de GSA para la estimación de los efectos principales por las siguientes razones:

1. Se puede aplicar independientemente del grado de correlación existente entre los factores input. En el caso de estudio los factores input presentan un cierto grado de correlación entre algunos de los indicadores, tal y como se ha visto en la tabla 6.4
2. Se puede aplicar incluso cuando la técnica de agregación utilizada en el modelo no es lineal. En el caso de estudio se han definido escenarios de ICs que siguen modelos no lineales en los indicadores simples. Estos escenarios son EUPV, E3, E6, E9, E10, E11, E12, E13 y E14
3. Se trata de un método de modelización mucho más eficiente en términos de coste computacional, ya que los métodos clásicos requieren de miles de evaluaciones del modelo y, por tanto, de costosos modelos computacionales
4. Se trata de un método de GSA que se puede utilizar con datos observables de los factores input sin la necesidad de generar muestras en las distribuciones de densidad de probabilidad de los intervalos de incertidumbre de los factores input

Dado que los indicadores simples del modelo presentan un cierto grado de correlación, se obtiene $\sum_{i=1}^Q S_i > 1$, por tanto se ha normalizado cada uno de los efectos para que la suma sea la unidad y, de esta manera, se puedan comparar con el conjunto de pesos objetivo proporcionado por el grupo de expertos de la CED.

A continuación se muestran los resultados del análisis de sensibilidad global aplicado a cinco de los escenarios de trabajo propuestos. Además de mostrar los resultados del análisis aplicado al escenario de la UPV, se muestran aquellos escenarios que han resultado tener un conjunto de pesos efectivos similar al conjunto de pesos objetivo propuesto por el grupo de expertos. Los resultados del análisis de sensibilidad realizado sobre el resto de escenarios se exponen en el anexo B.

ESCENARIO EUPV

En primer lugar, se ha aplicado el procedimiento de estimación no paramétrico al escenario de la UPV construido a partir de un modelo no lineal mediante el siguiente esquema de agregación:

$$IC_{UPV} = (0,3 * IAD_{FORMA} + 0,3 * IAD_{GUIA} + 0,3 * IAD_{INFORMA} + 0,3 * IAD_{MAT} + 0,7 * IAD_{DOC} + 0,7 * IAD_{TUT} + 0,7 * IAD_{EVA} + 0,7 * IAD_{OTRAS})(0,6 * K_1 + 0,3 * K_2 + 0,05 * K_3 + 0,05 * K_4)$$

Además, según su formulación se trata de un modelo no aditivo, puesto que existen interacciones entre los factores input del modelo.

En la tabla 6.22 se muestran los efectos principales que se han obtenido mediante el análisis de sensibilidad global aplicado al escenario de la UPV.

EUPV	S_i normalizado	Efectos principales S_i	Pesos objetivo	Diferencia
IAD_{FORMA}	0,062	6,2	12,0	-5,8
IAD_{GUIA}	0,098	9,8	8,0	1,8
$IAD_{INFORMA}$	0,013	1,3	5,0	-3,7
IAD_{MAT}	0,058	5,8	10,0	-4,2
IAD_{DOC}	0,250	25,0	15,0	10,0
IAD_{TUT}	0,132	13,2	7,0	6,2
IAD_{EVA}	0,129	12,9	8,0	4,9
IAD_{OTRAS}	0,168	16,8	10,0	6,8
K_1	0,073	7,3	18,0	-10,7
K_2	0,007	0,7	5,0	-4,3
K_3	0,000	0,0	1,0	-1,0
K_4	0,010	1,1	1,0	0,1
$ \Sigma $	1,000	100,0	100,0	

Tabla 6.22. Análisis de sensibilidad global aplicado al escenario EUPV

El coeficiente de correlación del modelo aditivo resultante es de:

$$R_{Y,aditivo}^2 = \frac{\sum_{t=1}^N (\sum_{i=1}^Q f_i(X_{ti}) - \bar{Y})^2}{\sum_{t=1}^N (Y_t - \bar{Y})^2} = 0,9965$$

donde Y es el IC, N es el número de profesores evaluados, Q es el número de indicadores simples, \bar{Y} es la media de Y , y $f_i(X_{ti})$ es una función que sólo depende del indicador X_i evaluada en el profesor t .

Este resultado informa que el modelo aditivo explica prácticamente toda la variabilidad del IC y, por tanto, no es necesario estudiar el efecto de las interacciones que viene dado por las funciones $(f_{i,j}(X_i, X_j), \dots, f_{1,2,\dots,Q}(X_1, \dots, X_Q))$, puesto que éstas explican una mínima parte de la varianza incondicional del IC, $V(Y)$.

Por tanto, el modelo UPV se resume al siguiente modelo aditivo, ya que la parte de interacción de las variables no aporta información relevante al Índice de Actividad Docente:

$$Y = IC \simeq f_{IAD_{FORMA}}(IAD_{FORMA}) + f_{IAD_{GUIA}}(IAD_{GUIA}) + \dots + f_{K_4}(K_4)$$

La tabla 6.22 muestra que los indicadores con mayor efecto principal sobre la formación del IC, con un total del 67,9%, son los indicadores de la dimensión *Desarrollo de la Enseñanza*

(IAD_{DOC} , IAD_{TUT} , IAD_{EVA} , IAD_{OTRAS}). Por el contrario, los indicadores de la dimensión *Resultados* (K_1 , K_2 , K_3 , K_4) tiene un efecto del 9,1% en la formación del índice.

La falta de normalización en los datos (normalización *min-max* que transforma los indicadores en un mismo rango de variación [0,1], o normalización *z-score* que transforma los indicadores en una misma escala adimensional con media 0 y varianza 1) puede ser causante de los resultados obtenidos en la tabla 6.22. Como consecuencia, los indicadores con mayor impacto sobre la formación del IC presentan un rango de variación y una varianza mucho mayor que el resto (véase tabla 6.2 y 6.22).

Si se comparan, por ejemplo, los indicadores IAD_{DOC} y K_2 se observa que el efecto o importancia que tiene el primero en la formación del IC_{UPV} es del 25%, mientras que la importancia del segundo en la formación del índice es de aproximadamente 0,7%. Por otro lado, se observa que la varianza del indicador K_2 ($\sigma_{K_2}^2 = 0,01$) es prácticamente despreciable comparada con la varianza de IAD_{DOC} ($\sigma_{IAD_{DOC}}^2 = 41,34$). Se concluye, por tanto, que el impacto que tienen los indicadores simples en la formación del IC viene condicionado por la varianza de los indicadores, por lo que sería conveniente la aplicación de alguna técnica de normalización.

Además, el tipo de ponderación utilizado en la construcción del IC_{UPV} también justifica los resultados obtenidos. En la construcción del índice se asignan pesos dentro de la primera dimensión que suman 1,2 puntos, mientras que los pesos asignados en la segunda dimensión suman 2,8. Por otro lado, en la tercera dimensión se distribuye un total de 1 punto en pesos. De esta forma, se asigna un peso más elevado a los indicadores de la segunda dimensión y este resultado también se refleja en los efectos principales obtenidos para los indicadores de dicha dimensión.

Si se compara el conjunto de efectos principales con el conjunto de pesos objetivo se observa que las diferencias son significativas para la mayoría de los indicadores. Por ejemplo, según el GSA, el indicador IAD_{DOC} tiene un efecto de un 25% en la formación del IC, mientras que según el grupo de expertos, la importancia deseada para este indicador es de un 15%. Esto significa que el indicador IAD_{DOC} tiene un mayor impacto en la formación del IC que el esperado o deseado por el grupo de expertos.

Por otra parte, el indicador de resultados K_1 que tiene un efecto principal de 7,3% es el que más impacto debería tener en la formación del IC con un peso del 18%.

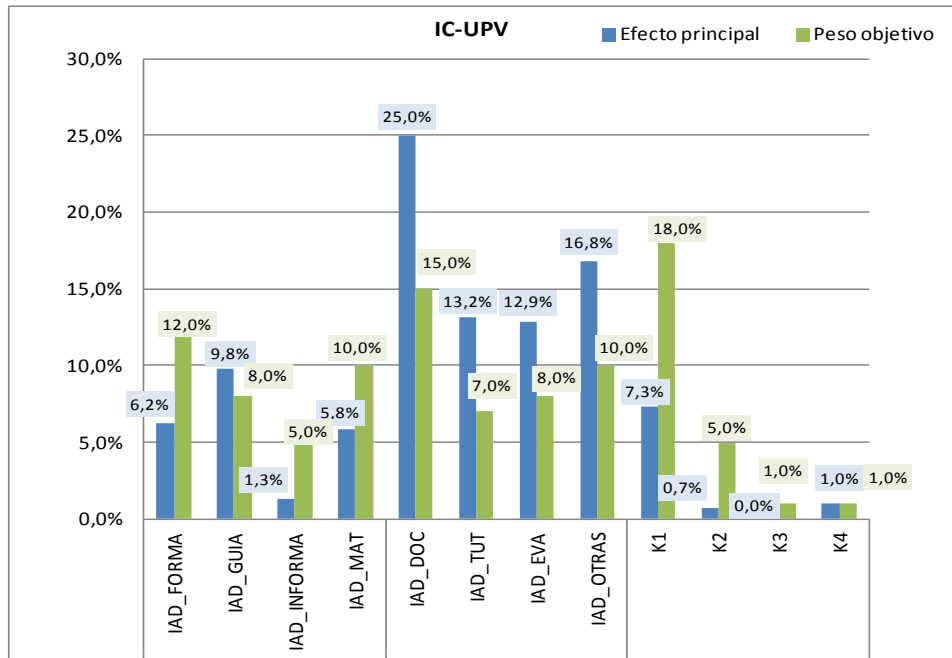


Figura 6.35. Efectos principales y pesos objetivo de los indicadores simples del IC_{UPV}

Si la UPV se replantease la construcción del indicador compuesto vigente y tuviese en consideración el conjunto de pesos objetivo para su construcción, se evidenciaría la necesidad de reestructurar la metodología utilizada, puesto que los resultados difieren del conjunto de pesos objetivo que muestra la importancia deseada de cada indicador en la formación del IC. El indicador IAD_{DOC} debería tener un menor efecto en la formación del IC, mientras que el indicador de resultados K_1 es el que mayor efecto debería tener sobre el IC, según el conjunto de pesos objetivo desarrollado por el grupo de expertos.

ESCENARIO E1

A continuación se presentan los resultados del escenario E1 en el que se ha aplicado una normalización z-score sobre los datos y el IC se ha construido mediante el siguiente modelo:

$$IC_{E1} = \frac{1}{3} * \left(\frac{1}{4} * IAD_{FORMA} + \frac{1}{4} * IAD_{GUIA} + \frac{1}{4} * IAD_{INFORMA} + \frac{1}{4} * IAD_{MAT} \right) + \frac{1}{3} * \left(\frac{1}{4} * IAD_{DOC} + \frac{1}{4} * IAD_{TUT} + \frac{1}{4} * IAD_{EVA} + \frac{1}{4} * IAD_{OTRAS} \right) + \frac{1}{3} * \left(\frac{1}{3} * K_1 + \frac{1}{3} * K_2 + \frac{1}{3} * K_4 \right)$$

E1	S_i normalizado	Efectos principales S_i	Pesos objetivo	Diferencia
IAD_{FORMA}	0,121	12,1	12,0	0,1
IAD_{GUIA}	0,090	9,0	8,0	1,0
$IAD_{INFORMA}$	0,061	6,1	5,0	1,1
IAD_{MAT}	0,084	8,4	10,0	-1,6
IAD_{DOC}	0,111	11,1	15,0	-3,9
IAD_{TUT}	0,129	12,9	7,0	5,9
IAD_{EVA}	0,057	5,7	8,0	-2,3
IAD_{OTRAS}	0,095	9,5	10,0	-0,5
K_1	0,138	13,8	18,0	-4,2
K_2	0,042	4,2	5,0	-0,8
K_3	0,000	0,0	1,0	-1,0
K_4	0,072	7,2	1,0	6,2
$ \Sigma $	1,000	100,0	100,0	

Tabla 6.23. Análisis de sensibilidad global aplicado al escenario E1

A partir de la tabla 6.23 se concluye que a pesar de asignar el mismo peso a cada una de las dimensiones y el mismo peso a los indicadores dentro de cada dimensión, el efecto principal de cada indicador difiere del peso nominal (véase figura 6.36), pero se consigue un conjunto de efectos principales con una diferencia menor que en el escenario de la UPV con respecto al conjunto de pesos objetivo formulado por el grupo de expertos, exceptuando estos resultados para los indicadores IAD_{TUT} y K_4 . Este resultado muestra que aunque se ponderen de forma equitativa las dimensiones del IC, los efectos principales de los indicadores sobre la formación del IC pueden ser diferentes debido a las correlaciones entre los indicadores simples. En este caso sí se pueden comparar los efectos principales con los pesos nominales, puesto que la forma lineal del modelo del escenario permite identificar un peso nominal asignado a cada indicador simple.

En la figura 6.36 se muestran de forma más clara las diferencias entre el conjunto de pesos nominales asignado en la construcción del índice, los efectos principales y el conjunto de pesos objetivo de cada indicador simple.

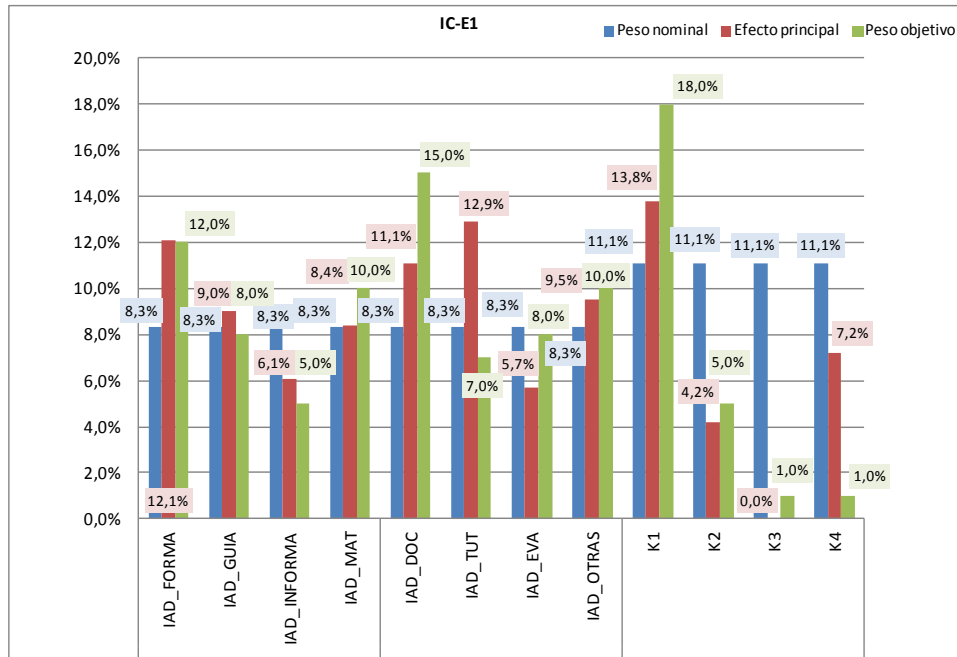


Figura 6.36. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC_{E1}

ESCENARIO E11

En la tabla 6.24 se muestran los resultados del escenario E11 en el que se ha aplicado una normalización min-max y se ha construido mediante el siguiente modelo:

$$IC_{E11} = (0,343 * IAD_{FORMA} + 0,228 * IAD_{GUIA} + 0,143 * IAD_{INFORMA} + 0,286 * IAD_{MAT})^{0,35} * (0,375 * IAD_{DOC} + 0,175 * IAD_{TUT} + 0,2 * IAD_{EVA} + 0,25 * IAD_{OTRAS})^{0,4} * (0,76 * K_1 + 0,2 * K_2 + 0,04 * K_4)^{0,25}$$

El coeficiente de correlación del modelo aditivo es $R_{Y,aditivo}^2 = 0,9879$, por lo que las interacciones entre los indicadores prácticamente no contribuyen en la formación del IC y es suficiente el cálculo de los efectos principales de los indicadores simples para analizar el impacto o importancia que tiene cada uno de ellos en la formación del IC.

E11	S_i normalizado	Efectos principales S_i	Pesos objetivo	Diferencia
IAD_{FORMA}	0,182	18,2	12,0	6,2
IAD_{GUIA}	0,074	7,4	8,0	-0,6
$IAD_{INFORMA}$	0,048	4,8	5,0	-0,2
IAD_{MAT}	0,111	11,1	10,0	1,1
IAD_{DOC}	0,097	9,7	15,0	-5,3
IAD_{TUT}	0,116	11,6	7,0	4,6
IAD_{EVA}	0,039	3,9	8,0	-4,1
IAD_{OTRAS}	0,114	11,4	10,0	1,4
K_1	0,196	19,6	18,0	1,6
K_2	0,009	0,9	5,0	-4,1
K_3	0,000	0,0	1,0	-1,0
K_4	0,014	1,4	1,0	0,4
$ \Sigma $	1,000	100,0	100,0	

Tabla 6.24. Análisis de sensibilidad global aplicado al escenario E11

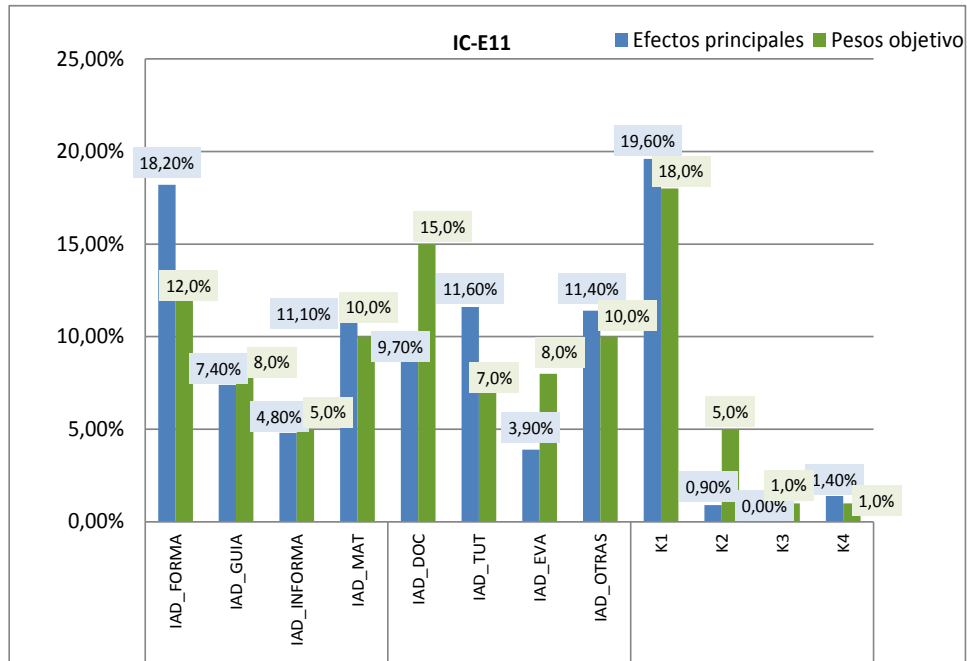


Figura 6.37. Efectos principales y pesos objetivo de los indicadores simples del IC_{E11}

Tal y como se aprecia en la tabla 6.24 y en la figura 6.37, el indicador que mide la formación del profesorado, IAD_{FORMA} , tiene un efecto principal mucho mayor que el deseado por el grupo de expertos, mientras que el indicador de docencia impartida, IAD_{DOC} , debería tener un mayor efecto en la construcción del índice.

ESCENARIO E7

En la tabla 6.25 se muestran los mismos resultados para el escenario E7 en el que se ha aplicado una normalización z-score y se ha construido mediante el siguiente modelo:

$$IC_{E7} = \frac{12}{100} * IAD_{FORMA} + \frac{8}{100} * IAD_{GUIA} + \frac{5}{100} * IAD_{INFORMA} + \frac{10}{100} * IAD_{MAT} + \frac{15}{100} * IAD_{DOC} + \frac{7}{100} * IAD_{TUT} + \frac{8}{100} * IAD_{EVA} + \frac{10}{100} * IAD_{OTRAS} + \frac{19}{100} * K_1 + \frac{5}{100} * K_2 + \frac{1}{100} * K_4$$

E7	S_i normalizado	Efectos principales S_i	Pesos objetivo	Diferencia
IAD_{FORMA}	0,168	16,8	12,0	4,8
IAD_{GUIA}	0,088	8,8	8,0	0,8
$IAD_{INFORMA}$	0,033	3,3	5,0	-1,7
IAD_{MAT}	0,083	8,3	10,0	-1,7
IAD_{DOC}	0,128	12,8	15,0	-2,2
IAD_{TUT}	0,110	11,0	7,0	4,0
IAD_{EVA}	0,055	5,5	8,0	-2,5
IAD_{OTRAS}	0,101	10,1	10,0	0,1
K_1	0,206	20,6	18,0	2,6
K_2	0,012	1,2	5,0	-3,8
K_3	0,000	0,0	1,0	-1,0
K_4	0,016	1,6	1,0	0,6
$ \Sigma $	1,000	100,0	100,0	

Tabla 6.25. Análisis de sensibilidad aplicado al escenario E7

En este caso, al asignar el peso propuesto por el grupo de expertos a cada uno de los indicadores previamente estandarizados y aplicar una agregación de tipo lineal compensatoria, se obtiene un mejor ajuste entre el conjunto de efectos principales y el conjunto de pesos objetivo (véase tabla 6.25 y figura 6.38).

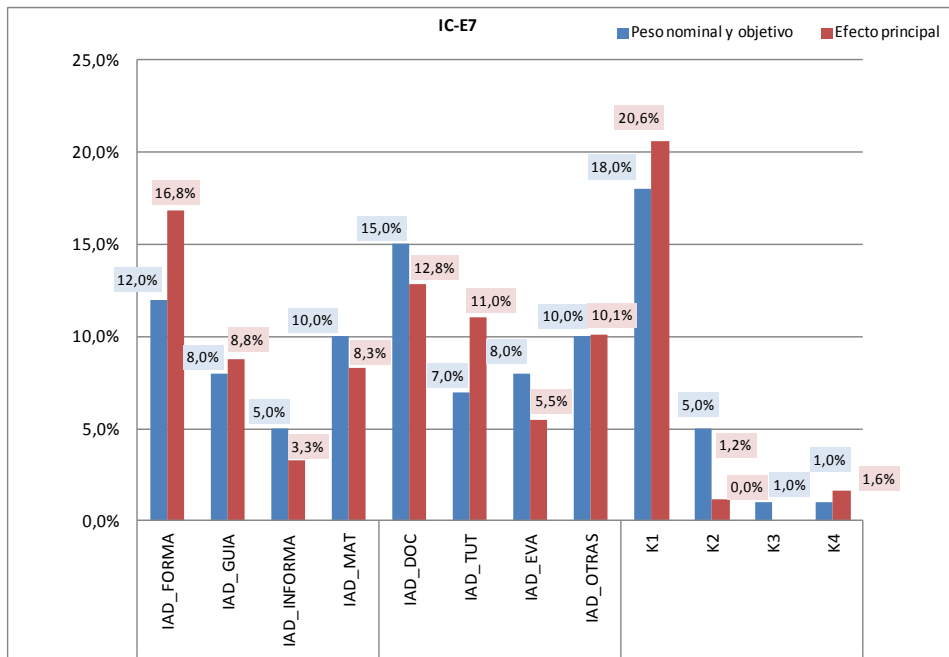


Figura 6.38. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC_{E7}

Paruolo, Saltelli y Saisana (2013) definen en su estudio una medida para la obtención de los pesos nominales de tal forma que los efectos principales sean los deseados. Sin embargo, uno de los supuestos para su aplicación es que $f_j(X_j)$ sea lineal en X_j . Los autores argumentan que encontrar una fórmula para el caso en el que no se cumpla esta linealidad supone un elevado coste computacional, por lo que no se ha definido ninguna medida hasta el momento. En el escenario E7, la mayoría de los indicadores simples no cumplen este supuesto de linealidad. Véase, por ejemplo, dicha relación con el indicador IAD_{DOC} .

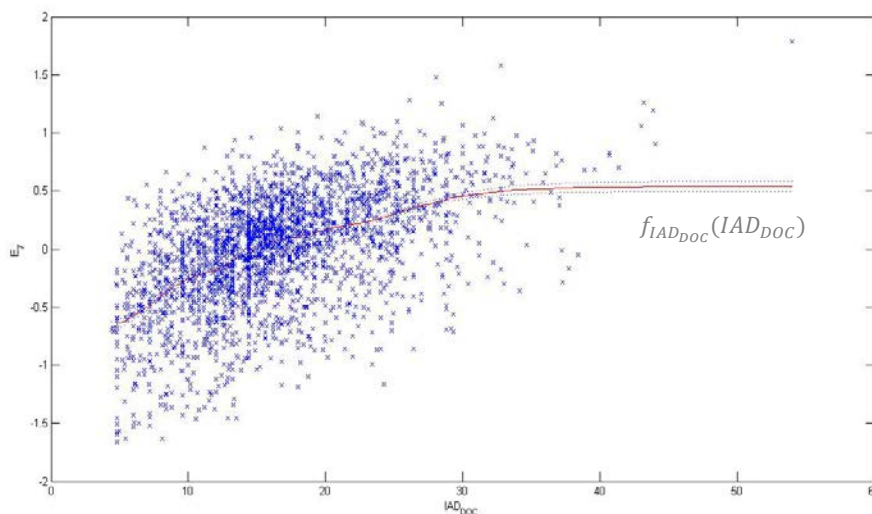


Figura 6.39. Relación entre $f_{IAD_{DOC}}(IAD_{DOC})$ y IAD_{DOC}

Por tanto, se ha construido un nuevo escenario E15 en base a la metodología de construcción del escenario E7, que es el que más se aproxima al objetivo del estudio según el grupo de expertos. De esta forma, se ha asignado un conjunto de pesos nominales para la construcción del indicador compuesto de modo que la diferencia entre el conjunto de efectos principales y el conjunto de pesos objetivo sea mínima. Así, la importancia de cada indicador sobre el IC es la deseada según el grupo de expertos.

No se trata del único escenario que se puede construir en base a los objetivos planteados, pero sí indica una aproximación de la metodología a seguir adaptada a las opiniones del grupo de expertos.

ESCENARIO E15

$$IC_{E15} = \frac{6}{100} * IAD_{FORMA} + \frac{7}{100} * IAD_{GUIA} + \frac{7}{100} * IAD_{INFORMA} + \frac{12}{100} * IAD_{MAT} + \frac{16}{100} * IAD_{DOC} + \frac{1}{100} * IAD_{TUT} + \frac{10}{100} * IAD_{EVA} + \frac{10}{100} * IAD_{OTRAS} + \frac{18}{100} * K_1 + \frac{12}{100} * K_2 + \frac{1}{100} * K_4$$

E15	Pesos nominales	S_i normalizado	Efectos principales S_i	Pesos objetivo	Diferencia
IAD_{FORMA}	6,0	0,124	12,4	12,0	0,4
IAD_{GUIA}	7,0	0,082	8,2	8,0	0,2
$IAD_{INFORMA}$	7,0	0,048	4,8	5,0	-0,2
IAD_{MAT}	12,0	0,102	10,2	10,0	0,2
IAD_{DOC}	16,0	0,153	15,3	15,0	0,3
IAD_{TUT}	1,0	0,084	8,4	7,0	1,4
IAD_{EVA}	10,0	0,074	7,4	8,0	-0,6
IAD_{OTRAS}	10,0	0,094	9,4	10,0	-0,6
K_1	18,0	0,179	17,9	18,0	-0,1
K_2	12,0	0,046	4,6	5,0	-0,4
K_3	0,0	0,000	0,0	1,0	-1,0
K_4	1,0	0,014	1,4	1,0	0,4
$ \Sigma $	100,0	1,000	100,0	100,0	

Tabla 6.26. Análisis de sensibilidad aplicado al escenario E15

Por ejemplo, para obtener unos valores ajustados entre el efecto principal del indicador IAD_{FORMA} sobre el IC y el peso objetivo para este indicador, se le asigna un peso nominal que coincide justamente con la mitad del peso objetivo. Esto se debe a la correlación que existe entre este indicador y K_1 ($r = 0,655$). La asignación de un peso elevado a K_1 también se ve reflejado en el peso efectivo del indicador IAD_{FORMA} . De esta forma, no se realiza un doble conteo del efecto principal del indicador IAD_{FORMA} .

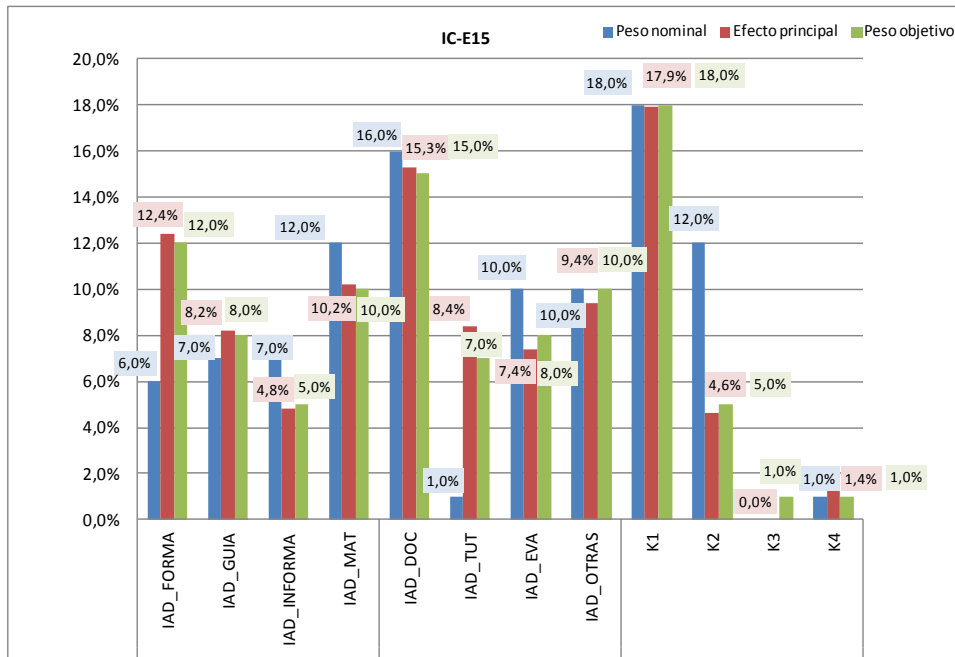


Figura 6.40. Pesos nominales, efectos principales y pesos objetivo de los indicadores simples del IC_{E15}

Según los resultados del análisis de sensibilidad, el escenario E15 es uno de los posibles escenarios que más se ajusta a los objetivos planteados en este estudio. Pero, ¿cómo varía el IC del escenario E15 al hacer variar en pequeñas cantidades los pesos nominales? El análisis de incertidumbre da respuesta a esta pregunta planteada.

El análisis de incertidumbre es un estudio interesante puesto que en muchas ocasiones se desea variar en pequeñas cantidades el peso nominal de los indicadores simples dependiendo del objetivo del estudio en cada año de evaluación. El indicador compuesto debe ser robusto a estas pequeñas variaciones y debe ordenar a los profesores de una manera similar tras los pequeños cambios introducidos en los pesos nominales.

Por tanto, para analizar la robustez y calidad de un indicador compuesto se deben aplicar los dos tipos de análisis. A continuación se muestran los resultados del análisis de incertidumbre del escenario E15 que es el que más se ajusta a los objetivos planteados por la CED para este trabajo y del escenario de la UPV que es la propuesta de índice de actividad docente desarrollada por la universidad. Los resultados del análisis de incertidumbre para el resto de escenarios se pueden encontrar en el anexo C.

6.2.9. Análisis de Incertidumbre

El objetivo del análisis de incertidumbre es cuantificar la variación del indicador compuesto al hacer variar las fuentes de incertidumbre del modelo. En el caso de estudio, las fuentes de incertidumbre del modelo pueden ser los pesos asignados, el tipo de normalización, el tipo de agregación y/o los datos originales.

Para el estudio se han considerado como fuentes de incertidumbre los pesos asignados, puesto que el tipo de normalización y el tipo de agregación se han fijado previamente en cada uno de los escenarios de trabajo definidos.

Se ha asignado un nivel de variación del 50% a aquellos pesos menores que 10 y del 25% a aquellos pesos mayores o iguales que 10, siguiendo el mismo criterio utilizado en el Índice de Competitividad Regional desarrollado recientemente por Annoni y Kozovska (2010). En las tablas 6.27 y 6.28 se presentan los intervalos de variación asignados a los pesos de cada escenario de trabajo.

	E1		E11		EUPV	
	Pesos nominales	Función de distribución de la variable	Pesos nominales	Función de distribución de la variable	Pesos nominales	Función de distribución de la variable
$W_{IAD_{FORMA}}$	0,08	U[0,041, 0,125]	0,343	U[0,257, 0,428]	0,3	U[0,225, 0,375]
$W_{IAD_{GUIA}}$	0,08	U[0,041, 0,125]	0,228	U[0,171, 0,285]	0,3	U[0,225, 0,375]
$W_{IAD_{INFORMA}}$	0,08	U[0,041, 0,125]	0,143	U[0,107, 0,178]	0,3	U[0,225, 0,375]
$W_{IAD_{MAT}}$	0,08	U[0,041, 0,125]	0,286	U[0,214, 0,357]	0,3	U[0,225, 0,375]
$W_{IAD_{DOC}}$	0,08	U[0,041, 0,125]	0,375	U[0,281, 0,468]	0,7	U[0,525, 0,875]
$W_{IAD_{TUT}}$	0,08	U[0,041, 0,125]	0,175	U[0,131, 0,218]	0,7	U[0,525, 0,875]
$W_{IAD_{EVA}}$	0,08	U[0,041, 0,125]	0,2	U[0,15, 0,25]	0,7	U[0,525, 0,875]
$W_{IAD_{OTRAS}}$	0,08	U[0,041, 0,125]	0,25	U[0,187, 0,312]	0,7	U[0,525, 0,875]
W_{K1}	0,11	U[0,083, 0,138]	0,76	U[0,57, 0,95]	0,6	U[0,45, 0,75]
W_{K2}	0,11	U[0,083, 0,138]	0,2	U[0,15, 0,25]	0,3	U[0,225, 0,375]
W_{K3}	0	–	0	–	0,05	U[0,025, 0,075]
W_{K4}	0,11	U[0,083, 0,138]	0,04	U[0,02, 0,06]	0,05	U[0,025, 0,075]

Tabla 6.27. Función de densidad de probabilidad asignada a cada fuente de incertidumbre en los escenarios E1, E11, EUPV

	E15		E7	
	Pesos nominales	Función de distribución de la variable	Pesos nominales	Función de distribución de la variable
$w_{IAD_{FORMA}}$	0,06	U[0,03, 0,09]	0,12	U[0,09, 0,15]
$w_{IAD_{GUIA}}$	0,07	U[0,035, 0,105]	0,08	U[0,04, 0,12]
$w_{IAD_{INFORMA}}$	0,07	U[0,035, 0,105]	0,05	U[0,025, 0,075]
$w_{IAD_{MAT}}$	0,12	U[0,09, 0,15]	0,1	U[0,075, 0,125]
$w_{IAD_{DOC}}$	0,16	U[0,12, 0,2]	0,15	U[0,1125, 0,1875]
$w_{IAD_{TUT}}$	0,01	U[0,005, 0,015]	0,07	U[0,035, 0,105]
$w_{IAD_{EVA}}$	0,1	U[0,075, 0,125]	0,08	U[0,04, 0,12]
$w_{IAD_{OTRAS}}$	0,1	U[0,075, 0,125]	0,1	U[0,075, 0,125]
w_{K1}	0,18	U[0,135, 0,225]	0,19	U[0,1425, 0,2375]
w_{K2}	0,12	U[0,09, 0,15]	0,05	U[0,025, 0,075]
w_{K3}	0	–	0	–
w_{K4}	0,01	U[0,005, 0,015]	0,01	U[0,005, 0,015]

Tabla 6.28. Función de densidad de probabilidad asignada a cada fuente de incertidumbre en los escenarios E15, E7

La asignación del nivel de incertidumbre se debe realizar de tal forma que los pesos tomen valores en un rango lo suficientemente amplio de la distribución uniforme y se obtenga una variación en las puntuaciones de los rankings lo más pequeña posible. Dicho de otro modo, el objetivo es elegir un nivel de incertidumbre amplio sobre las fuentes de incertidumbre de tal forma que las puntuaciones finales del IC no se vean fuertemente afectadas por dichas variaciones.

El análisis de incertidumbre se ha realizado para los cinco escenarios sobre los que se ha descrito más detalladamente el análisis de sensibilidad (EUPV, E1, E11, E7, E15). No obstante, se van a mostrar e interpretar con mayor profundidad los resultados de los escenarios E15, por ser el escenario que presenta una distancia mínima entre el conjunto de pesos efectivos y el conjunto de pesos objetivo definido por la CED y el escenario EUPV, por ser el escenario planteado por la universidad. Los resultados del resto de escenarios se muestran en el anexo C.

Las etapas del procedimiento Monte-Carlo que se han aplicado para el desarrollo del análisis de incertidumbre en cada uno de los escenarios se describen a continuación:

1. Se han generado 2^{14} muestras de cada una de las funciones de densidad de probabilidad asignadas mediante la técnica de muestreo cuasialeatoria con secuencias de baja discrepancia de Sobol', LP_{τ} (véase anexo D).

El uso de la técnica de muestreo cuasialeatoria de baja discrepancia es recomendable, ya que tiene la propiedad de cubrir un espacio de dimensión n de forma más uniforme que una secuencia de puntos aleatorios (Saltelli, Chan y Scott, 2000). Además, cuanto mayor es el número de muestras simuladas, menor es la discrepancia entre los puntos. En este caso se han realizado 2^{14} muestras para cada conjunto de pesos. El cálculo de 2^{15} simulaciones ha

resultado mucho más costoso computacionalmente, por lo que se ha considerado suficiente el número de muestras obtenido.

- Supóngase que se consideran n pesos en el escenario. La matriz final de pesos obtenida con las 2^{14} muestras se define como sigue:

$$W := \begin{bmatrix} \widehat{w}_{1,1} & \dots & \widehat{w}_{1,n} \\ \dots & \dots & \dots \\ \widehat{w}_{2^{14},1} & \dots & \widehat{w}_{2^{14},n} \end{bmatrix}$$

donde $\widehat{w}_{i,j}$ representa el peso obtenido en la simulación i para el peso nominal j

Finalmente se ha trabajado con 2^{14} conjuntos de pesos diferentes, que corresponden a las 2^{14} filas de la matriz W . Se va a nombrar a cada conjunto de pesos por $W_i = \{\widehat{w}_{i,1}, \dots, \widehat{w}_{i,n}\}$ para $i = 1, \dots, 2^{14}$

- Se han normalizado los conjuntos de pesos de acuerdo a las restricciones que presenta cada uno de los escenarios y que se exponen a continuación:

Escenario	Restricciones en los pesos
E1 E15 E7	$w_{IADFORMA} + w_{IADGUIA} + w_{IADINFORMA} + w_{IADMAT} + w_{IADDOC} + w_{IADTUT} + w_{IADEVA} + w_{IADOTRAS} + w_{K1} + w_{K2} + w_{K4} = 1$
E11	$w_{IADFORMA} + w_{IADGUIA} + w_{IADINFORMA} + w_{IADMAT} = 1$ $w_{IADDOC} + w_{IADTUT} + w_{IADEVA} + w_{IADOTRAS} = 1$ $w_{K1} + w_{K2} + w_{K4} = 1$
EUPV	$w_{IADFORMA} + w_{IADGUIA} + w_{IADINFORMA} + w_{IADMAT} = 1,2$ $w_{IADDOC} + w_{IADTUT} + w_{IADEVA} + w_{IADOTRAS} = 2,8$ $w_{K1} + w_{K2} + w_{K4} = 1$

Tabla 6.29. Restricciones en los pesos de cada escenario de trabajo

- Se ha calculado la puntuación del IC para cada observación y para cada conjunto de pesos, obteniendo la siguiente matriz:

$$M := \begin{bmatrix} IC_{1,W_1} & \dots & IC_{1,W_{2^{14}}} \\ \dots & \dots & \dots \\ IC_{2689,W_1} & \dots & IC_{2689,W_{2^{14}}} \end{bmatrix} \in \mathbb{R}^{2689 \times 2^{14}}$$

- Para cada una de las columnas de la matriz M se ha calculado la posición que ocupa cada observación, de tal manera que por filas se obtendrá la posición que ocupa cada observación según los distintos IC calculados con los distintos conjuntos de pesos. Sea P la matriz que define las posiciones de cada observación en los distintos conjuntos de pesos:

$$P := \begin{bmatrix} \text{Ranking}_{1,W_1} & \dots & \text{Ranking}_{1,W_{2^{14}}} \\ \dots & \dots & \dots \\ \text{Ranking}_{2689,W_1} & \dots & \text{Ranking}_{2689,W_{2^{14}}} \end{bmatrix} \in \mathbb{R}^{2689 \times 2^{14}}$$

El objetivo del análisis de incertidumbre es obtener un conjunto de posiciones para cada profesor con los 2^{14} conjuntos de pesos lo más parecido posible entre ellas y lo más diferente posible a las posiciones del resto de profesores. De esta forma, se mantiene el orden de las observaciones con respecto a las posiciones obtenidas en el escenario referencial.

Técnicas de visualización de los resultados del análisis de incertidumbre

DIAGRAMA DE CAJAS

Para visualizar los resultados de este estudio se ha representado mediante un gráfico de cajas la variación de las posiciones de cada uno de los profesores y la posición en cada escenario referencial. Puesto que se trabaja con un número de observaciones elevado, únicamente se han representado los 50 “primeros” y “últimos” profesores de los escenarios E15 y EUPV.

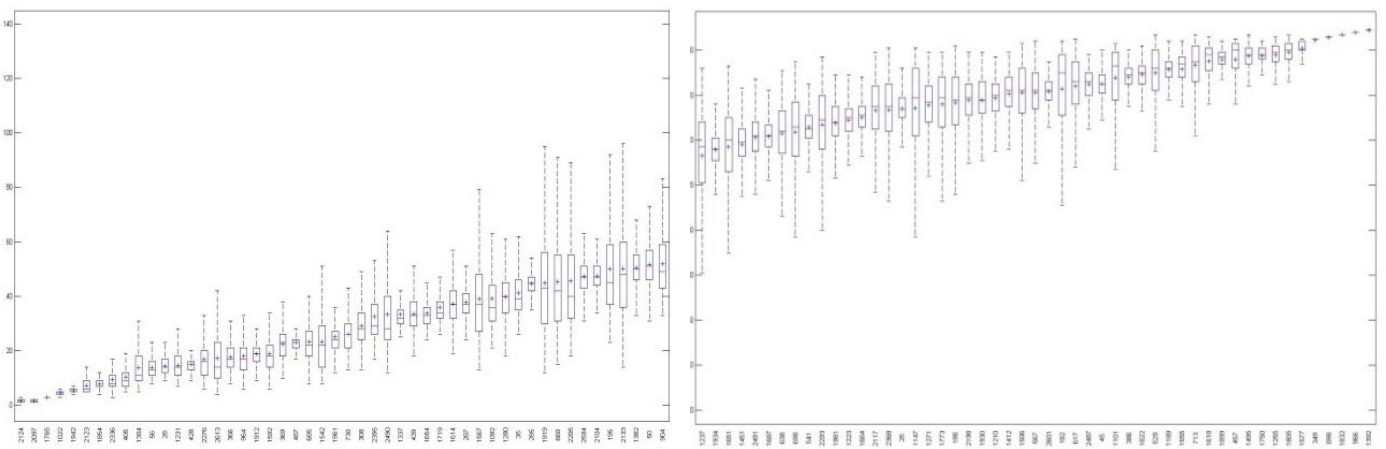


Figura 6.41. Resultados del AI donde se muestra el E15 referencial (cruces azules), la media (marcas rojas) y el correspondiente P_5 y P_{95} de la distribución del AI para los 50 “primeros” y “últimos” profesores

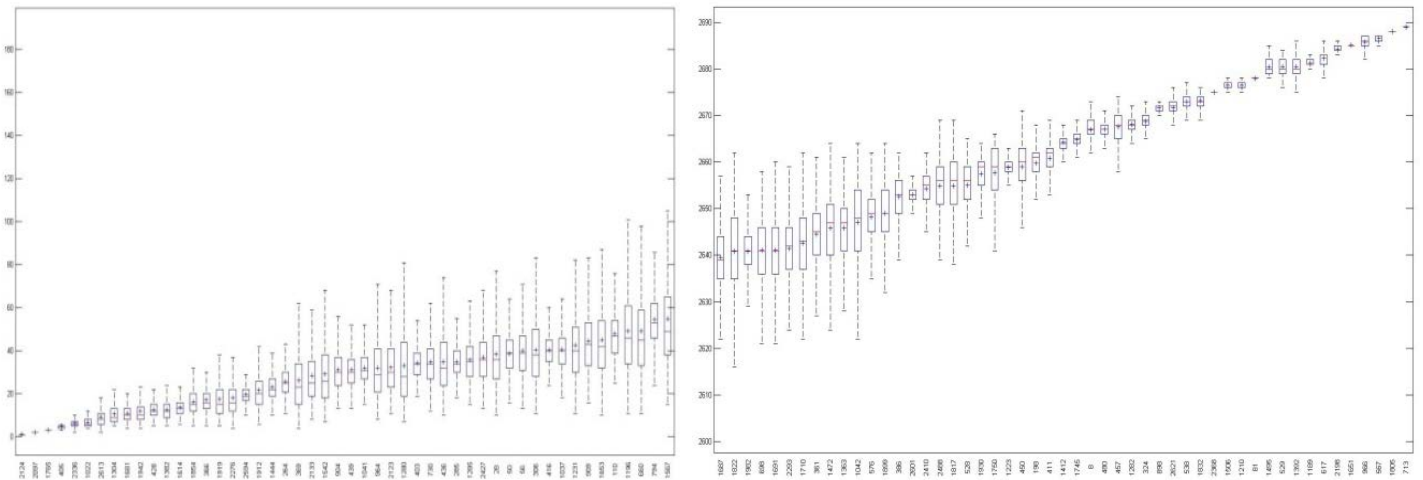


Figura 6.42. Resultados del AI donde se muestra el EUPV referencial (cruces azules), la media (marcas rojas) y el correspondiente P_5 y P_{95} de la distribución del AI para los 50 “primeros” y “últimos” profesores

De modo visual se observa que las posiciones de los profesores con “mejor” y “peor” puntuación de actividad docente son muy robustas frente a la incertidumbre asignada al modelo en ambos escenarios.

Además, para la mayoría de los profesores representados en ambos escenarios se cumple que el valor del escenario referencial es muy similar a la media de la distribución de cada profesor obtenida al variar los pesos nominales en el intervalo de incertidumbre asignado. Este resultado informa que los escenarios referenciales, a pesar de estar contruidos con un único conjunto de pesos, representan una imagen poco sesgada de la actividad docente del profesorado de la UPV.

Esta técnica de visualización de los resultados del análisis de incertidumbre es muy informativa si se tienen menos observaciones. Sin embargo, al tener tantas observaciones en el estudio la representación de la incertidumbre mediante el diagrama de cajas no permite obtener conclusiones más específicas del análisis de incertidumbre aplicado a los escenarios de trabajo. A continuación se presentan diferentes metodologías propuestas para la representación de los resultados del análisis de incertidumbre aplicado en este estudio.

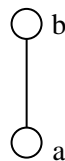
DIAGRAMA DE HASSE

Una relación \preceq en un conjunto A se define de *orden parcial* si \preceq es reflexiva, asimétrica y transitiva. El conjunto A junto con el orden \preceq se denomina conjunto *parcialmente ordenado* y se denota por (A, \preceq) . Los órdenes parciales \preceq más comunes son las relaciones \geq, \leq en Z y N . La teoría de la ordenación parcial se puede encontrar en (Neggers y Kim, 1998).

Sea $M_{n,k}$ una matriz formada por k indicadores observados en n observaciones y sea (P, \leq) la relación de orden parcial definida donde P es el conjunto de todas las observaciones n y \leq es la relación de orden utilizada. Sean a y b dos elementos del conjunto P ; $a \in P$ y $b \in P$. Se dice que $a \leq b$ si las puntuaciones de b son mayores que las de a para todos los indicadores k . Si todas las puntuaciones de los indicadores de a y b son iguales entonces se dice que a y b son equivalentes, $a \cong b$. En cualquier otro caso se dice que a y b son incomparables $a \parallel b$.

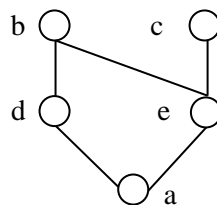
El diagrama de Hasse, desarrollado por Helmut Hasse en 1967, es una herramienta que permite representar las relaciones de orden definidas en el conjunto P . Se trata de una representación gráfica simplificada de un conjunto parcialmente ordenado finito. La teoría de la ordenación parcial es una herramienta muy utilizada para la generación de rankings usando múltiples indicadores (Ghanima, 2010).

La idea principal del diagrama de Hasse es realizar una ordenación de las n observaciones utilizando los k indicadores disponibles en lugar de aplicar una técnica de agregación lineal. Si una observación b es \geq que otra observación a ($a \leq b$) para todos los indicadores simples k , entonces las dos observaciones se dibujan en nodos diferentes, una encima de otra y separadas por una línea denominada arista.



En el caso en que $a \not\leq b \forall k$, entonces las dos observaciones son incomparables. El diagrama de Hasse elimina la necesidad de representar lazos en las aristas puesto que la relación parcialmente ordenada es reflexiva y, además, la transitividad permite prescindir de mostrar líneas entre elementos que tengan un elemento intermedio relacionado.

Sea, por ejemplo, el siguiente diagrama de Hasse:



En este caso, $b \geq d \forall k$, $b \geq e \forall k$ y $b \geq a \forall k$ (por transitividad); $d \geq a \forall k$ y $e \geq a \forall k$; $c \geq e \forall k$ y $c \geq a \forall k$ (por transitividad). Sin embargo $b \parallel c$ y $d \parallel e$.

La ventaja que presenta la aplicación del diagrama de Hasse es que permite agrupar los objetos en distintas categorías o niveles. En el caso del ejemplo anterior, las observaciones *b* y *c* están en la categoría superior, mientras que la observación *a* está en la categoría inferior. Además, la aplicación de esta técnica no exige ningún tipo de relación predeterminada (lineal, etc.) entre los indicadores simples, puesto que se trata de un método no paramétrico. Por otra parte, el diagrama de Hasse es uno de los métodos de clasificación menos subjetivo, puesto que no se realizan juicios de valor en su proceso de construcción. Sin embargo, la aplicación de esta técnica no proporciona un orden total de las observaciones y puede que clasifique dos observaciones en la misma categoría a pesar de que una de ellas presente puntuaciones mucho más elevadas en la mayoría de los indicadores simples. Además, los resultados de la clasificación no permiten cuantificar el grado de diferencias existente entre las observaciones (Ghanima, 2010). Annoni, Bruggemann y Saltelli (2012) aplican el diagrama de Hasse para representar mediante un orden parcial a un conjunto de países según un conjunto de indicadores simples.

En este trabajo se ha aplicado esta técnica de ordenación sobre los resultados del análisis de incertidumbre de los escenarios mediante el software Dart 2.05 (*Decision Analysis by Ranking Techniques*). Sin embargo, los resultados no proporcionan información interesante para analizar la incertidumbre del IC, puesto que los distintos niveles están formados por muchos profesores y no se pueden comparar entre ellos.

En la figura 6.43 se muestran los resultados del análisis de incertidumbre del escenario E15 para los 50 profesores con mejores posiciones con respecto al ranking del escenario referencial. Se observa, por ejemplo, que el profesor “1” ocupa una posición menor que cualquier otro profesor en los 2^{14} rankings derivados de los 2^{14} conjuntos de pesos simulados, lo cual significa que es el profesor con mayor actividad docente. Igual ocurre con los profesores “2”, “3” y “4”. Sin embargo, conforme se aumenta el número de profesores el diagrama no permite realizar comparaciones entre ellos, puesto que los va colocando en la misma categoría. Por tanto, los resultados obtenidos sólo permiten diferenciar a los “mejores” y a los “peores” profesores del resto, al igual que la representación del análisis de incertidumbre mediante el diagrama de cajas. Es por eso que el diagrama de Hasse no ofrece resultados relevantes para interpretar los resultados del análisis de incertidumbre en este estudio.

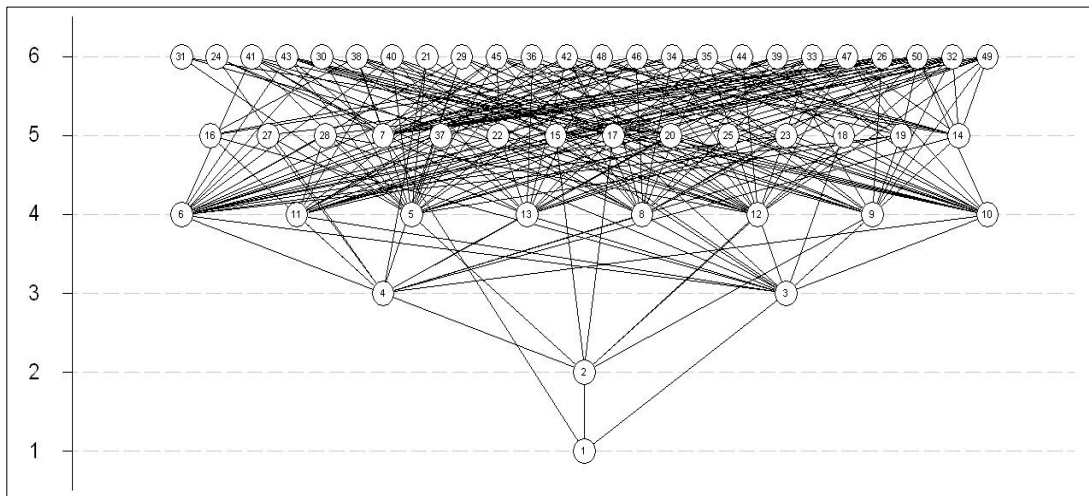


Figura 6.43. Diagrama de Hasse aplicado a los resultados del A.I. del escenario E15

EL MÉTODO DE COPELAND

El método de Copeland (1951) es una herramienta no paramétrica de agregación que se utiliza para establecer un orden en las observaciones, según las puntuaciones de éstas en diferentes indicadores. Se trata de una extensión del método de Condorcet y es un método muy utilizado en el campo político para ordenar candidatos según un conjunto de votaciones.

Recientemente, Ghanima (2010) ha aplicado el modelo de Copeland para contextos sobre elecciones mediante votaciones, en el que los candidatos de las votaciones son las unidades de análisis (observaciones) y las votaciones son los indicadores simples. El método de Copeland selecciona la observación con mayor índice de Copeland, siendo este índice el número de veces que una observación es mejor que las otras menos el número de veces que es peor (comparando las observaciones por pares).

El método construye una matriz de comparaciones C , de tal forma que se compara un indicador tras otro (X_i , $i = 1, \dots, Q$) para cada par de observaciones (x_j, x_k , $j, k = 1, \dots, N$) aplicando la siguiente regla:

- Si $X_i(x_j) > X_i(x_k)$, $C(j, k) = C(j, k) + 1$
- Si $X_i(x_j) < X_i(x_k)$, $C(j, k) = C(j, k) - 1$, y
- Si $X_i(x_j) = X_i(x_k)$, $C(j, k) = C(j, k) + 0$

Sumando todas las puntuaciones $C(j, k)$ para todo $k = 1, \dots, M$ se obtiene $CS(x_j)$. Posteriormente las observaciones se ordenan según el valor CS .

A continuación se muestra un ejemplo muy sencillo en el que se aplica el método de Copeland. Supóngase que se tienen siete unidades de análisis a las que se les ha medido tres indicadores:

	X ₁	X ₂	X ₃
A	9	10	2
B	1	11	1
C	4	11	5
D	3	10	1
E	8	12	4
F	6	10	6
G	3	9	8

Por ejemplo, si se comparan las observaciones A y B, se parte de $C(A, B) = 0$ y aplicando la regla se obtiene:

- $X_1(A) > X_1(B)$, $C(A, B) = C(A, B) + 1 = 0 + 1 = 1$
- $X_2(A) < X_2(B)$, $C(A, B) = C(A, B) - 1 = 1 - 1 = 0$
- $X_3(A) > X_3(B)$, $C(A, B) = C(A, B) + 1 = 0 + 1 = 1$

Al realizar estas comparaciones para cada par de observaciones se obtiene la siguiente matriz de comparaciones C y el índice de Copeland para cada observación:

	A	B	C	D	E	F	G	CS
A	0	1	-1	2	-1	0	1	$\Sigma=2$
B	-1	0	-2	0	-3	-1	-1	-8
C	1	2	0	3	-1	-1	1	5
D	-2	0	-3	0	-3	-2	0	-10
E	1	3	1	3	0	1	1	10
F	0	1	1	2	-1	0	1	4
G	-1	1	-1	0	-1	-1	0	-3

Por tanto, el ranking queda de la siguiente forma:

Observación	Ranking
E	1
C	2
F	3
A	4
G	5
B	6
D	7

Para una revisión más precisa de los métodos multi-criterio no compensatorios véanse las siguientes referencias bibliográficas (Vincke, 1992; Munda, 1995; Roy, 1996; Munda, 2005).

En este trabajo se ha considerado adecuado aplicar la técnica de agregación multi-criterio no compensatoria de Copeland a los resultados del análisis de incertidumbre, puesto que la cantidad tan elevada de observaciones no permite obtener buenas interpretaciones con las otras técnicas de visualización del análisis de incertidumbre. Además, se trata de una técnica que tiene en cuenta las posiciones de cada observación en los diferentes rankings y las compara con las posiciones del resto de observaciones.

Sea P la matriz resultante de la aplicación del análisis de incertidumbre definida como sigue:

$$P: = \begin{pmatrix} \text{Ranking}_{1,W_1} & \dots & \text{Ranking}_{1,W_{2^{14}}} \\ \dots & \dots & \dots \\ \text{Ranking}_{2689,W_1} & \dots & \text{Ranking}_{2689,W_{2^{14}}} \end{pmatrix} \in \mathbb{R}^{2689 \times 2^{14}}$$

donde $\text{Ranking}_{i,W_j}$ es la posición del profesor i calculada con el conjunto de pesos w_j . La posición se obtiene comparándose con el IC del resto de profesores calculado con el mismo conjunto de pesos w_j . Para aplicar Copeland se ha considerado que los conjuntos de pesos son los indicadores y los profesores son las observaciones, de tal forma que se evalúa la posición de cada profesor en todos los conjuntos de pesos calculados:

	W_1	...	$W_{2^{14}}$
P1	$\text{Ranking}_{1,W_1}$...	$\text{Ranking}_{1,W_{2^{14}}}$
...
P2689	$\text{Ranking}_{2689,W_1}$...	$\text{Ranking}_{2689,W_{2^{14}}}$

En la figura 6.44 se comparan las posiciones obtenidas mediante el escenario referencial y las obtenidas una vez aplicada la técnica Copeland sobre los resultados del análisis de incertidumbre en los escenarios E15 y EUPV.

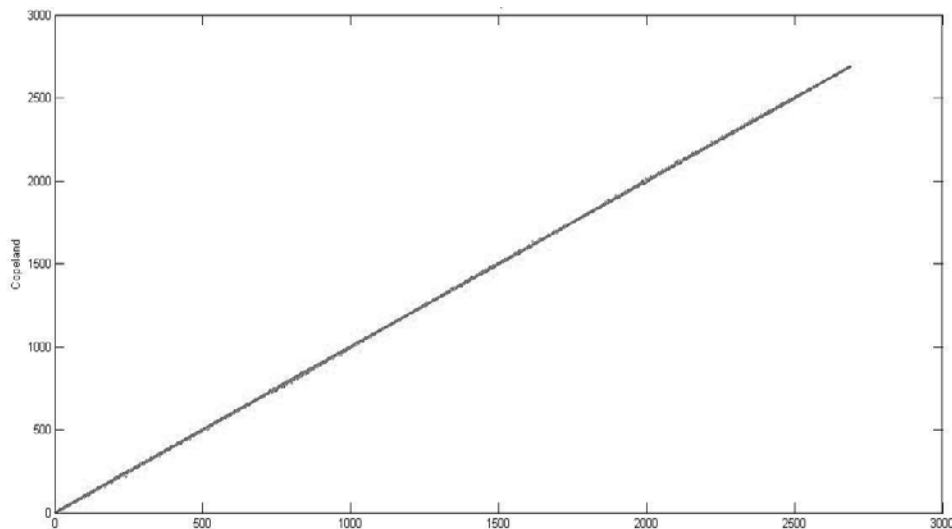


Figura 6.44. Comparación de las posiciones del escenario referencial E15 y las posiciones aplicando Copeland

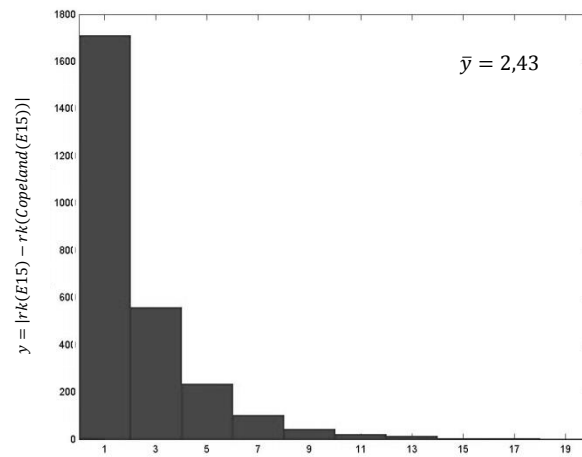


Figura 6.45. Histograma de las diferencias en posiciones del escenario referencial E15 y las posiciones aplicando Copeland

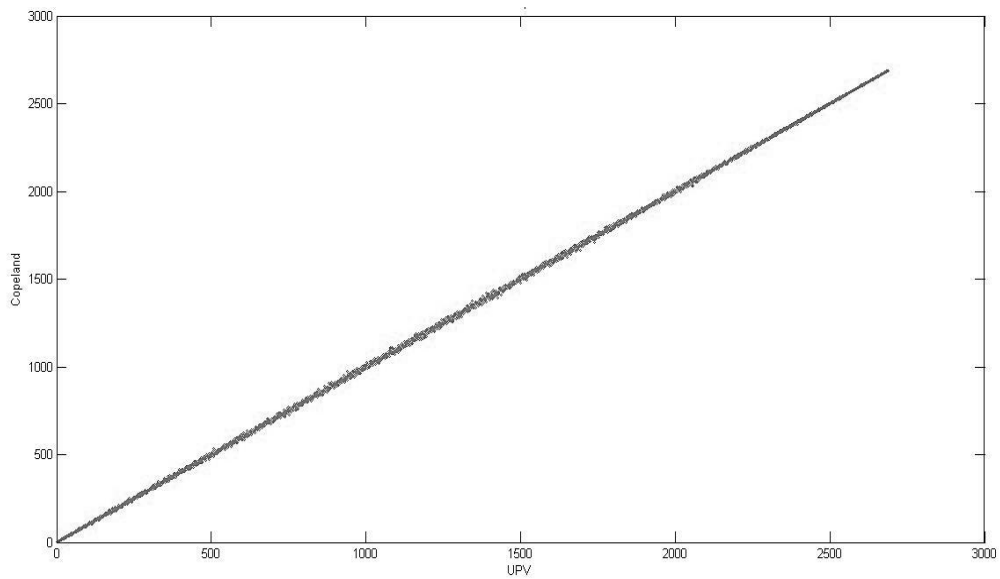


Figura 6.46. Comparación de las posiciones del escenario referencial EUPV y las posiciones aplicando Copeland

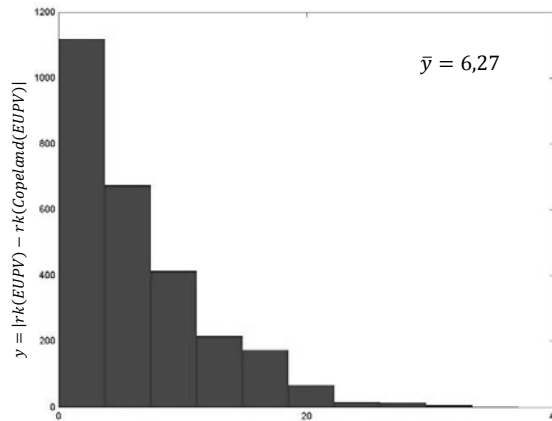


Figura 6.47. Histograma de las diferencias en posiciones del escenario referencial EUPV y las posiciones aplicando Copeland

Según las figuras 6.44 y 6.46, los escenarios E15 y EUPV son robustos frente a la incertidumbre asignada al conjunto de pesos nominales, ya que las posiciones de los profesores en general se mantienen al aplicar un análisis de incertidumbre.

Los histogramas informan de la distribución de la variable $y = |rk(E15) - rk(Copeland(E15))|$ donde $rk(E15)$ es la posición de cada profesor en el escenario E15 y $rk(Copeland(E15))$ es la posición de cada profesor al aplicar la técnica Copeland sobre los resultados del análisis de incertidumbre del escenario E15. La variable y se ha definido de forma análoga para el escenario EUPV.

Se ha considerado la media de la variable definida y para identificar el escenario más robusto al aplicar un análisis de incertidumbre sobre los pesos. Una media próxima a cero, $\bar{y} \cong 0$, indica que las posiciones de los profesores al aplicar el análisis de incertidumbre no varían prácticamente en nada con respecto a las posiciones en el escenario referencial y , por tanto, que el escenario es robusto frente a una pequeña variación en los pesos asignados a los indicadores simples.

En este caso el escenario propuesto E15 es el más robusto para la construcción de un IC que mida la actividad docente del profesorado de la UPV, según las orientaciones del grupo de expertos de la CED. Además, según los resultados del análisis de sensibilidad es el escenario que mejor aproxima el conjunto de efectos principales al conjunto de pesos objetivo definido por el grupo de expertos. No obstante, el indicador compuesto definido por la UPV para la medición de la actividad docente del profesorado de la UPV también es muy robusto según las figuras 6.46 y 6.47.

La metodología propuesta para la construcción del escenario E15 podría ser de gran utilidad para la universidad, puesto que si se replantease orientar la actividad docente hacia otra

dirección de forma que se aumentara o disminuyera el peso de un(os) indicador(es) o dimensión(es) en concreto, los resultados del indicador compuesto E15 no se verían fuertemente afectados por un pequeño cambio.

6.3. Resumen y conclusiones

Este capítulo se ha centrado en la búsqueda de diferentes estrategias metodológicas para la construcción de un IC que mida la actividad docente del profesorado con el fin de disponer de herramientas eficientes para asegurar la garantía de calidad de la enseñanza. Para ello, se han aplicado diversas técnicas de análisis de datos y se han definido escenarios de trabajo sobre los que finalmente se ha realizado un análisis de sensibilidad y de incertidumbre para evaluar su calidad y robustez, y conseguir que los procesos de evaluación sean más transparentes en la construcción de ICs.

En primer lugar, el análisis de tratamiento de datos ha permitido identificar la heterogeneidad de la base de datos. En este sentido, se sugiere la aplicación de una técnica de normalización previamente a la construcción de un IC con el objetivo de disponer de indicadores comparables. Se trata de una de las posibles mejoras a introducir en el desarrollo del IC propuesto por la UPV en el que se ha realizado una simple transformación sobre los indicadores para obtener una misma unidad de medida, llamada puntos IAD. Además, el rango de variación y la variabilidad de cada uno de los indicadores presentan comportamientos heterogéneos que afectan a la importancia de cada indicador en la formación del IC, tal y como se demuestra posteriormente en el análisis de sensibilidad. No obstante, el análisis realizado muestra que mediante la construcción del Índice de Actividad Docente vigente en la UPV se han obtenido unos resultados coherentes con los objetivos planteados por la universidad.

La aplicación del análisis de componentes principales ha permitido analizar la estructura subyacente de la base de datos del estudio para la construcción del IC. Los resultados han identificado cinco dimensiones que definen las actividades docentes medidas con los indicadores simples. Estas dimensiones se han denominado como *Volumen de actividad docente*, *Resultados de la evaluación del profesorado*, *Volumen de actividades adicionales al desempeño habitual de la actividad docente*, *Gestión y Rendimiento académico del alumnado*. Este análisis propone un marco conceptual teórico alternativo al programa DOCENTIA que está basado en el grado de correlación existente entre los indicadores. Sin embargo, en el enfoque de este programa se definen las dimensiones de acuerdo a los procesos académicos que se siguen en la universidad: *Planificación de la docencia*, *Desarrollo de la enseñanza* y *Resultados*.

La aplicación del análisis de consistencia interna ha demostrado que la eliminación de cada uno de los indicadores simples no hace aumentar de forma considerable la fiabilidad del conjunto de datos y, por tanto, se concluye que todos los indicadores son imprescindibles para la medición de la actividad docente. El análisis de consistencia interna confirma la correcta selección del conjunto de indicadores simples para la medición de las distintas actividades que definen la actividad docente, lo que aporta nuevos indicios de validez del sistema de medida.

Por otra parte, se han identificado diferentes perfiles de profesorado según su dedicación a cada tipo de actividad docente. Además, en algunos casos se ha podido relacionar el perfil definido con una categoría profesional de profesorado. Para obtener la categorización sobre los datos disponibles el algoritmo de clasificación en dos fases ha mostrado ser eficiente.

El análisis de sensibilidad global basado en el cálculo de varianzas ha permitido medir la discrepancia entre el efecto principal y la importancia deseada (o peso objetivo) proporcionada por el grupo de expertos de la CED para cada indicador simple en los diferentes escenarios propuestos. La estimación de los efectos principales se ha calculado mediante el procedimiento no paramétrico *State Dependent Parameter* (SDP) por las ventajas que ofrece: i) su coste computacional es menor que los métodos de GSA clásicos, ii) no requiere la generación de simulaciones de los factores input, iii) se puede aplicar independientemente del grado de correlación entre los factores input y, finalmente, iv) se puede emplear como un método no invasivo, ya que su aplicación no altera los resultados del IC ni la estructura de correlación de los indicadores.

Los resultados del análisis de sensibilidad muestran que los efectos principales de los indicadores de cada escenario definido discrepan, en menor o mayor medida, del conjunto de pesos objetivo definido por el grupo de expertos. En concreto, en el escenario de la UPV se ha podido comprobar cómo la ausencia de normalización y la estructura de correlación afectan al efecto principal de los indicadores. Finalmente, se ha definido un nuevo escenario E15 con una discrepancia entre el efecto principal y la importancia deseada de cada indicador mínima. Se trata, por tanto, de un escenario que cumple con el objetivo deseado para este estudio por el grupo de expertos. Si cambiase el criterio de la CED sobre la importancia asignada a los indicadores simples el IC se podría reformular aplicando la metodología propuesta en este capítulo.

El método de análisis de sensibilidad aplicado a este estudio evidencia la necesidad de reflexionar sobre la interpretación que se le da a los pesos nominales como medidas de importancia en el proceso de construcción de un IC en el que se emplean técnicas de agregación compensatorias. Además, este método proporciona una herramienta de evaluación de la calidad de un IC basada en el cálculo de los efectos principales de cada uno de los indicadores simples sobre el IC. Se trata de un procedimiento de medición orientado hacia la mejora en la construcción de ICs y que ofrece a la universidad información fiable para una posible reestructuración del conjunto de pesos nominales en algunos indicadores simples.

Por último, el análisis de incertidumbre ha permitido cuantificar la variación de las posiciones de cada uno de los profesores al asignar un nivel de incertidumbre en los pesos de los indicadores simples. La herramienta de visualización más efectiva para la interpretación de los resultados ha resultado ser la técnica Copeland, puesto que proporciona una ordenación del profesorado teniendo en cuenta la incertidumbre asignada a los pesos. Los resultados de esta técnica muestran que el escenario E15 es robusto frente a pequeñas variaciones asignadas a los pesos, es decir, que si se realiza una pequeña reestructuración en el conjunto de pesos nominales los resultados del IC no se verían fuertemente afectados. Este resultado constituye una

aportación muy interesante en el caso en que el IC se modificase cada cierto tiempo según los objetivos estratégicos de las universidades.

En conclusión, este capítulo proporciona una metodología de construcción de indicadores compuestos adaptada a la gestión de la educación superior que es eficiente según los criterios planteados por la universidad y que permite modificar los ICs de modo que se ajusten a los objetivos de la institución.

Capítulo 7

La evaluación de la actividad académica del profesorado universitario

7.1. Introducción

A pesar del interés creciente por parte de las IES por disponer de herramientas de medición para evaluar la actividad académica del profesorado universitario, el tema sigue siendo objeto de constantes discrepancias. Uno de los motivos que incita la aparición de estos debates es la falta de una metodología consistente que justifique el proceso de evaluación y que sea consensuada y aceptada por la mayoría de expertos en el ámbito de la educación superior.

Son muchas las cuestiones sin resolver acerca de los criterios a seguir para evaluar las actividades del profesorado universitario. Algunas de estas cuestiones se plantean ¿qué actividades se deben incluir en la evaluación de un profesor universitario?, ¿qué se entiende por profesor eficaz en el desarrollo de su actividad académica?, ¿cómo se pueden ponderar y agregar los distintos aspectos que miden las actividades académicas del personal docente e investigador (PDI) universitario?, ¿cuál es la mejor herramienta que permite proporcionar a los profesores un *feedback* de su actividad realizada?, etc.

Este capítulo, y la tesis en general, no trata de resolver todas las cuestiones anteriores relativas a los criterios teóricos que se deben tener en cuenta en la medición de las distintas actividades que desarrolla un docente universitario. El problema abordado es el . Para ello, se va a adoptar el modelo teórico descrito por la universidad compuesto por tres grandes dimensiones: docencia, investigación y gestión académica del profesorado de la UPV (véase capítulo 5).

Como ya se ha mencionado en el capítulo anterior, la ponderación de los indicadores es una de las etapas más criticadas en la metodología de construcción de indicadores compuestos. La importancia asignada a cada una de las dimensiones de estudio y la metodología de agregación de éstas es un tema discutido abiertamente por muchos investigadores, entre los que destacan Arneson (1990), Fleurbaey (1992) y Rawls (1999) en la literatura sobre la elección social. Estos autores debaten la idea de proponer un conjunto de pesos fijo para todas las unidades de estudio independientemente de las preferencias de cada unidad, o la idea de asignar a cada unidad un

conjunto de pesos individualizado según sus preferencias. Es lo que ellos definen como el *dilema de indexación*.

En el caso particular del Índice de Actividad Académica, este razonamiento se traduce en la posibilidad de asignar un conjunto de pesos fijo e igual para todo el profesorado sin tener en cuenta los distintos perfiles de profesorado o, por contra, asignar un conjunto de pesos personalizado a cada profesor según su perfil académico. En el primer caso de ponderación, el modelo no tiene en cuenta las diferentes formas que tiene el profesorado de desarrollar su actividad académica, ya que se fija a priori un peso para cada dimensión que es el mismo para todo el profesorado. Mientras que en el segundo caso, el modelo sí tiene en cuenta las diferentes contribuciones del profesorado al desarrollo de la actividad académica.

Muchos otros autores del ámbito de la educación superior critican la idea de asignar un conjunto de pesos fijo e igual para todo el PDI, como por ejemplo Marsh y Roche (1994), Pritchard et al. (1998), Avital y Collopy (2001), Emery, Kramer y Tian (2003). De esta forma, los autores argumentan que las actividades de docencia e investigación son muy complejas como para simplificarlas asignando un peso fijo a cada una de ellas para todo el profesorado. Aunque las dimensiones de docencia, investigación y gestión sean importantes, puesto que se han considerado para la medición de la evaluación académica del profesorado, cada profesor percibe la importancia de cada una de ellas de forma diferente según sus concepciones de profesor excelente.

Como se observa, la etapa correspondiente a la ponderación de los distintos indicadores o dimensiones es una de las principales críticas que recibe la construcción del Índice de Actividad Académica. Los profesores, expertos⁹ y tomadores de decisiones tienen diferentes visiones y opiniones de lo que significa ser un profesor eficaz o eficiente en el desarrollo de actividades realizadas en las dimensiones de docencia, investigación y gestión con respecto al resto del profesorado.

Por tanto, la falta de conocimiento *a priori* sobre la correcta elección del conjunto de pesos en las dimensiones de estudio justifica el uso de una metodología flexible con pocas hipótesis de partida, aplicada en este último capítulo para la evaluación de la actividad académica del profesorado de la UPV.

Este capítulo pretende contribuir a los estudios existentes sobre la evaluación de la actividad académica del profesorado universitario mediante la metodología de medición aplicada. La técnica utilizada se conoce como el enfoque “Beneficio de la duda – *Benefit of the Doubt*” (BOD) y tiene sus raíces en la técnica no paramétrica Análisis Envolvente de Datos (DEA, por sus siglas en inglés), utilizada para la estimación del rendimiento de un conjunto de observaciones.

⁹ Un caso ilustrativo de los diferentes puntos de vista de un grupo de expertos se expone en el escenario propuesto E2 del apartado 7.2.5

Para obtener los resultados expuestos a lo largo del capítulo se han empleado diferentes programas estadísticos. El tratamiento de datos se ha realizado mediante el software SPSS 16.0, mientras que para la aplicación de la metodología BOD se ha utilizado el software libre R 2.11.1. Finalmente, la visualización de los resultados de cada profesor se ha realizado mediante la programación de un código en TikZ, que es un paquete de dibujo de LaTeX.

7.2. Índice de Actividad Académica del profesorado de la UPV

7.2.1. Desarrollo del marco conceptual teórico y definición de los indicadores del estudio

El objetivo del Índice de Actividad Académica (IAA) es medir en un único valor las diferentes actividades del profesorado de la UPV. Se define mediante la medición de las actividades de docencia, I+D+i y gestión. Inicialmente, la universidad planteó incorporar la formación del profesorado como otra actividad adicional a las anteriores. Sin embargo, la dificultad de medir esta actividad y la falta en la UPV de un indicador que la evalúe, no permite incorporarla como un indicador independiente al resto de actividades del profesorado en la UPV (2012).

Las actividades de docencia, I+D+i y gestión son, a su vez, conceptos multidimensionales. Por tanto, para evaluar la actividad académica del profesorado se parte de un conjunto de ICs definidos previamente por la UPV que en el proceso de construcción del IC toman el papel de los indicadores simples (véase capítulo 3).

La **actividad docente** se evalúa mediante el conjunto de tareas que hay que realizar para que los alumnos adquieran las competencias y habilidades definidas en los planes de estudio. Es por eso que la UPV ha considerado adecuado definir el Índice de Actividad Docente (**IAD**) para recoger las distintas actividades docentes (Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV, 2010). Este índice se describe en detalle en el capítulo 5 de la presente Tesis Doctoral.

La **actividad investigadora** se valora fundamentalmente por los resultados. Además, la UPV considera que la actividad que produce un resultado puede extenderse hasta varios años anteriores. Por ello, para valorar esta dimensión se tiene en cuenta la Valoración de la Actividad Investigadora Personalizada (**VAIP**) que en el año de estudio X se calcula como sigue:

$$VAIP_X = IAI_X + 0,75 * IAI_{X-1} + 0,5 * IAI_{X-2} + 0,25 * IAI_{X-3}$$

El Índice de Actividad Investigadora (IAI) desarrollado por la UPV se puede encontrar con más detalle en el reglamento publicado por la universidad y aprobado por Consejo de Gobierno en 2008 (Universitat Politècnica de València, 2008).

La **actividad de gestión (AG)** queda definida en la Normativa de Organización Docente (NOD) de la UPV. Las actividades principales son: la reducción de créditos lectivos a los cargos académicos y la dedicación del profesorado a las sesiones de las comisiones y demás órganos colegiados necesarios para la dirección y gestión de la universidad y sus unidades académicas. La última actividad se contabiliza en términos de créditos de gestión asignando un número de horas de dedicación al año a cada tarea. Cada crédito de gestión equivale a $1.560/33=47,27$ horas y a un crédito lectivo reducido en la NOD.

7.2.2. Tratamiento de datos

Para este estudio se dispone de una base de datos formada por 2.063 observaciones que corresponden al Personal Docente e Investigador de la UPV en plantilla en el año académico 2011/2012. Los datos han sido anonimizados y proceden del Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV. Se ha excluido de la construcción del Índice de Actividad Académica al profesorado con categoría laboral de Asociado, puesto que el tipo de contrato correspondiente no considera la eventual actividad de investigación y de gestión.

En primer lugar, se ha realizado un análisis descriptivo de los tres indicadores propuestos por la UPV para la medición de la actividad académica de su profesorado. Según los resultados obtenidos, el indicador que mide la actividad de docencia tiene un rango de variación de $[0;68,5]$, por tanto, el límite inferior del intervalo indica que existe al menos un profesor exento de docencia. Según la normativa docente vigente de la UPV debe quedar excluido de evaluación el profesorado que, por distintas razones, no tuvo docencia asignada en el curso objeto de valoración (Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV, 2009). En la tabla 7.1 se muestran los estadísticos básicos calculados tras la eliminación del profesorado con una puntuación nula en la dimensión de docencia.

	N	Media	Desv. Típ.	Rango de variación		Asimetría	Curtosis
				Mínimo	Máximo		
IAD	2020	47,11	10,64	9,41	68,50	0,03	-0,27
VAIP	2020	49,69	53,63	0,00	208,00	1,50	1,65
AG	2020	2,00	3,55	0,00	20,00	2,45	6,28

Tabla 7.1. Estadísticos descriptivos de las dimensiones del IAA

De esta forma, la base de datos del estudio se reduce a 2.020 profesores evaluados. Además, en la base de datos se observa que el valor máximo para el indicador de docencia y para el indicador de investigación es alcanzado por un grupo de profesores. Esto se debe a que la UPV ha acotado el Índice de Actividad Docente y la Valoración de la Actividad Investigadora Personalizada del profesorado para que no existan valores atípicos, considerando que un profesor realiza favorablemente las actividades de docencia y de investigación si alcanza dichos límites. Así pues, la universidad ha aplicado la siguiente acotación en las dimensiones:

- Para la dimensión IAD la puntuación se limita al valor de IAD superado por el 3% del profesorado de la UPV. Este valor corresponde a una puntuación de 68,5 puntos IAD

- En la dimensión VAIP la puntuación se limita al valor del indicador VAIP que es superado por el 3% del profesorado de la UPV. Este valor corresponde a una puntuación de 208 puntos IAI
- En la dimensión de gestión la puntuación se limitan a 20 créditos

Esta acotación *a priori* de las dimensiones de docencia e investigación evita posibles valores atípicos en estas dimensiones que pueden ser referentes para el cálculo del Índice de Actividad Académica del profesorado.

Adicionalmente al estudio de los descriptivos básicos de las dimensiones se ha analizado la relación existente entre éstas mediante el cálculo de correlaciones.

	IAD	VAIP	AG
IAD	1	-0,104**	-0,063**
VAIP	-0,104**	1	0,168**
AG	-,063**	0,168**	1

** La correlación es significativa al nivel 0,01 (bilateral)

Tabla 7.2. Coeficientes de correlación entre las dimensiones de estudio

El coeficiente de correlación de Pearson muestra que la relación entre los indicadores es significativa a un nivel de 0,01. Sin embargo, este índice de correlación está condicionado por un número de observaciones elevado. En este caso, el número de profesores es muy elevado (2.020) y, por tanto, no se considera que exista un alto grado de correlación entre las dimensiones del estudio, ya que todos los coeficientes de correlación son inferiores a 0,3 en valor absoluto (Hair et al., 2007). Así pues, se puede afirmar que los indicadores medidos muestran actividades diferentes realizadas por el profesorado y que el conjunto de indicadores es relevante y adecuado para la construcción del Índice de Actividad Académica.

7.2.3. Análisis del Índice de Actividad Académica propuesto por la UPV

En el capítulo 5 se ha descrito la metodología propuesta por la UPV para la evaluación académica del profesorado universitario. En esta sección se ha analizado de forma descriptiva el escenario propuesto por la UPV, denotado como IAA_UPV.

Análisis descriptivo del escenario IAA_UPV

En primer lugar, se ha realizado un estudio descriptivo del indicador compuesto propuesto por la UPV para la evaluación de la actividad académica. Tal y como se muestra en el capítulo 5, el modelo que sigue la UPV para el desarrollo del IAA de un profesor c es el siguiente:

$$IAA_UPV_c = p_{doc} * \frac{IAD_c^*}{IAD_{med,UPV}} + p_{inv} * \frac{VAIP_c^*}{VAIP_{med,UPV}} + p_{ges} * \frac{AG_c^*}{AG_{med,UPV}} \text{ con } c = 1, \dots, 2.020$$

$$IAA_UPV_c = 67,60 * \frac{IAD_c^*}{43,48} + 26,40 * \frac{VAIP_c^*}{38,11} + 6 * \frac{AG_c^*}{1,56}$$

donde:

$IAD_{med,UPV}$: valor medio del indicador de docencia para todo el conjunto de profesores de la UPV

$VAIP_{med,UPV}$: valor medio del indicador de investigación para todo el conjunto de profesores de la UPV

$AG_{med,UPV}$: valor medio del indicador de gestión para todo el conjunto de profesores de la UPV

La figura 7.1 muestra la distribución del escenario IAA_UPV y los principales estadísticos básicos del indicador compuesto. Se ha considerado el IAA_UPV normalizado, IAA_UPV_N, a una escala de [0,1] para poder comparar su comportamiento con los escenarios que se proponen en la sección 7.2.5. Según este gráfico, el modelo propuesto por la UPV para la evaluación académica del profesorado asigna una puntuación muy baja a la mayoría de los profesores evaluados, siendo la media global del indicador compuesto 0,36.

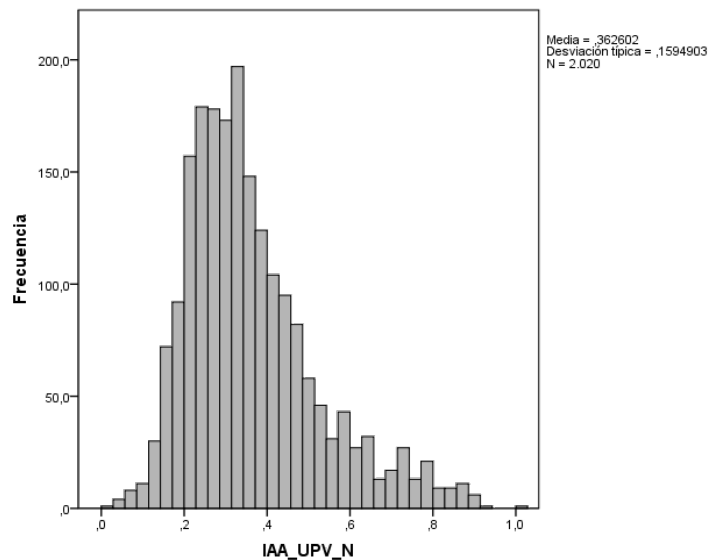


Figura 7.1. Histograma del escenario IAA_UPV_N

La figura 7.2 muestra la distribución del escenario IAA_UPV_N según las categorías laborales de profesorado. Se observa que el profesorado que posee, en promedio, una actividad académica mayor es Catedrático de Universidad. Por el contrario, los profesores Ayudantes y Ayudantes Doctores son los que desarrollan menor actividad académica en promedio, según los

resultados del escenario IAA_UPV_N. Además, la mayoría del profesorado tiene un IAA_UPV_N inferior a 0,5, siendo el valor máximo 1. Este último resultado se confirma con el histograma de la figura 7.1 y los estadísticos básicos del escenario que se muestran.

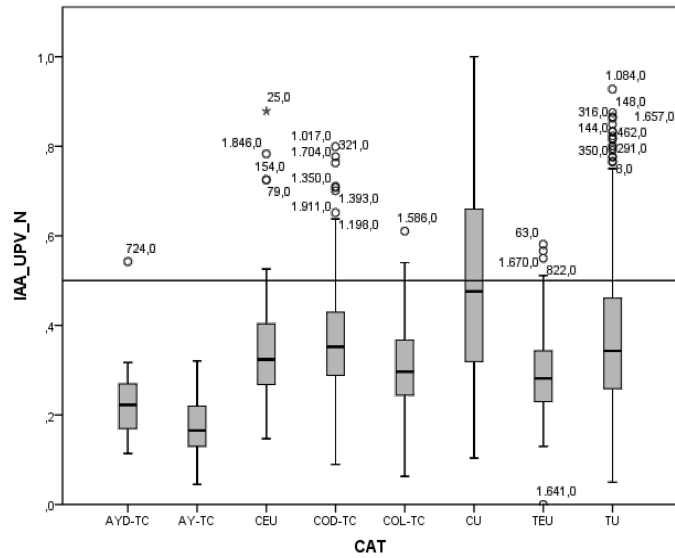


Figura 7.2. Distribución del escenario IAA_UPV_N por categoría laboral

La figura 7.3 muestra la contribución de la docencia, investigación y gestión a la formación del Índice de Actividad Académica para los distintos grupos de profesorado según la categoría laboral. La contribución se ha definido como la parte del IAA_UPV que corresponde a cada una de las dimensiones. En el caso de la docencia, por ejemplo, la contribución para el profesor c se ha obtenido mediante la siguiente ecuación:

$$Contribución_{DOC_c} = \frac{67,60 * IAD_c^*}{IAA_{UPV_c}}$$

siendo el valor IAD_c^* la puntuación obtenida por el profesor c en la dimensión de docencia y el valor IAA_{UPV_c} la puntuación del índice de actividad académica obtenida por el profesor c en el escenario de la UPV. De forma análoga, se ha calculado el valor para $Contribución_{INV_c}$ y $Contribución_{GES_c}$.

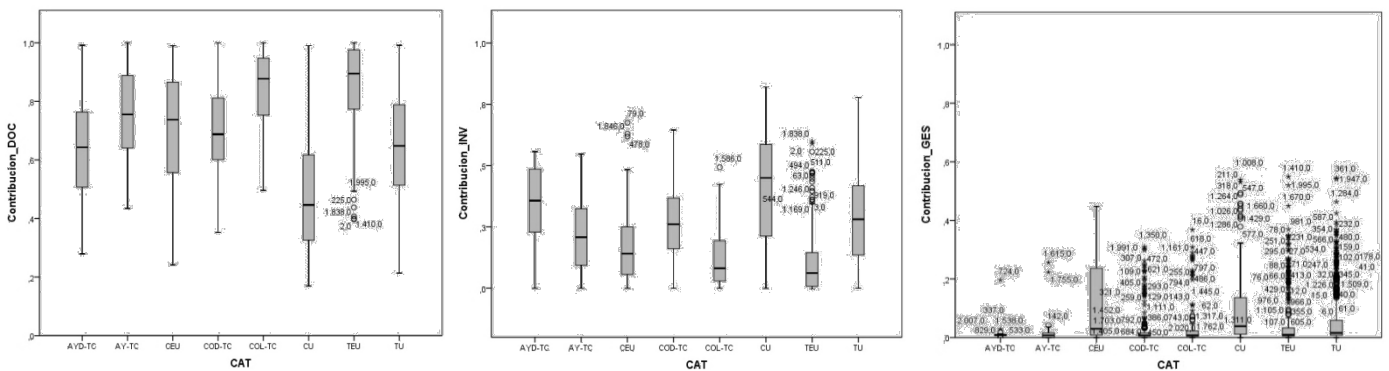


Figura 7.3. Contribución de las dimensiones al IAA por categoría laboral para el escenario de la UPV

Como puede apreciarse en la figura 7.3, la dimensión de la docencia es la que más contribuye, en promedio, al IAA_UPV de los profesores Titulados de Escuela Universitaria y Colaboradores. Por el contrario, la docencia no contribuye en gran medida a la formación del IAA_UPV para los profesores Catedráticos de Universidad. No obstante, en general se observa que los profesores tienen un IAA_UPV en el que contribuye en gran medida la dimensión docente.

Por lo que respecta a la I+D+i, esta dimensión contribuye en gran medida al IAA_UPV de los profesores Catedráticos de Universidad, siendo en este caso el profesorado Titular de Escuela Universitaria y Colaborador el que presenta un IAA_UPV con una baja contribución de la investigación.

La última parte de la figura 7.3 muestra la contribución de la gestión al IAA_UPV del profesorado. En general, se observa que existe una menor contribución de esta dimensión. Cabe destacar a los Catedráticos de Universidad y Catedráticos de Escuela Universitaria por la variabilidad existente dentro de sus grupos en cuanto a la contribución de la gestión en la formación de su índice.

Representación del IAA_UPV en coordenadas baricéntricas

Con el objetivo de analizar de forma más detallada los diferentes perfiles de profesorado obtenidos mediante el escenario IAA_UPV, se han transformado las coordenadas cartesianas de las contribuciones a las dimensiones (α, β, γ) a coordenadas baricéntricas (x, y) para obtener una representación gráfica que ayude a interpretarlos. Para ello, se han aplicado las siguientes transformaciones:

$$\begin{aligned}x &= \alpha + 2 * \beta + 1,5 * \gamma \\y &= \alpha + 1 * \beta + (2 + \sqrt{3})/2 * \gamma\end{aligned}$$

con $\alpha + \beta + \gamma = 1$. La figura 7.4 muestra los resultados obtenidos para el escenario IAA_UPV.

Las coordenadas baricéntricas permiten parametrizar, mediante $n + 1$ números reales en un intervalo prefijado, el interior de un n -simplex. En este caso se tienen 3 números reales para cada profesor correspondientes a las dimensiones de docencia, investigación y gestión, y se representan en un 2-simplex equivalente a un triángulo equilátero. Estos 3 números reales representan la contribución de las dimensiones a la formación del IAA del profesorado.

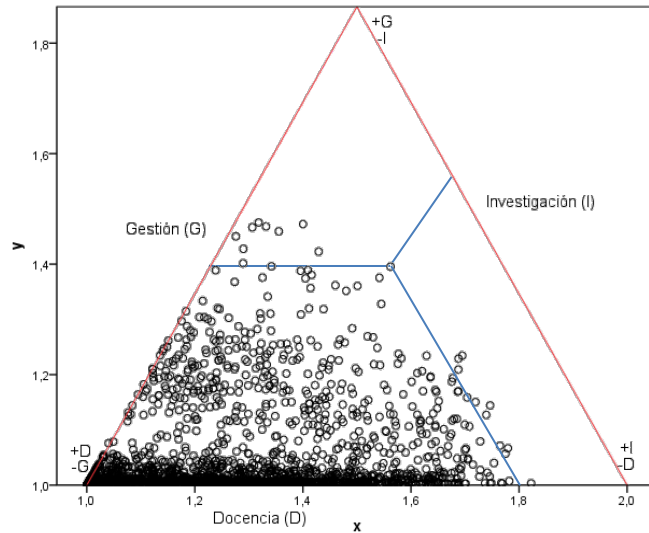


Figura 7.4. Contribuciones de las dimensiones al IAA_UPV en coordenadas baricéntricas

En la figura 7.4 se aprecia que la mayoría del profesorado evaluado, mediante el modelo de la UPV, tiene asignado un IAA_UPV en el que la dimensión de la docencia presenta una alta contribución, mientras que las dimensiones de investigación y gestión contribuyen en menor medida a la formación de éste, teniendo en cuenta las proyecciones de los puntos en los ejes correspondientes.

Una forma de estudiar cómo funciona el escenario IAA_UPV es observar lo que ocurre con el profesorado con puntuaciones más altas y más bajas. Para ello, se ha analizado con más detalle a los profesores que han obtenido una alta puntuación en el IAA_UPV considerando, por ejemplo, una valoración superior a 0,8. La figura 7.5 ilustra que se trata principalmente de profesores que tienen un perfil investigador, es decir, que la dimensión de investigación contribuye en gran medida a la formación de sus índices. Se trata de profesores que obtienen una elevada puntuación en el IAA_UPV, por lo que son profesores que realizan mucha actividad académica según el escenario de la UPV.

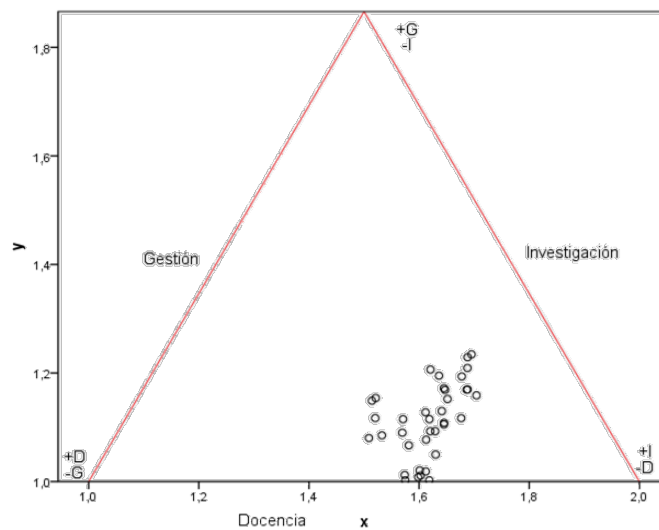


Figura 7.5. Contribuciones de las dimensiones para profesores con IAA_UPV > 0,8 en coordenadas baricéntricas

La figura 7.6 muestra las distintas categorías laborales a las que pertenece el grupo de profesores con una alta puntuación en el IAA_UPV. Estos son Catedráticos de Escuela Universitaria, Catedráticos de Universidad y Titulares de Universidad, y representan un 1,8% del total del profesorado evaluado. De lo que se deduce que el modelo no asigna a ningún profesor AY, AYD, TEU, COL y COD una puntuación IAA superior a 0,8.

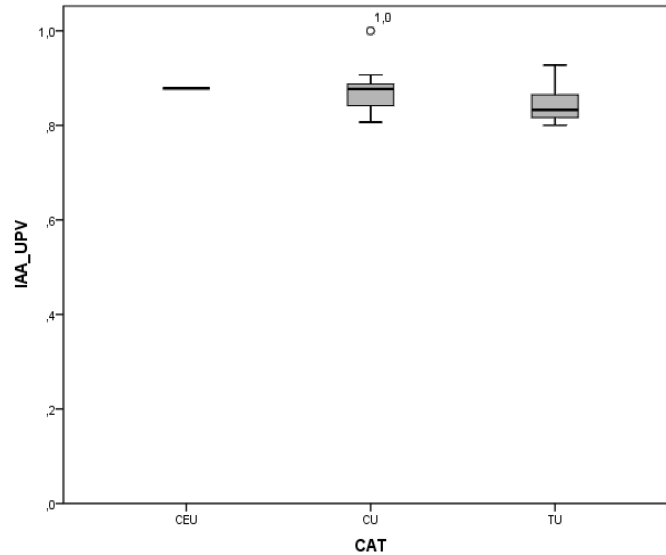


Figura 7.6. IAA_UPV por categoría laboral para el profesorado con IAA_UPV>0,8

En la figura 7.7 se observa claramente el perfil de este conjunto de profesores. Se trata de profesores que presentan un IAA_UPV con una baja contribución de las tareas docente y de gestión a éste, mientras que la contribución de las actividades de investigación es un poco más elevada.

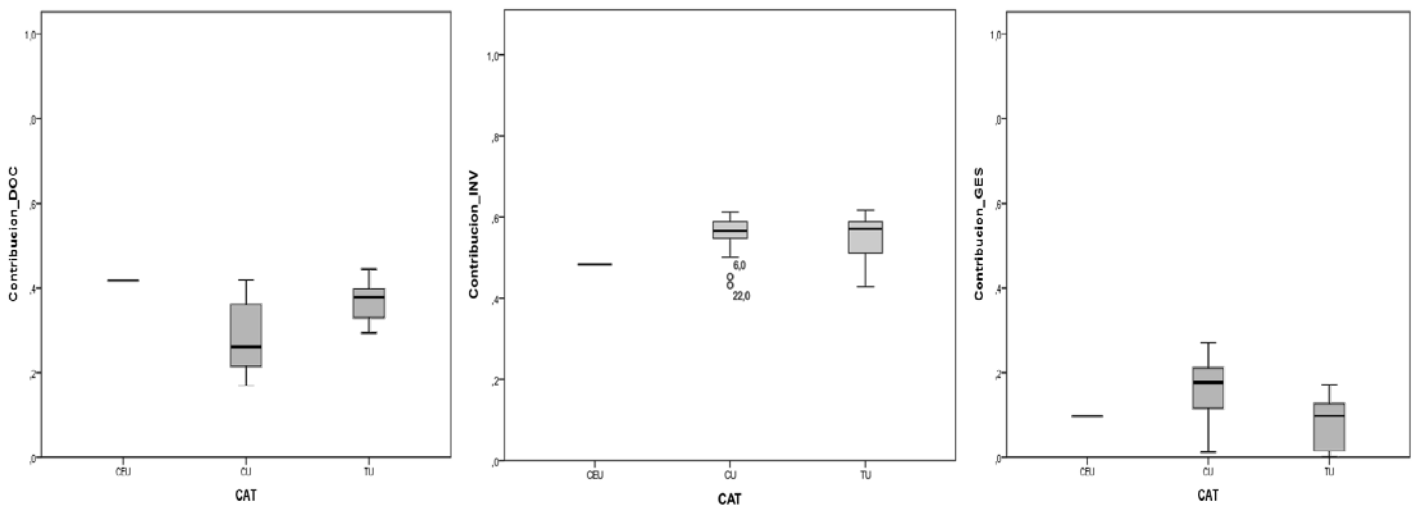


Figura 7.7. Contribución de las dimensiones para profesores con IAA_UPV>0,8

De forma análoga, se ha analizado el perfil de aquellos profesores que han obtenido una baja puntuación IAA_UPV, considerando por ejemplo una puntuación inferior a 0'2.

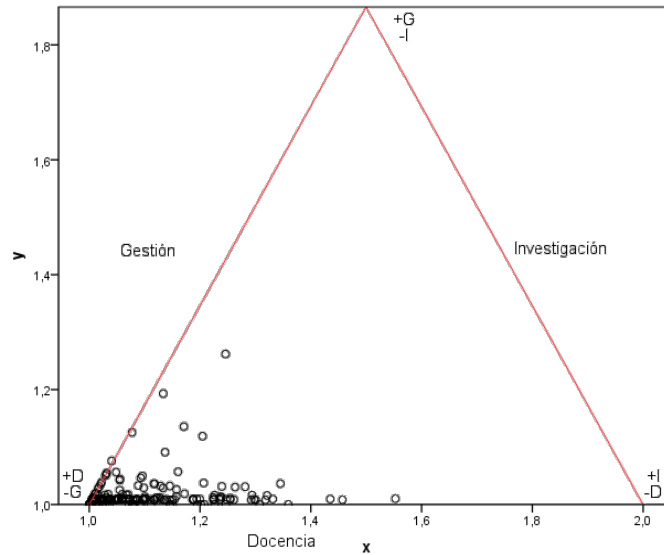


Figura 7.8. Contribuciones de las dimensiones para profesores con $IAA_UPV < 0,2$ en coordenadas baricéntricas

En este caso se trata de profesores con un perfil claramente docente y con baja contribución de las actividades de investigación y gestión a la formación de su IAA_UPV . Estos representan un 10,8% del profesorado evaluado. A continuación se muestra la distribución del IAA_UPV por categoría laboral y la contribución de las dimensiones a la formación del IAA_UPV para este grupo formado por profesores de todas las categorías laborales.

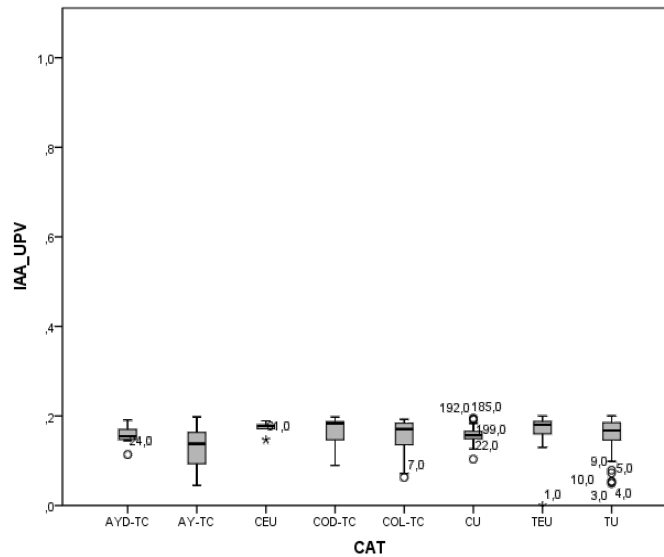


Figura 7.9. IAA_UPV por categoría laboral para el profesorado con $IAA_UPV < 0,2$

Cabe destacar que a diferencia del análisis del profesorado con un IAA_UPV superior a 0,8, en este caso el modelo asigna una valoración académica por debajo de 0,2 a profesores de todas las categorías laborales. La figura 7.10 muestra que se trata de profesores con una alta contribución de las actividades de docencia a la formación del índice y una baja contribución de las actividades de investigación y gestión.

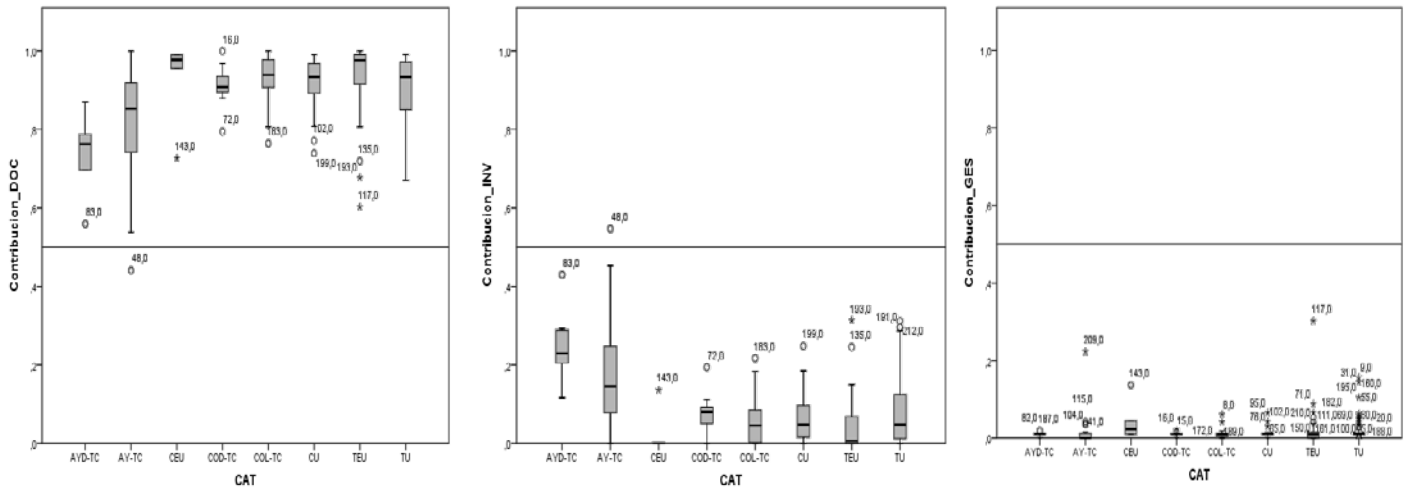


Figura 7.10. Contribución de las dimensiones por categoría laboral para profesores con IAA_UPV<0,2

Limitaciones del modelo IAA propuesto por la UPV

A continuación se exponen algunas de las limitaciones principales que presenta la construcción del IAA propuesto por la UPV y que justifican la propuesta metodológica de este último capítulo de la Tesis Doctoral.

Según el modelo de construcción propuesto por la UPV, la puntuación de cada dimensión correspondiente a cada profesor se normaliza dividiendo con respecto a los valores medios de las dimensiones. Teniendo en cuenta la figura 7.11, se observa que las dimensiones de gestión y de investigación presentan una clara asimetría en sus distribuciones.

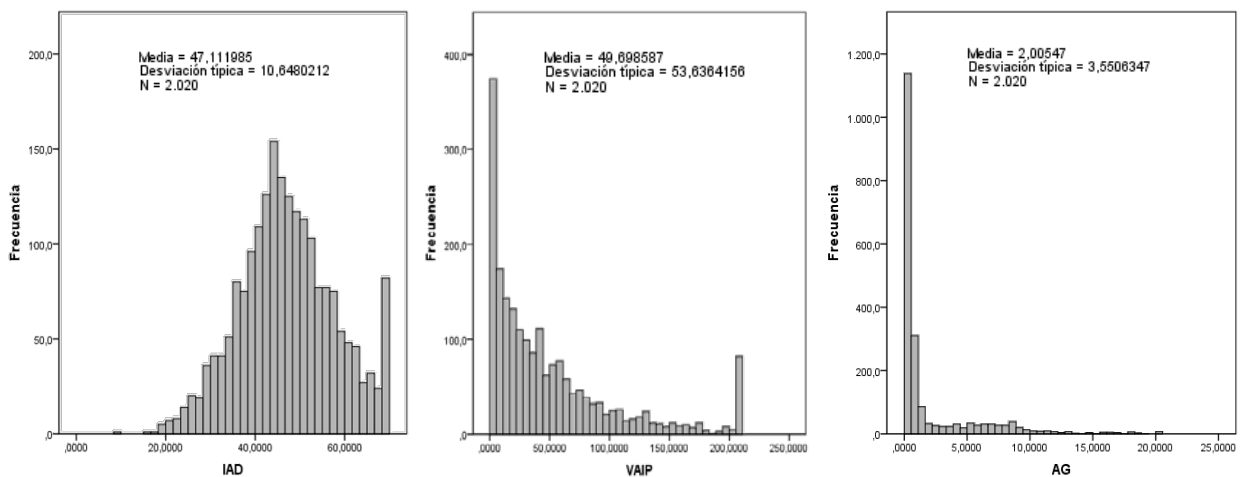


Figura 7.11. Histogramas de las dimensiones IAD, VAIP y AG

Por tanto, el estadístico de la media utilizado por la universidad para la normalización de los indicadores no es un buen parámetro de posición para normalizar las distribuciones de gestión e investigación.

Por otra parte, en el modelo planteado por la universidad se asignan unos pesos fijos para todo el profesorado en cada una de las dimensiones. Dichos pesos se formulan según los objetivos estratégicos de la universidad siendo para el año de evaluación 2011/2012 los siguientes: $p_{doc} = 67,60$, $p_{inv} = 26,40$ y $p_{ges} = 6$.

Una de las limitaciones que presenta la asignación fija de pesos es que puede favorecer a un grupo de profesores a la vez que perjudica a otro (Rogge, 2011). Es decir, si la universidad asigna un peso elevado a una de las dimensiones, el IAA puede favorecer a aquellos profesores con una alta puntuación en esa dimensión. Este resultado puede crear un gran debate crítico por parte del profesorado con respecto al modelo metodológico empleado en la construcción del IAA de la universidad, ya que todo el profesorado no tiene una misma concepción de la importancia de las distintas dimensiones de la actividad académica del PDI. Es por eso que los modelos propuestos en este capítulo se han basado en una metodología que asigna un conjunto de pesos particulares para cada profesor de forma que el IAA resultante sea máximo para el profesor evaluado con respecto al resto de profesores. Así, ningún profesor puede estar desconforme con el modelo de construcción del índice, ya que es máximo para cada uno de ellos según su perfil docente, investigador y/o gestor.

Finalmente, se ha aplicado un análisis de sensibilidad para evaluar la calidad del indicador compuesto IAA_UPV. El cálculo de los coeficientes de sensibilidad S_i ha permitido analizar si la importancia asignada por la universidad, en forma de peso nominal p_{doc} , p_{inv} , p_{ges} , a cada una de las dimensiones coincide con el efecto principal de cada una de éstas en la construcción del IAA.

La relación existente entre el IC y las dimensiones es lineal ($R^2 \cong 1$) y, además, el grado de correlación entre los indicadores es muy bajo según la tabla 7.3, por tanto, los coeficientes de sensibilidad de primer orden S_i , o conocidos también como efectos principales, serán muy similares a los coeficientes de regresión al cuadrado estandarizados de cada una de las dimensiones en la regresión múltiple (véase capítulo 3). Con el fin de comprobar esta afirmación se han calculado los coeficientes S_i mediante regresión lineal y mediante el procedimiento no paramétrico *State Dependent Parameter* (SDP). Para ello, se ha considerado el siguiente modelo en el que los pesos suman la unidad y se pueden comparar con los coeficientes de primer orden obtenidos mediante ambos procedimientos:

$$IAA_UPV_i = 0,6760 * \left(\frac{IAD_i^*}{43,48} * 100 \right) + 0,2640 * \left(\frac{VAIP_i^*}{38,11} * 100 \right) + 0,06 * \left(\frac{AG_i^*}{1,56} * 100 \right)$$

donde $i = 1, \dots, N$ (número de profesores evaluados).

Una vez normalizados a la unidad los coeficientes de primer orden obtenidos mediante el procedimiento SDP y mediante regresión lineal, las diferencias con respecto a los coeficientes de importancia asignados en el proceso de construcción del IC se muestran en la figura 7.12. En

la tabla 7.3 se observa que los resultados obtenidos mediante ambos métodos son prácticamente iguales.

	$S_{i,Regresion}$	$S_{i,SDP}$
IAD	0,194	0,197
VAIP	0,700	0,699
AG	0,098	0,103

Tabla 7.3. Efectos principales de las dimensiones del IAA

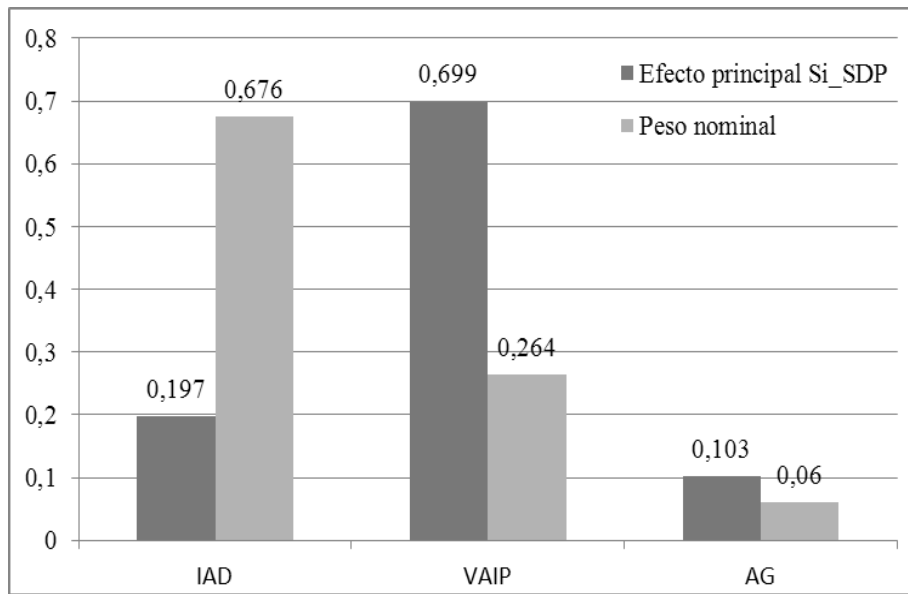


Figura 7.12. Pesos nominales y efectos principales de las dimensiones

Los resultados revelan que la propuesta del IAA_UPV asigna a la dimensión de gestión una importancia del 6% con respecto al total de la actividad académica, sin embargo, el impacto o efecto real de esta dimensión sobre el IAA es del 10,3%. De igual forma, la dimensión de investigación tiene un efecto real mayor (69,9%) que el esperado por el modelo para el desarrollo del IC (26,4%). De forma contraria, el indicador que mide la actividad docente del profesorado tiene un efecto sobre la formación del IC del 19,7% en la construcción del IAA y, según la universidad, debería de tener una importancia del 67,6%.

Las figuras 7.13, 7.14 y 7.15 que se muestran a continuación confirman los resultados anteriores.

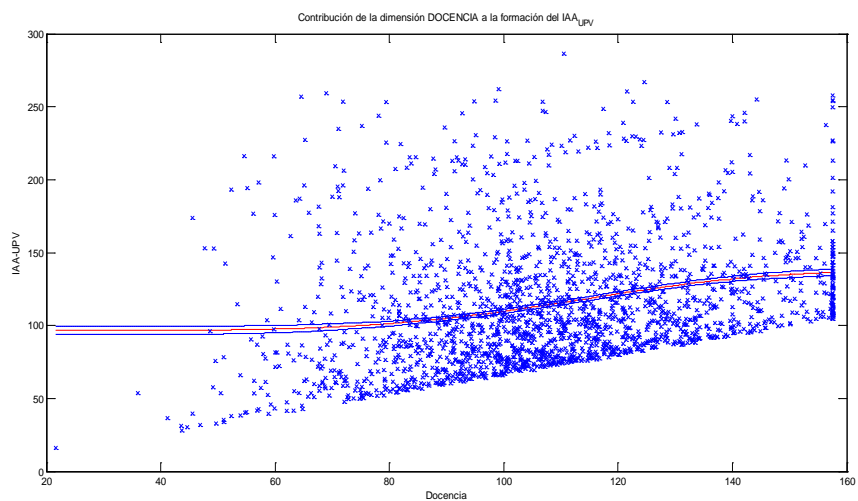


Figura 7.13. Efecto principal de la dimensión IAD a la formación del IAA_UPV

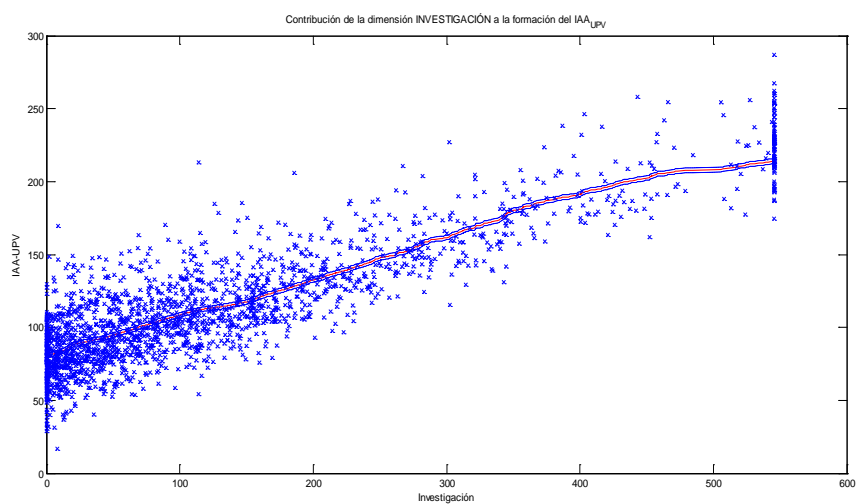


Figura 7.14. Efecto principal de la dimensión VAIP a la formación del IAA_UPV

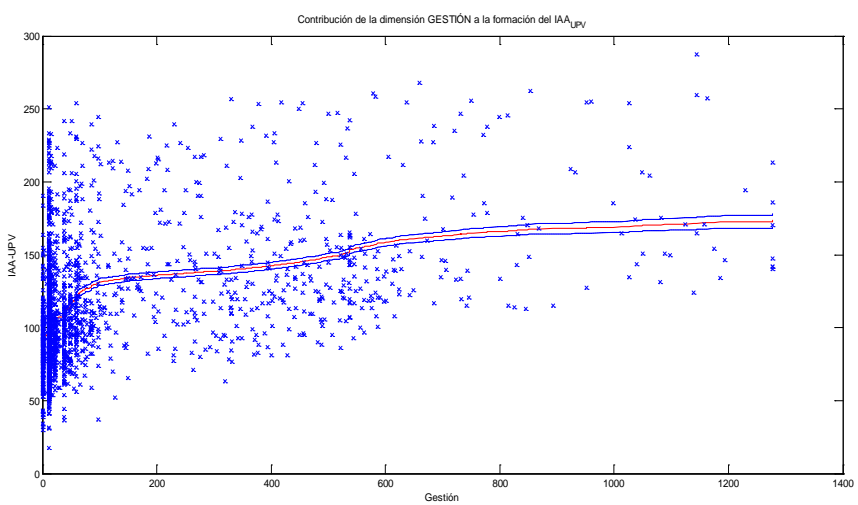


Figura 7.15. Efecto principal de la dimensión AG a la formación del IAA_UPV

Según el análisis de calidad del IAA_UPV realizado mediante el cálculo de los coeficientes de sensibilidad de primer orden, se concluye que las dimensiones evaluadas presentan una importancia sobre el IC diferente a la esperada por el modelo.

Estas observaciones justifican la necesidad de plantear modelos de construcción del IAA alternativos al propuesto por la UPV y basados en una metodología conocida como 'Beneficio de la Duda' descrita a continuación.

7.2.4. Ponderación y agregación mediante la técnica DEA-BOD

Introducción a la técnica DEA en el ámbito de la educación superior

La metodología DEA se ha aplicado en muchos trabajos de diversos ámbitos de estudio entre los que destacan el sector bancario, la educación superior, la salud, etc. En particular, son muchos los estudios que aplican esta técnica para medir la eficiencia de las universidades. Sin embargo, según los distintos objetivos de estudio, cada uno de ellos se basa en un marco teórico diferente y, por tanto, en una selección de *inputs* y *outputs* diferente para medir las distintas actividades de las universidades. La mayoría de estos estudios empíricos se centran en comparar la eficiencia productiva de las distintas universidades dentro de un mismo país: en 1988, Ahn, Charnes y Cooper (1988) estudian la eficiencia de 161 universidades de doctorado (*doctoral-granting universities*) en Estados Unidos; Athanassopoulos y Shale (1997) estudian la eficiencia en costes y eficiencia técnica de 45 universidades para el curso 1992-93 en Reino Unido; McMillan y Datta (1998) evalúan la eficiencia relativa de 45 universidades canadienses; Ng y Li (2000) evalúan la eficiencia relativa de 84 IES de China en el período 1993-95; Abbot y Doucouliagos (2003) evalúan en el año 1995 la eficiencia de 36 universidades australianas; Warning (2004) estudia el rendimiento de 73 universidades públicas alemanas; Carrington, Coelli y Prasada (2005) miden la eficiencia productiva de 35 universidades australianas en el período 1996-2000; Johnes y Johnes (2006; 2009) evalúan la eficiencia de universidades en el Reino Unido y, finalmente, el estudio de Gómez y Mancebón (2012) se centra en evaluar en el año 2000 la eficiencia técnica de las 47 universidades públicas españolas divididas en tres grandes grupos de tipología de universidades: técnicas, no técnicas y generalistas.

Estudios alternativos se centran en comparar la eficiencia de las universidades en distintos países, como es el caso de la investigación dirigida por Agasisti y Pérez-Esparrells (2010) en la que se compara la eficiencia de las universidades españolas e italianas.

Otro grupo de estudios tienen como objetivo evaluar la eficiencia a nivel departamental o de área de conocimiento, como es el caso del trabajo pionero de Tomkins y Green (1988) en el que se analiza la eficiencia docente e investigadora de 20 departamentos de contabilidad británicos. Además, estos autores concluyen en su trabajo la siguiente afirmación:

“This experiment offers some hope that DEA, carefully and sensitively used, can offer additional insights on performance which are not available from other methods of assessment.”

En esta línea, Johnes y Johnes (1995) evalúan la eficiencia de 36 departamentos de economía del Reino Unido, pero sólo teniendo en cuenta la actividad investigadora. Otro ejemplo es el estudio realizado por Beasley (1995) en el que se evalúa la eficiencia mediante las actividades de docencia e investigación de 52 departamentos de química y 50 de física en universidades del Reino Unido.

En España la mayoría de investigaciones en las que se aplica la técnica DEA pertenecen a este último grupo de estudios. En los estudios por área de conocimiento destacan, entre muchos otros, el trabajo de Pina y Torres (1995) en el cual se evalúan 22 departamentos de contabilidad en el curso 1991-92 teniendo en cuenta la función docente e investigadora, y el trabajo de Martínez Cabrera en el cual se analiza la eficiencia investigadora de una muestra de departamentos de economía de las universidades públicas españolas (Martínez, 2000). También se han realizado trabajos referidos a departamentos que pertenecen a una misma universidad, por ejemplo: el estudio realizado en la Universidad de La Laguna donde se analiza el grado de eficiencia docente e investigadora de los departamentos (Martín, 2008), el trabajo realizado por Giménez y Martínez (2006) en el cual se estudia la eficiencia docente e investigadora de 42 departamentos de la Universitat Autònoma de Barcelona y el estudio realizado por Ramón, Ruiz y Sirvent (2010) en el que se evalúa la eficiencia docente de 20 departamentos de la Universidad Miguel Hernández.

De forma alternativa a los trabajos revisados, en este estudio se han considerado como unidades de análisis a los profesores de la UPV con el objetivo de evaluar su eficacia relativa con respecto al resto de profesores en términos de su actividad académica desarrollada. Para ello, se ha aplicado el enfoque BOD descrito a continuación.

El enfoque BOD en la construcción de Indicadores Compuestos

El enfoque BOD, utilizado para la construcción de indicadores de rendimiento de conceptos multidimensionales, tiene sus raíces en el DEA, una técnica ampliamente conocida en el campo de la estimación de unidades productivas en cualquier ámbito de aplicación y, en particular, en la educación superior. La metodología DEA es una técnica de programación lineal no paramétrica que evalúa la eficiencia de un conjunto de unidades productivas homogéneas (empresas, facultades, individuos, profesores, etc.) mediante el uso de variables input y output en un entorno caracterizado por no tener información sobre el sistema de pesos de los input y output, y no conocer la ‘forma funcional’ de la función que relaciona las variables. Una descripción gráfica de esta metodología se presenta en el capítulo 3 y una descripción analítica en el anexo A.

El paralelismo de esta técnica con la construcción de ICs resulta evidente. En el caso de la metodología DEA, se pretende medir la eficiencia de unidades de análisis mediante un modelo formulado con variables input y output del que no se conoce información sobre los pesos. Por otra parte, en el caso de la construcción de un IC el modelo está formulado por distintos criterios (o indicadores) de los cuales no se conoce un peso exacto¹⁰. La única diferencia que existe con respecto al modelo tradicional DEA es que la construcción de ICs sólo se fija en el logro de los objetivos, variables output, sin tener en cuenta de forma explícita las dimensiones de los parámetros de entrada, variables input. Por tanto, en el modelo aplicado para la construcción de ICs se asume que todas las unidades de análisis tienen una variable ‘dummy input’ igual a la unidad. Según Lovell, Pastor y Turner (1995), el input igual a 1 para todas las unidades de análisis puede ser interpretado como una unidad referencial que persigue varios objetivos en el tema de estudio. Esta técnica aplicada al campo de los ICs se conoce como enfoque del “Beneficio de la Duda” (BOD) y fue introducida en 1991 por Melyn y Moesen (1991).

Existen muchos estudios en los que se han construido ICs mediante el uso de la metodología BOD. Algunos ejemplos son: la evaluación del rendimiento de los profesores del departamento de negocios y administración de la Universidad HU de Bruselas (Rogge, 2010; De Witte y Rogge, 2011), la construcción de un índice que evalúe el progreso en materia de banda ancha de internet (Badasyan, Shideler y Silva, 2011), la construcción del índice de mercado interno (Cherchye et al., 2007a), la construcción del índice basado en la economía del conocimiento (Cherchye et al., 2011), la evaluación del rendimiento de compañías de construcción en Portugal (Horta, Camanho y Moreira da Costa, 2012), la construcción de un índice de desempeño ambiental (Rogge, 2012), la construcción de un índice que mide la presencia de economías de alcance entre la docencia y la investigación (De Witte et al., 2013), la construcción de un IC con el fin de establecer una política de inclusión social (Cherchye, Moesen y Van Puyenbroeck, 2004), propuesta de construcción del índice de desarrollo humano (Despotis, 2005), propuesta de construcción del índice de desarrollo tecnológico del informe de desarrollo humano (Cherchye et al., 2008).

El modelo BOD se caracteriza por seleccionar los pesos de los indicadores de forma endógena en el conjunto de datos observable, siempre que no se conozca información detallada sobre el esquema de ponderación de los indicadores. Por tanto, la idea básica es situar el rendimiento de una unidad de análisis en relación con el resto de unidades y detectar aquellos indicadores que presentan una alta y baja puntuación. De esta forma, el modelo asigna un mayor peso a los primeros indicadores y un menor peso a los segundos, es decir, maximiza el impacto de los indicadores con alta puntuación y minimiza la influencia de los indicadores con poca puntuación¹¹. De ahí deriva el nombre del enfoque BOD, ya que el modelo selecciona el

¹⁰ Se puede disponer de un peso exacto en el caso en que exista un acuerdo unánime sobre la importancia asignada a cada indicador simple

¹¹ Cherchye et al. (2007b) remarcan que esto no significa que el modelo BOD asigna automáticamente todo el peso a aquel indicador en el que la unidad de análisis tiene una mayor puntuación (comparada con el resto de indicadores) y cero al resto. Puede ocurrir que asignando un peso distinto de cero a todos los

conjunto de pesos más favorable para cada unidad de análisis. Esto significa que el modelo concede a las unidades de análisis el beneficio de la duda cuando asigna los pesos a los correspondientes criterios o indicadores. Los pesos resultantes del modelo BOD, $w_{j,i}$, son óptimos en el sentido que son elegidos con el objetivo de maximizar el IC de cada unidad de análisis con respecto al resto de unidades. Otro esquema de ponderación de los indicadores simples para cada unidad de análisis proporcionaría una puntuación para el IC más baja.

Por tanto, el esquema de incertidumbre sobre los pesos en el modelo BOD tiene una clara identificación con la construcción de ICs cuando no se llega a un acuerdo sobre el peso asignado a cada indicador. La ausencia de un consenso en el conjunto de pesos es un tema bastante habitual en el desarrollo de ICs.

De forma analítica, el modelo se expresa como sigue (Cherchye et al., 2007b):

$$IC_c^* = \max_{(w_{c,1}, \dots, w_{c,q})} \frac{\sum_{i=1}^q w_{c,i} y_{c,i}}{\max_{(y_{1,i}, \dots, y_{n,i})} \sum_{i=1}^q w_{c,i} y_{j,i}} \quad (7.1)$$

s. a

$$\sum_{i=1}^q w_{c,i} y_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, n \quad (7.1a)$$

$$w_{c,i} \geq 0 \quad i = 1, \dots, q \quad (7.1b)$$

donde:

IC_c^* es la puntuación óptima del indicador compuesto para la unidad de análisis c

$y_{c,i}$ es la puntuación del indicador i para la unidad de análisis c

$w_{c,i}$ es el peso óptimo del modelo BOD asignado al indicador i de la unidad de análisis c

q es el número de indicadores del estudio

n es el número de unidades de análisis en el conjunto Y

La forma ratio de la función objetivo muestra la idea de *benchmarking* del modelo, puesto que la puntuación del IC de la unidad de análisis bajo evaluación, c , no viene dada por la suma ponderada de las puntuaciones de sus indicadores, sino por el ratio con respecto a una unidad de análisis *benchmark* con un IC calculado bajo el mismo esquema de ponderación. Sin embargo, la restricción (7.1a) permite reformular el modelo de tal forma que la puntuación del IC se calcule como la suma ponderada de las puntuaciones de los indicadores (la restricción de normalización establece una cota para el IC de la unidad benchmark igual a la unidad, por lo que el denominador de (7.1) desaparece).

Nótese que la unidad de referencia o benchmark del modelo (7.1) se obtiene del conjunto de unidades de análisis ($y_{j,i} \in Y$). Por tanto, el IC de la unidad de análisis bajo estudio se calcula teniendo en cuenta el resto de unidades de análisis de la muestra Y , en lugar de calcularse con

indicadores se consiga una puntuación en el IC más alta que la puntuación obtenida al asignar todo el peso a un mismo indicador, aunque este se comporte como el ‘mejor’.

unidades de referencia externas. En particular, el benchmark del denominador de (7.1) se obtiene mediante un proceso de optimización y corresponde a la unidad de análisis del conjunto Y que, bajo el esquema de ponderación óptima para la unidad evaluada c , obtiene una puntuación máxima para el IC. Esto significa que el benchmark se elige mediante un procedimiento endógeno y puede ser diferente para cada unidad de análisis evaluada.

La forma ratio de la función objetivo también facilita la interpretación del problema, de tal forma que $IC_c = 1$ indica un nivel de eficacia relativa¹² global para la unidad de análisis c igual a la eficacia de su benchmark. Esto significa que no existe ninguna unidad de análisis observada que sea más eficaz, con el conjunto de pesos óptimo de la unidad evaluada c , que la misma unidad de análisis bajo estudio. En este caso se puede decir que la unidad evaluada actúa como su propio benchmark y que se trata de una unidad que manifiesta la “mejor práctica”. Por otra parte, un $IC_c < 1$ revela que existe al menos una observación que con los pesos óptimos de la unidad c es más eficaz que la propia unidad evaluada. La diferencia entre el valor obtenido y el valor máximo posible muestra las carencias de dicha unidad y su potencial horizonte de mejora.

El modelo BOD básico con las restricciones (7.1a) y (7.1b) se caracteriza por ser muy flexible en cuanto a la elección de los pesos óptimos para cada unidad de análisis. En concreto, en el modelo se deben satisfacer sólo dos restricciones. La primera restricción de normalización (7.1a) indica que cuando se aplica el conjunto de pesos óptimos $w_{c,i}$ para la unidad evaluada c al resto de unidades de la muestra Y ninguna debe obtener una puntuación del IC superior a 1¹³. Por tanto, las puntuaciones de todas las unidades de análisis se encuentran entre 0 y 1 ($0 < IC_j \leq 1, \forall j \in Y$) donde puntuaciones más altas indican una mayor eficacia relativa con respecto al resto. Una vez más esta restricción subraya la idea de benchmarking del modelo. El conjunto de pesos óptimo de la unidad bajo evaluación c se aplica al resto de unidades de la muestra Y . De esta forma, se observan las unidades que son menos, igual y más eficaces que la unidad evaluada aplicando el conjunto de pesos óptimo para ésta. Si existen unidades más eficaces que la unidad c (con un IC mayor) se puede decir que existe evidencia de que c es “superada” por otras unidades de la muestra en términos del concepto medido, por lo que la unidad con mayor rendimiento se considera un benchmark para la unidad evaluada c . Nótese que esta interpretación es muy intuitiva y sencilla para transmitir al público objetivo. Cuanto más similar sea la puntuación IC_c a la unidad, más cerca estará la unidad c a su benchmark. Si IC_c es próximo a cero la unidad c deberá mejorar su eficacia relativa para conseguir la misma eficacia que su unidad benchmark.

Por otro lado, la segunda restricción (7.1b) indica que el IC es una función creciente de los indicadores parciales. Por tanto, una mejora en un indicador en concreto implicará una mejora en la puntuación del IC final.

¹² Nótese que en este trabajo se considera el término ‘eficacia relativa’ en el modelo BOD puesto que las unidades de análisis se diferencian entre ellas por las variables output de resultados, ya que la variable input es constante para todas

¹³ Este número es arbitrario. Se podría fijar cualquier otra cota de normalización

Considerando que por construcción (en particular, la restricción (7.1a)) la unidad benchmark siempre alcanza el valor 1 para el IC, la función objetivo del modelo se puede reformular como sigue:

$$IC_c^* = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^q w_{c,i} y_{c,i} \quad (7.2)$$

s. a

$$\sum_{i=1}^q w_{c,i} y_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, n \quad (7.2a)$$

$$w_{c,i} \geq 0 \quad i = 1, \dots, q \quad (7.2b)$$

De esta forma, el modelo se convierte en un problema de optimización lineal mucho más sencillo de computar. Por tanto, el modelo (7.2) es equivalente al modelo CCR-DEA de orientación input con múltiples outputs y constantes inputs, en el que los indicadores representan los diferentes outputs y se asigna como input una variable dummy con valor 1 a cada unidad de análisis (véase anexo A).

Ventajas del modelo BOD

El modelo BOD presenta ventajas tanto a nivel conceptual como a nivel metodológico. La primera ventaja del modelo es la **flexibilidad en cuanto a la modelización** de la técnica subyacente. Las técnicas paramétricas deben asumir una forma paramétrica para la función de producción o de coste que relaciona las variables independientes con la(s) variable(s) dependiente(s), mientras que el modelo BOD no requiere de una forma conocida (Charnes et al., 1994).

Por otra parte, la **flexibilidad en la ponderación** de los indicadores es una característica muy atractiva de la técnica. El modelo BOD asigna un conjunto de pesos para cada unidad de análisis de forma relativa al resto de unidades de la muestra Y , de tal forma que se obtiene una puntuación en el IC óptima para cada unidad. Por tanto, la aplicación del modelo es adecuada en el caso en que no se llegue a un acuerdo en los pesos asignados a los criterios o indicadores. Esta flexibilidad en los pesos provoca una gran aceptación del método por el público objetivo. Boussofiane, Dyson y Thanassoulis (1991) consideran que dicha flexibilidad en la elección de los pesos es una fortaleza y una debilidad del modelo DEA. Según los autores, se trata de una fortaleza porque si una unidad es ineficiente incluso cuando se han incorporado los pesos más favorables en su medida de eficacia relativa, entonces el argumento de que los pesos no son apropiados no es justificable. Por otra parte, afirman que es una debilidad porque una elección poco sensata de los pesos puede permitir calificar como eficaz a una unidad, aunque esto tenga más que ver con la elección de pesos que con cualquier eficacia relativa inherente.

Además, los pesos se calculan de tal forma que cada unidad de análisis obtenga una **puntuación óptima para el IC**. De forma concreta, el modelo asigna un conjunto de pesos que

maximiza el rendimiento de cada unidad de análisis comparada con el resto. Esta propiedad puede resultar muy atractiva y provocar menos rechazo que otros enfoques en la construcción de ICs. Esto se debe a que el modelo no exige que todas las unidades concedan la misma importancia a cada indicador parcial y, por tanto, se respetan las peculiaridades de cada unidad permitiendo que éstas escojan los pesos que más las beneficien con respecto al valor del índice.

Aunque esta flexibilidad en la ponderación pueda parecer inicialmente excesiva, el enfoque BOD permite graduarla con la incorporación de **restricciones adicionales** sobre los pesos (véase sección titulada “El modelo BOD con restricciones en los pesos”). Esta propiedad puede ser de gran utilidad en el caso en que se disponga de información adicional sobre los pesos (bien sea la opinión de expertos en el campo de estudio, resultados de investigaciones empíricas, etc.). Tal y como afirman Foster y Sen (1997): “*while the possibility of arriving at a unique set of weights is rather unlikely, that uniqueness is not really necessary to make agreed judgments in many situations*”. Esta “libertad controlada” en la fijación de ponderaciones sobre los indicadores constituye una de las principales ventajas de la construcción de ICs mediante la técnica BOD.

Por otra parte, el modelo BOD **selecciona los pesos de forma endógena** con el fin de maximizar el IC de cada unidad de análisis. Por tanto, el esquema de pesos refleja las fortalezas y debilidades de las unidades de análisis en cada indicador evaluado. En concreto, valores elevados en los pesos indican que la unidad de análisis evaluada tiene un rendimiento alto en los correspondientes indicadores con respecto al resto de unidades y viceversa.

El modelo BOD, además de proporcionar una puntuación para cada unidad de análisis, presenta la ventaja de **identificar la contribución de los factores**, es decir identificar aquellos que explican en mayor y menor medida la puntuación del IC de cada unidad de análisis. En este sentido, los dos resultados proporcionados por el modelo son de gran utilidad. Por una parte, la puntuación del IC combina toda la información y proporciona una “imagen general” del concepto multidimensional que se está midiendo útil para los tomadores de decisiones. Por otra parte, la contribución de los factores proporciona información personalizada para cada unidad de análisis, de forma que ayuda a diagnosticar las deficiencias en los distintos criterios medidos y a definir estrategias de mejora. Así pues, se suele destacar la riqueza de los resultados ofrecidos por esta técnica de evaluación, resultados que pueden constituir una base de sumo interés en la implementación de estrategias de mejora en el objeto de estudio.

Otra característica importante de la técnica BOD es la **invarianza en la unidad de medida**. El valor del IC es independiente de las unidades de medida de los sub-indicadores. Una demostración matemática de esta propiedad de invarianza se puede encontrar en (Cooper, Seiford y Tone, 2000). La razón fundamental por la que el modelo cumple la propiedad de invarianza procede de la forma endógena en la elección de los pesos que implica flexibilidad y, por tanto, libertad para que los pesos se adapten a las unidades de medida de los indicadores. Por consiguiente, no es necesario aplicar ningún tipo de normalización previa sobre los indicadores medidos en escalas ratio (precios, porcentajes,...) para hacerlos comparables, puesto que la técnica BOD es invariante en la unidad de medida de los indicadores (Cherchye et

al., 2007b). Sin embargo, aunque el IC derivado de la aplicación de la técnica BOD es invariante en la unidad de medida, se debe tener cuidado en la interpretación de los pesos, porque estos siguen dependiendo de la unidad de medida. Además, si se añaden restricciones adicionales puede resultar muy difícil la interpretación de éstas cuando cada indicador tiene una unidad de medida diferente. Existen dos formas de lidiar con este problema: la primera es aplicar una técnica de normalización para poder interpretar los pesos, tal y como han aplicado los autores Cherchye, Moesen y Van Puyenbroeck (2004) en la construcción del IC de inclusión social, y la segunda es centrar la interpretación en las “porciones de los indicadores” (“*pie-shares*”, en adelante)¹⁴, que son independientes de las unidades de medida. Los *pie-shares* son el producto del valor original de cada indicador $y_{c,i}$ y su peso asignado por el modelo BOD $w_{c,i}$. Por tanto, cambiar la unidad de medida de los indicadores no afecta al valor de la porción de indicador $P_{c,i}^* = y_{c,i} * w_{c,i}$.

Por tanto, volviendo al modelo (7.2), el IC de cada unidad de análisis puede interpretarse como la suma de los *pie-shares* de los indicadores:

$$IC_c^* = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^q w_{c,i} y_{c,i} = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^q P_{c,i}^* ;$$

donde $P_{c,i}^*$ indica la contribución de cada indicador i ($i = 1, \dots, q$) al IC global para la unidad de análisis c (Cherchye et al., 2007b).

Limitaciones del modelo

El modelo BOD también presenta algunas limitaciones en su metodología. Por un lado, el IC se construye como agregación lineal de los indicadores parciales, por lo que implícitamente se está aplicando una compensación entre éstos. Además, cuando se introducen las restricciones adicionales para reducir la flexibilidad en la ponderación se están introduciendo opiniones subjetivas en la construcción del modelo. Aunque, tal y como apuntan Pedraja-Chaparro, Salinas-Jimenez y Smith (1997), el modelo DEA-BOD básico ya presenta un cierto grado de subjetividad al elegir los inputs y outputs que se deben incluir en el problema de optimización para medir el rendimiento de las unidades de análisis. Sin embargo, los autores afirman que este grado de subjetividad no invalida el uso de la metodología DEA-BOD, puesto que cualquier análisis de eficacia relativa siempre se basa en opiniones subjetivas.

Además, tal y como señala Doménech (1992), la metodología DEA-BOD es sensible a la existencia de observaciones extremas y toda desviación respecto de la frontera es tratada como ineficiencia, lo que puede derivar en una sobreestimación de la misma. Por tanto, se debe realizar un tratamiento de datos previamente a la aplicación del modelo.

¹⁴ En DEA se conocen como “*virtual outputs*” y en BOD como “*pie-shares*” (Cherchye et al., 2007b)

Por otra parte, tal y como ya se ha mencionado, permitir un cierto grado de flexibilidad en la asignación de pesos es deseable desde un punto de vista práctico. En el caso en que los expertos, tomadores de decisiones, unidades evaluadas etc. no lleguen a un consenso en los pesos es adecuado que éstos se obtengan de tal forma que se optimice el rendimiento de cada unidad de análisis evaluada con respecto al resto. Así, el proceso es aceptado por el público evaluado, puesto que nadie puede estar en desacuerdo con el conjunto de pesos que se le ha asignado para el cálculo del IC.

Sin embargo, la flexibilidad excesiva en la ponderación del modelo (7.2) puede producir escenarios extremos no justificables para ciertas unidades de análisis. En concreto, el modelo puede asignar pesos que no estén en concordancia con las opiniones de los expertos. Por ejemplo, el modelo básico BOD puede asignar un peso de cero, o prácticamente cero, a indicadores en los que las unidades de análisis tienen un rendimiento bajo, puesto que la restricción del modelo (7.2b) lo permite. Por el contrario, puede ocurrir que el modelo asigne un peso muy elevado a aquellos indicadores en los que las unidades tienen un alto rendimiento, siendo estos indicadores menos importantes¹⁵ según la opinión de expertos. Como consecuencia, se pueden dar situaciones en las que las unidades de análisis se evalúen ignorando algunos indicadores y focalizando la atención en sólo unos pocos. Por tanto, al aplicar el modelo básico BOD se corre el riesgo de obtener puntuaciones IC para unidades de análisis basadas en un subconjunto de los indicadores observables y no en la totalidad del conjunto de indicadores definidos para la medición del concepto que se pretende medir. Tal y como apuntan Roll, Cook y Golany (1991) es muy difícil aceptar que después de una selección justificada de los indicadores por parte de los expertos se asigne un peso de cero a alguno de ellos y, por tanto, no se tengan en cuenta para el cálculo del índice global. Además, esto puede producir que las unidades de análisis se centren en mejorar el rendimiento de un subconjunto de indicadores e ignorar el resto de ellos. No obstante, muchas veces la asignación de ceros en algunos de los indicadores está justificada desde el punto de vista teórico.

Por otra parte, según Pedraja-Chaparro, Salinas-Jimenez y Smith (1997) es inaceptable asumir que la importancia relativa de cada input y output sea muy diferente entre las unidades de análisis, puesto que éstas son homogéneas en el sentido de producir la misma cantidad de outputs y tener el mismo objetivo. En esta línea, Schmidt (1986) apunta que la flexibilidad total no tiene sentido cuando las unidades de análisis emplean las mismas tecnologías, pagan con precios similares por los inputs y reciben similares outputs.

Por tanto, si se quiere medir un concepto multidimensional definido por un conjunto de indicadores que individualmente aportan una parte de información del concepto, el modelo básico BOD no es un instrumento válido para evaluar la eficacia relativa de las unidades de análisis. A continuación se introduce un ajuste del modelo básico BOD que mejora las deficiencias de éste y permite medir la eficacia relativa de unidades en conceptos

¹⁵ “Menos importante” no implica que no sea importante para la medición del concepto que se está evaluando. Si el ítem o indicador no fuese importante no se hubiese incluido en la lista de indicadores para medir el concepto multidimensional bajo evaluación. En este caso “menos importante” significa que existen indicadores que, según el grupo de expertos, son menos importantes que otros en el modelo

multidimensionales con mejores resultados, incorporando limitaciones en la flexibilidad de los pesos (Pedraja-Chaparro, Salinas-Jimenez y Smith, 1997).

El modelo BOD con restricciones en los pesos

Una forma de evitar los escenarios extremos es limitar la libertad en la elección de los pesos incorporando en el modelo las opiniones de los expertos. La inclusión de las opiniones de los expertos en el modelo ha atraído una considerable atención en la literatura de las técnicas DEA-BOD (Allen et al., 1997; Thanassoulis, Portela y Allen, 2004; Cherchye et al., 2007b; Cherchye et al., 2008). La metodología más común que se utiliza es la incorporación de restricciones adicionales en los pesos. De forma analítica, esta incorporación supone añadir la siguiente restricción general al modelo básico BOD:

$$w_{c,i} \in W_e \quad i = 1, \dots, q; e \in E \quad (7.2c)$$

donde W_e es el conjunto de los valores permitidos para los pesos basado en las opiniones de los expertos seleccionados $e \in E$. El resultado de la inclusión de (7.2c) en el modelo (7.2) constituye el modelo BOD restringido, en el que se calculan los pesos óptimos bajo el cumplimiento de unas restricciones definidas *a priori* que reflejan la opinión de las partes interesadas. El modelo BOD restringido se considera una herramienta muy valiosa para la construcción de ICs cuando no se obtiene un consenso en los pesos de los indicadores simples por parte de los expertos (Cherchye et al., 2007b).

En la literatura DEA-BOD se han propuesto distintos tipos de restricciones (p.ej. (Wong y Beasley, 1990; Thompson et al., 1990; Roll y Golany, 1993; Thanassoulis y Allen, 1998)). Cherchye et al. (2007b), siguiente entre otros a Wong y Beasley (1990), argumentan a favor de la incorporación de restricciones sobre los pie-shares, $P_{c,i}^*$, y no sobre los pesos absolutos, $w_{c,i}$, como introdujeron Dyson y Thanassoulis (1988) para un modelo con un input y varios outputs que fue generalizado en 1991 por Roll, Cook y Golany (1991). Esto se debe a la facilidad de interpretar los pie-shares, ya que no dependen de la unidad de medida e indican la contribución de cada indicador sobre el IC, y a la sencillez de informar a las unidades de análisis de su situación con respecto al resto y de las posibles mejoras en su actuación. En el campo de los ICs la invarianza del método frente a un cambio en la unidad de medida de los indicadores es una ventaja, ya que se puede trabajar con el conjunto de datos original y, por tanto, se evita la elección de un tipo de normalización (Cherchye et al., 2007b). Además, es mucho más fácil para los expertos asignar cotas que sean interpretadas como la contribución de cada indicador sobre el IC que asignar cotas sobre los pesos cuando cada indicador se mide en una unidad de medida diferente.

A continuación se enumeran y describen los tipos de restricciones sobre los pie-shares que más se nombran y aplican en la literatura DEA-BOD.

RESTRICCIONES ABSOLUTAS

Las restricciones absolutas sobre los pie-shares acotan la contribución de cada pie-share en un rango específico, determinado por una cota superior y una cota inferior (α_i y β_i , respectivamente):

$$\alpha_i \leq w_{c,i} y_{c,i} \leq \beta_i$$

Este tipo de restricciones es útil para evitar que los indicadores estén sobre o infravalorados en el cálculo del IC. Por ejemplo, la aplicación de esta restricción puede evitar que se asigne un peso nulo a algún indicador y, por tanto, evitar que se ignore alguno de los indicadores definidos para evaluar el concepto multidimensional bajo estudio. Sin embargo, el uso de este tipo de restricciones puede provocar incompatibilidades en el modelo y, por tanto, que no haya una solución factible en el problema de optimización. Esto se debe al efecto indirecto que tiene la asignación de una cota a uno de los indicadores sobre las cotas del resto de indicadores (véase Podinovski (1999)). En este caso, se deberían de ampliar los rangos de los pie-shares hasta obtener una solución al problema de optimización (Roll, Cook y Golany, 1991).

RESTRICCIONES ORDINALES

Supóngase que los expertos no llegan a un consenso sobre los pesos exactos que se deben asignar a cada pie-share, pero sí acuerdan que el indicador X debe ser “al menos tan importante como” Y, y éste a su vez “tan importante como” Z, etc. Golany (1988) propuso la inclusión de restricciones ordinales sobre los pesos absolutos, $w_{c,i}$, y Cherchye et al. (2007b) lo aplican al caso de los pie-shares como sigue:

$$w_{c,i1} y_{c,i1} \leq w_{c,i5} y_{c,i5} \leq w_{c,i3} y_{c,i3} \leq \dots \leq w_{c,i2} y_{c,i2}$$

Este tipo de restricciones todavía permite un alto grado de libertad al modelo en la elección de los pesos óptimos de cada unidad de análisis.

RESTRICCIONES RELATIVAS

Las restricciones relativas sobre los pie-shares consisten en asignar una cota máxima y una cota mínima al ratio entre pares de pie-shares (i, j) de los indicadores (Pedraja-Chaparro, Salinas-Jimenez y Smith, 1997):

$$\alpha_i \leq \frac{w_{c,i} y_{c,i}}{w_{c,l} y_{c,l}} \leq \beta_i$$

Este tipo de restricciones facilita a los expertos la asignación de pesos y permite realizar comparaciones de pie-shares de los indicadores por pares. Se suele aplicar cuando se tiene

información de la importancia relativa de cada pie-share con respecto al resto, p.ej. al aplicar el procedimiento de ponderación participativo de jerarquía analítica (*Analytic Hierarchy Process*). Sin embargo, al incluir este tipo de restricciones se debe tener en cuenta que se cumple la propiedad transitiva en la importancia relativa de los pie-shares de los indicadores. Ramón, Ruiz y Sirvent (2010) aplican este tipo de restricciones sobre los pesos absolutos en el modelo BOD con la combinación de las opiniones de expertos mediante el procedimiento de jerarquía analítica.

RESTRICCIONES DE PROPORCIÓN

Wong y Beasley (1990) propusieron el siguiente tipo de restricciones para facilitar a los expertos cuantificar su opinión en términos de valores de porcentajes:

$$\alpha_i \leq \frac{w_{c,i}y_{c,i}}{\sum_{i=1}^q w_{c,i}y_{c,i}} \leq \beta_i$$

Al aplicar este tipo de restricción la propiedad de invarianza en las unidades de medida se mantiene. Cherchye et al. (2007b) argumentan que las restricciones de proporción son muy útiles y atractivas cuando se consulta a un grupo de expertos sobre la contribución de cada una de las dimensiones propuestas en el estudio. Este tipo de restricciones facilita a los expertos o *stakeholders* cuantificar su opinión en términos de valores porcentuales (p.ej. utilizando la técnica de ponderación participativa Budget Allocation). Un ejemplo de la aplicación de esta técnica para la construcción de un IC puede verse en (Cherchye et al., 2005).

RESTRICCIONES DE PROPORCIÓN POR CATEGORÍAS

Este tipo de restricciones es similar a las restricciones de proporción. La única diferencia es que las restricciones de proporción por categorías se aplican dentro de cada dimensión definida en lugar de aplicarse sobre los indicadores simples que definen las diferentes dimensiones.

$$\alpha_a \leq \frac{\sum_{i \in D_a} w_{c,i}y_{c,i}}{\sum_{i=1}^q w_{c,i}y_{c,i}} \leq \beta_a$$

7.2.5. Escenarios de trabajo propuestos mediante la técnica BOD

En esta sección se han definido escenarios de trabajo alternativos al propuesto por la UPV para la evaluación académica del profesorado universitario. Dichos escenarios se han basado en la metodología BOD y se han definido según las restricciones del modelo. A continuación se presentan los resultados de los escenarios más interesantes para la UPV. En el anexo E se describen todos los diferentes escenarios propuestos que se han trabajado en esta fase de la investigación.

Escenario E1. Modelo BOD sin restricciones (IAA_E1)

En primer lugar, se ha propuesto un escenario de trabajo en el que se aplica la metodología BOD sin restricciones para la formulación del Índice de Actividad Académica del profesorado de la UPV. En este escenario se permite una máxima flexibilidad en el cálculo de los pesos asignados a cada una de las tres dimensiones, de forma que a cada profesor se le asigna un índice máximo según la actividad realizada en cada una de las dimensiones en relación con el resto del profesorado.

La formulación matemática del modelo del escenario IAA_E1 para el profesor c es la siguiente:

$$\begin{aligned}
 IAA_E1_c^* &= \max_{(w_{c,1}, \dots, w_{c,3})} \sum_{i=1}^3 w_{c,i} I_{c,i} \\
 & \quad s. a \\
 \sum_{i=1}^3 w_{c,i} I_{j,i} &\leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020 \\
 w_{c,i} &\geq 0 \quad i = 1, \dots, 3
 \end{aligned}$$

$IAA_E1_c^*$: puntuación máxima del IAA para el profesor c con respecto al resto de profesores

$w_{c,i}$: pesos asignados por el modelo BOD para el profesor c

$I_{c,1}$: puntuación de la dimensión de docencia IAD para el profesor c

$I_{c,2}$: puntuación de la dimensión de investigación VAIP para el profesor c

$I_{c,3}$: puntuación de la dimensión de gestión AG para el profesor c

La principal ventaja que presenta este modelo es que suele ser aceptado por las partes evaluadas, en este caso el profesorado, puesto que asigna un peso óptimo a cada una de las dimensiones según los distintos perfiles de profesorado; docente, investigador y/o gestor con respecto al resto de profesores.

En la figura 7.16 se comparan las puntuaciones del IAA obtenidas en el escenario E1 y las puntuaciones obtenidas con el modelo propuesto por la UPV.

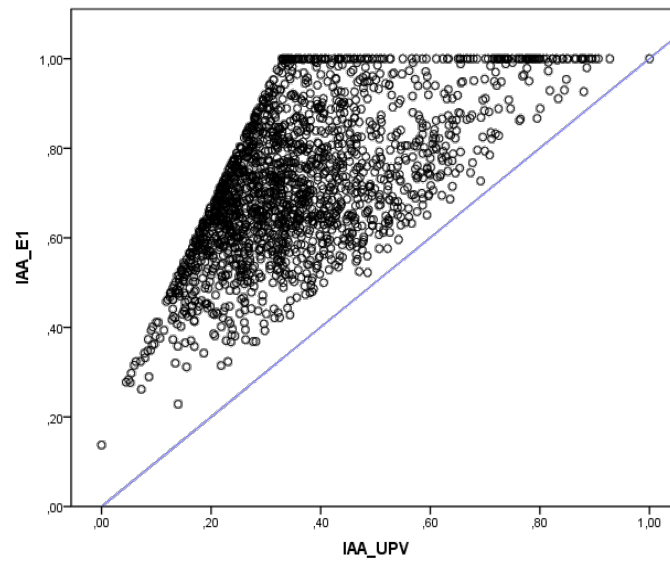


Figura 7.16. Comparación del IAA_E1 y IAA_UPV

Se observa que el IAA formulado en el escenario E1 asigna una puntuación académica mayor que el modelo propuesto por la UPV para todo el profesorado. Por tanto, el uso de pesos adaptables según el perfil de cada profesor hace que el modelo sea más aceptable y menos criticable por parte del profesorado, ya que cualquier otro conjunto de pesos distinto genera un IAA menor para cada profesor.

Sin embargo, la principal limitación del modelo libre de restricciones es que puede conducir a resultados inapropiados y difíciles de justificar desde el punto de vista teórico. En este caso el modelo asigna una puntuación máxima a aquellos profesores que alcanzan el valor máximo en al menos una de las dimensiones y cero en el resto. Por tanto, el modelo calcula el IAA_E1 teniendo en cuenta una dimensión e ignorando el resto. En la tabla 7.4 se muestra un claro ejemplo:

	$I_{c,1}$	$I_{c,2}$	$I_{c,3}$	$w_{c,1}^*$	$w_{c,2}^*$	$w_{c,3}^*$	$\%w_{c,1}^*I_{c,1}$	$\%w_{c,2}^*I_{c,2}$	$\%w_{c,3}^*I_{c,3}$	$IAA_{E1}_c^*$
Profesor A	68,500	43,510	20,000	0,000	0,000	0,050	0,000%	0,000%	100,000%	1,000
Profesor B	68,500	0,170	5,580	0,015	0,000	0,000	100,000%	0,000%	0,000%	1,000
Profesor C	45,350	55,860	20,000	0,000	0,001	0,048	0,000%	3,164%	96,836%	1,000
Profesor D	29,960	208,000	17,930	0,000	0,005	0,000	0,000%	100,000%	0,000%	1,000
Profesor E	48,080	208,000	17,910	0,000	0,005	0,000	0,000%	100,000%	0,000%	1,000
Profesor F	26,020	208,000	8,320	0,000	0,005	0,000	0,000%	100,000%	0,000%	1,000

Tabla 7.4. Resultados del modelo BOD para un conjunto de profesores

La expresión $\%w_{c,1}^*I_{c,1}$ representa el concepto de *contribución* descrito anteriormente, es decir, la contribución del indicador que en este caso es $I_{c,1}$ a la formación del índice $IAA_{E1}_c^*$ y se calcula como sigue:

$$\%w_{c,1}^*I_{c,1} = \left(\frac{w_{c,1}^* * I_{c,1}}{IAA_{E1}_c^*} \right) * 100$$

En la tabla 7.4 se exponen los resultados de seis profesores. Para cada profesor c ($c = A, B, C, D, E, F$) se muestra: la puntuación de cada dimensión ($I_{c,1}, I_{c,2}, I_{c,3}$), el conjunto de pesos óptimo asignado por el modelo a cada dimensión una vez comparado con el resto de profesores ($w_{c,1}^*, w_{c,2}^*, w_{c,3}^*$), la contribución o porción de cada dimensión en la formación del IAA ($\%w_{c,1}^*I_{c,1}, \%w_{c,2}^*I_{c,2}, \%w_{c,3}^*I_{c,3}$) y la puntuación final asignada por el modelo a cada profesor ($IAA_E1_c^*$).

Se observa que la puntuación del indicador compuesto para todos los profesores es la misma, la unidad, siendo ésta la puntuación máxima que puede alcanzar un profesor según el modelo planteado. Este resultado indica que no hay ningún otro profesor que con el conjunto de pesos óptimo del profesor evaluado tenga una eficacia relativa¹⁶ mayor que la del profesor evaluado.

Así pues, el modelo asigna un peso muy elevado a aquellas dimensiones en las que el profesor tiene una mayor puntuación y asigna un peso nulo o prácticamente cero a las dimensiones con una baja actividad. Por tanto, el modelo BOD sin restricciones se centra en aquella dimensión que tiene una puntuación elevada, asignándole un 100% en su contribución y 0% al resto. Así, el modelo ignora una o más dimensiones del indicador compuesto y sobreestima otra u otras dimensiones en las que el profesorado tiene una alta actividad. Esto conlleva a que el uso del modelo BOD sin restricciones corra el riesgo de construir el IAA teniendo en cuenta y premiando sólo una parte de las dimensiones definidas para su evaluación e ignorando y no penalizando la baja actividad en el resto de dimensiones. Como consecuencia, el profesorado puede focalizar su esfuerzo en una de las dimensiones para salir favorecido en la puntuación del IAA, ignorando el resto de dimensiones consideradas también importantes para medir la actividad académica del profesorado. El siguiente histograma muestra que el escenario E1 asigna una puntuación máxima a muchos de los profesores que tienen una alta puntuación en alguna de las dimensiones.

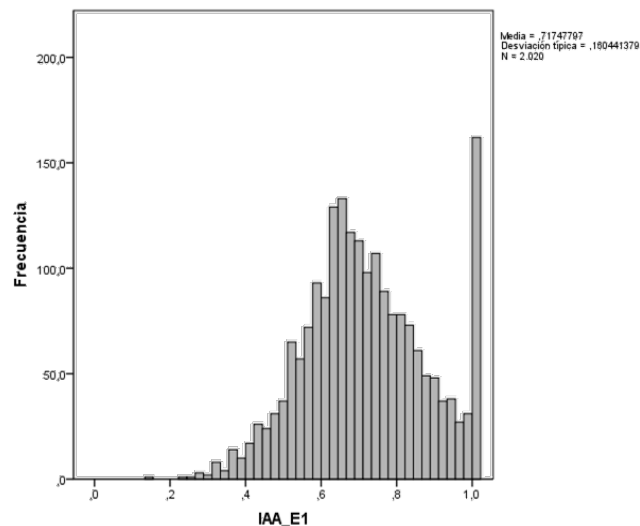


Figura 7.17. Histograma del escenario IAA_E1

¹⁶Nótese que en este caso siempre se habla de eficacia relativa en términos de actividad académica realizada por un profesor con respecto a la actividad académica realizada por el resto del profesorado

Además, el modelo asigna una misma eficacia relativa a profesores en los que se evidencia una gran diferencia en su actividad académica. Véase por ejemplo el profesor E y F en la tabla 7.4. Estos dos profesores tienen una misma actividad de investigación, sin embargo, el profesor E tiene mucha más actividad de docencia y de gestión. En este caso el escenario E1 asigna un 100% a la dimensión de investigación ignorando el resto de dimensiones y coloca a ambos profesores en un mismo nivel de eficacia relativa. Sin embargo, se observa claramente que el profesor E debería tener una puntuación más alta que el profesor F.

Por tanto, a pesar de las ventajas que ofrece el modelo BOD sin restricciones, la principal limitación mencionada anteriormente hace que el uso del escenario E1 no sea adecuado para medir la actividad académica del profesorado de la UPV. Esta limitación se puede corregir incluyendo un conjunto de restricciones expuestas a continuación en el escenario E2.

Escenario E2. Modelo BOD con restricciones de tipo I (IAA_E2)

El escenario E2 introduce una modificación al escenario E1 que permite limitar la flexibilidad en los pesos asignados en el modelo BOD sin restricciones. Los objetivos de esa limitación son dos: i) tener en cuenta todas las dimensiones definidas en el marco conceptual teórico para la evaluación de la actividad académica del profesorado de la UPV e ii) introducir opiniones generales de un conjunto de expertos sobre la ponderación adecuada de las dimensiones. Una forma de limitar esta libertad en la elección de los pesos del escenario E1 es el uso de restricciones en los pesos basadas en opiniones de expertos o tomadores de decisiones.

El uso de restricciones en el modelo resulta muy interesante en el caso en que no se llegue a un consenso sobre la importancia que debe tener cada una de las dimensiones de estudio. En este caso resulta muy difícil asignar una importancia fija e igual a todo el profesorado para cada una de las dimensiones que miden la actividad académica, puesto que existen distintos perfiles de profesorado más enfocados a unas dimensiones u otras. Es por eso que se ha empleado la técnica de ponderación participativa conocida como *Budget Allocation* (BA) para consultar a un grupo de expertos sobre la contribución que debe tener cada dimensión evaluada en la formación del IAA del profesorado.

El grupo de expertos está formado por 13 miembros de la UPV concedores de los indicadores definidos en la universidad y del funcionamiento de las actividades del profesorado de la UPV. La técnica de ponderación BA se ha basado en pedir a cada experto que distribuya 100 puntos entre las dimensiones de estudio, según su opinión sobre la contribución de éstas a la formación del IAA_E2. Los resultados se muestran en la figura 7.19. Se trata de una de las técnicas más simples y utilizadas para consultar opinión a un grupo de expertos. En este caso, cada experto ha proporcionado un gráfico de sectores como el que sigue:

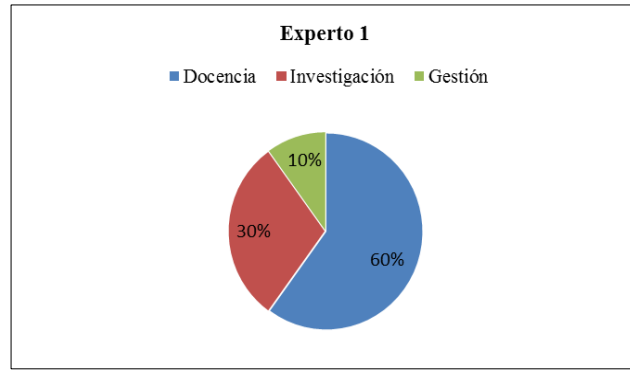


Figura 7.18. Contribución de cada dimensión al IAA_E2 según el experto 1

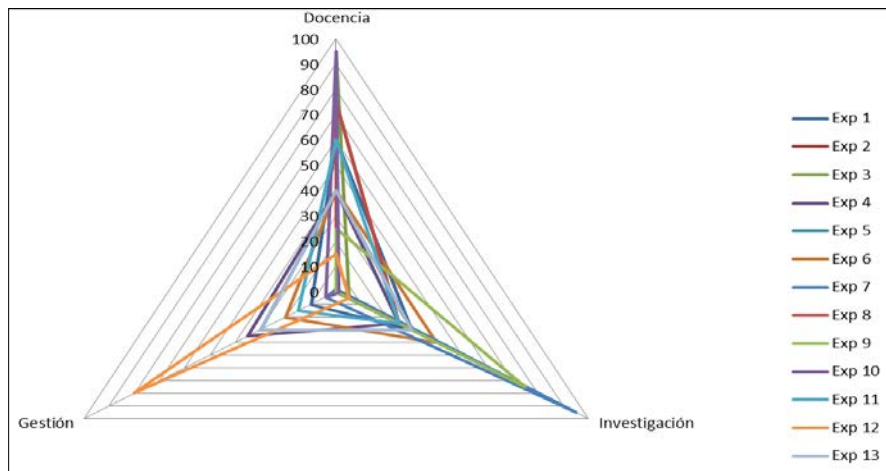


Figura 7.19. Contribución de cada dimensión según la opinión del grupo de expertos

Además, para construir el escenario E2 se propuso a los expertos que llegaran a un consenso sobre la variación permitida en la contribución de cada una de las dimensiones. En este caso, les resultó más fácil llegar a un consenso sobre la variación que puede tener cada dimensión que asignar un único valor a cada una de ellas, puesto que había opiniones distintas sobre las contribuciones de la docencia, investigación y gestión al IAA. A este conjunto de restricciones se les ha nombrado como restricciones de tipo I, siendo la formulación matemática del modelo BOD con estas restricciones la siguiente:

$$IAA_E2_c^* = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^3 w_{c,i} I_{c,i}$$

s. a

$$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020$$

$$0,25 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,75$$

$$0,25 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,75$$

$$0 \leq \frac{w_{c,3}I_{c,3}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}y_{c,3}} \leq 0,5$$

$$w_{c,i} \geq 0 \quad i = 1, \dots, 3$$

$IAA_E2_c^*$: puntuación máxima del IAA para el profesor c con respecto al resto de profesores

$w_{c,i}$: pesos asignados por el modelo BOD para el profesor c

$I_{c,1}$: puntuación de la dimensión de docencia IAD para el profesor c

$I_{c,2}$: puntuación de la dimensión de investigación VAIP para el profesor c

$I_{c,3}$: puntuación de la dimensión de gestión G para el profesor c

Nótese que se han incorporado al modelo las restricciones sobre las contribuciones de cada dimensión ($\frac{w_{c,i}I_{c,i}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}y_{c,3}}$), conocidas como restricciones de proporción y descritas en el apartado 7.2.4., ya que son las restricciones más sencillas de interpretar por el grupo de expertos. Resulta más fácil que los expertos proporcionen información sobre la contribución de cada una de las dimensiones que sobre el peso que debe tener cada una de éstas. Por tanto, en este caso se ha trabajado sobre lo que se conoce como pie-shares, es decir sobre las porciones o contribuciones de cada una de las dimensiones a la formación del índice. Tal y como se ha explicado en el apartado anterior, la ventaja que presenta trabajar sobre los pie-shares es que no es necesario realizar una normalización previa sobre las dimensiones, puesto que se cumple la propiedad de invarianza en la unidad de medida.

La tabla 7.1 de estadísticos descriptivos muestra que las dimensiones de investigación y gestión pueden ser nulas para algunos profesores, puesto que el valor mínimo para estas dimensiones es cero. Sin embargo, la dimensión de docencia no puede ser nula, ya que cualquier profesor con contrato en la universidad debe tener asignado un mínimo de créditos docentes. En el caso de la dimensión de gestión, si algún profesor toma el valor cero la contribución de esta dimensión sería nula, ya que se trata de un posible valor aceptable según las restricciones definidas en el modelo IAA_E2 por el grupo de expertos. Sin embargo, si un profesor presenta una puntuación nula en la dimensión de investigación el valor $I_{c,2}$ sería cero y, por tanto, la restricción

$$0,25 \leq \frac{w_{c,2}I_{c,2}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}y_{c,3}} \leq 0,75$$

sería inconsistente. Para resolver esta inconsistencia se ha realizado una transformación que no altera la restricción pero que asigna una contribución nula a la dimensión de investigación en el caso en que la puntuación para $I_{c,2}$ sea cero. Esta corrección se ha basado en multiplicar la restricción por el factor $I_{c,2}$. Aun así, si las contribuciones de las dimensiones de investigación y gestión son cero, el máximo asignado a la docencia no puede ser 100% ya que su restricción está acotada a un 75%. Por tanto, el modelo es inconsistente para aquellos profesores que presentan puntuaciones nulas en estas dos dimensiones.

Para solucionar esta limitación se ha calculado el IAA de este conjunto de profesores, con puntuaciones nulas en investigación y gestión, previamente a la aplicación del modelo E2 para el resto de profesores. Para ello, se ha aplicado la siguiente ecuación:

$$IAA_E2_k^* = \max_{w_{k,1}} \frac{w_{k,1} I_{k,1}}{\max_{(I_{1,1}, \dots, I_{j,1}, \dots, I_{n,1})} w_{k,1} I_{j,1}} * 0,75 = \frac{I_{k,1}}{\max_{(I_{1,1}, \dots, I_{j,1}, \dots, I_{n,1})} I_{j,1}} * 0,75 \quad (7.3)$$

donde $k \in A$ y A : conjunto de profesores con una puntuación nula en las dimensiones de investigación y gestión.

En este caso, la puntuación IAA_E2 para cada profesor k del conjunto A se ha definido como el cociente entre la puntuación en la dimensión docencia y la puntuación máxima en dicha dimensión teniendo en cuenta todo el profesorado. Además, dicho cociente se ha penalizado por un factor para que la dimensión docencia pese como máximo un 75%.

Sin embargo, una vez realizados todos los cálculos se observa que algunos de los resultados no se pueden justificar a nivel teórico. En la tabla 7.5 se muestra un claro ejemplo de ello:

	$I_{c,1}$	$I_{c,2}$	$I_{c,3}$	$w_{c,1}^*$	$w_{c,2}^*$	$w_{c,3}^*$	$\%w_{c,1}^* I_{c,1}$	$\%w_{c,2}^* I_{c,2}$	$\%w_{c,3}^* I_{c,3}$	$IAA_E2_c^*$
Profesor X	19,020	0,000	0,000	0,011	0	0	75%	0%	0%	0,208
Profesor Y	19,420	0,000	0,170	0,001	0,000	0,045	75%	0%	25%	0,031

Tabla 7.5. Resultados de las puntuaciones de los profesores X, Y en el modelo del escenario E2

En este caso particular el IAA_E2 correspondiente al profesor X se ha calculado mediante la fórmula (7.3), mientras que la puntuación del indicador compuesto para el profesor Y se ha calculado con el modelo del escenario E2. Nótese que a pesar de que el profesor Y realiza más actividad de docencia y de gestión su puntuación en el IAA_E2 es mucho menor, resultado que no se puede justificar a nivel teórico.

En conclusión, este escenario aporta una gran ventaja como es la introducción de restricciones sobre las contribuciones de las dimensiones basadas en opiniones de expertos, sin embargo, el modelo es inconsistente frente a situaciones en las que el profesorado no realiza actividades en dos de las dimensiones, en este caso investigación y gestión. La presencia conjunta de varias puntuaciones nulas en los indicadores es un tema de investigación en el que se continúa trabajando.

Escenario E3. Modelo BOD con restricciones de tipo I y acotación en la dimensión Investigación (IAA_E3)

El escenario E3 se ha planteado con el mismo modelo funcional que el escenario E2 para el cálculo del Índice de Actividad Académica. Por tanto, se basa en la metodología BOD con restricciones de tipo I consensuadas por el grupo de expertos.

$$IAA_E3_c^* = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^3 w_{c,i} I_{c,i}$$

s. a

$$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020$$

$$0,25 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,75$$

$$0,25 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,75$$

$$0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,5$$

$$w_{c,i} \geq 0 \quad i = 1, \dots, 3$$

$IAA_E3_c^*$: puntuación máxima del IAA para el profesor c con respecto al resto de profesores

$w_{c,i}$: pesos asignados por el modelo BOD para el profesor c

$I_{c,1}$: puntuación de la dimensión de docencia IAD para el profesor c

$I_{c,2}$: puntuación de la dimensión de investigación VAIP para el profesor c

$I_{c,3}$: puntuación de la dimensión de gestión G para el profesor c

Tal y como muestran los resultados del escenario E2, el modelo IAA_E2 ordena al profesorado evaluado de forma no justificable desde el punto de vista teórico debido a la presencia de profesores con puntuaciones nulas en las dimensiones de investigación y de gestión. Como solución a este problema, en el escenario E3 se ha introducido una cota inferior para todo el profesorado en la dimensión de investigación. La incorporación de esta cota se ha consultado previamente con el grupo de expertos, los cuales han concluido que parece razonable asignar un valor prefijado de 5 puntos a aquellos profesores que no tienen actividad de investigación en la dimensión VAIP o tienen una puntuación muy baja en ésta, ya que cualquier docente ha realizado durante su carrera académica alguna actividad de investigación, aunque ésta no se haya recogido en el índice de investigación (VAIP) definido por la UPV. La presencia de ceros en la base de datos es una limitación importante en la metodología BOD. Existen diversos estudios que tratan esta problemática e incorporan posibles modificaciones en el modelo, sin embargo, muchos otros optan por sustituir los ceros por una cota inferior justificada (Byrnes et al., 1984; Ferrier y Lovell, 1990; Thompson, Dharmapala y Thrall, 1993).

Por otra parte, la puntuación en la dimensión de gestión se ha mantenido igual que en el escenario E2, puesto que la gestión puede presentar una contribución de cero según el grupo de expertos (la cota inferior en la restricción de la dimensión de gestión puede tomar el valor cero). En este caso el profesorado no está obligado a realizar tareas de gestión y puede que haya profesores que en ningún momento de su carrera académica hayan realizado actividades de esta dimensión.

Al comparar la distribución del escenario E3 con la del escenario E1 se observa una gran diferencia. Mientras que el escenario E1 asigna la máxima puntuación a la mayoría del profesorado, el escenario E3 resalta muy bien aquellos profesores que son excelentes en todo.

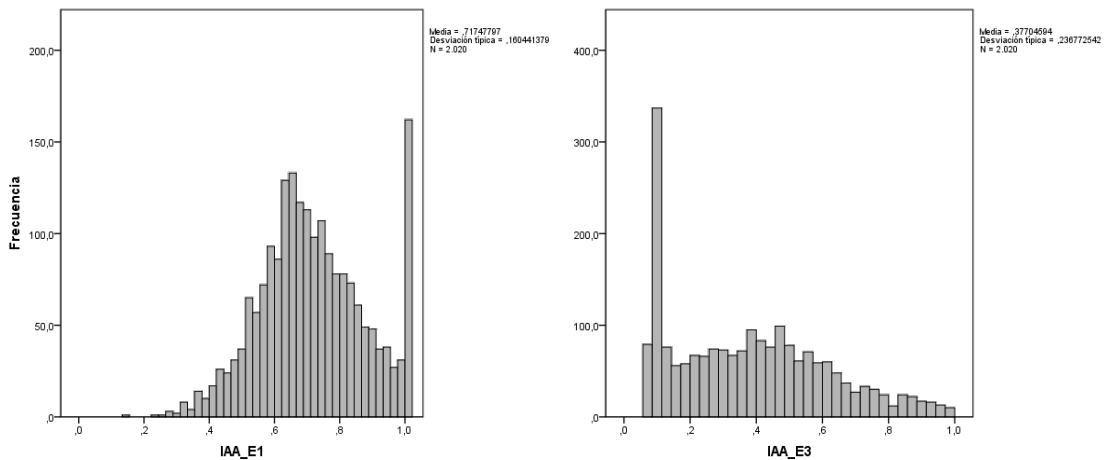


Figura 7.20. Histogramas de los escenarios E1 y E3

En la figura 7.21 se comparan los resultados del escenario E3 con el escenario de la UPV.

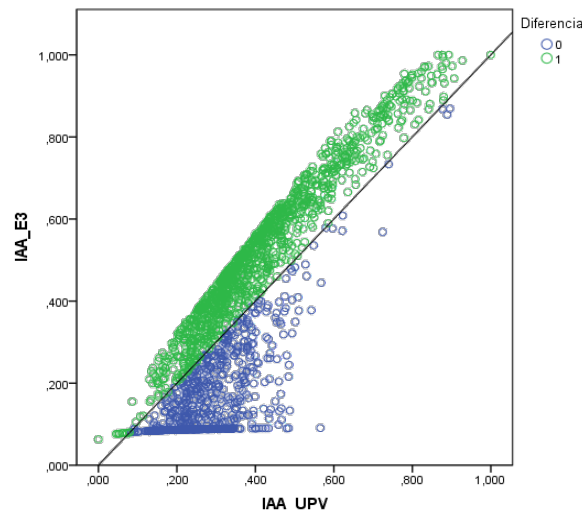


Figura 7.21. Comparación del IAA_E3 y IAA_UPV

Se observa que el escenario E3 asigna una puntuación IAA mayor que el escenario de la universidad para la mayoría de profesores (valores representados en color verde). Sin embargo, existe un grupo de profesores para los que la puntuación del IAA es mayor en el escenario de la

UPV (valores representados en color azul). Se trata de profesores a los que el modelo IAA_UPV les ha asignado una contribución en las dimensiones que no cumple las restricciones definidas en el escenario E3. En la tabla 7.6 se describen dos situaciones posibles que muestran que los escenarios no son comparables:

	Escenario UPV			Escenario UPV			Comparación escenarios		
	$I_{c,1}$	$I_{c,2}$	$I_{c,3}$	$\%w_{c,1}^*I_{c,1}$	$\%w_{c,2}^*I_{c,2}$	$\%w_{c,2}^*I_{c,2}$	$IAA_{UPV_c}^*$	$IAA_{UPV}_{[0,1]_c}^*$	$IAA_{E3_c}^*$
Profesor X	68,50	11,29	0	93,16%	6,84%	0,00%	114,31	0,36	0,19
Profesor Y	68,50	15,18	8,25	71,69%	7,08%	21,24%	148,55	0,49	0,24

Tabla 7.6. Profesores X, Y con un IAA_UPV mayor que IAA_E3

Los profesores X e Y presentan una puntuación para el IAA_UPV mayor que en el escenario E3 ($0,36 > 0,19$ y $0,49 > 0,24$, siendo 0,36 y 0,49 el valor para IAA_UPV normalizado en [0,1], respectivamente para el profesor X e Y). Sin embargo, las contribuciones de las dimensiones no cumplen las restricciones del escenario E3. En el caso del profesor X la contribución de la dimensión de docencia es del 93,16%, siendo la cota máxima para esta dimensión un 75%. Por otra parte, el profesor Y presenta una contribución de la dimensión de investigación del 7,08%, valor que no alcanza la cota mínima para esta dimensión (25%).

De esta forma, los profesores en el modelo IAA_UPV que presentan contribuciones de las dimensiones que incumplen las restricciones del modelo E3 no se pueden comparar con los resultados del escenario E3. Éstos corresponden a todos los profesores marcados con un cuadro rojo en las figuras 7.22 y 7.23:

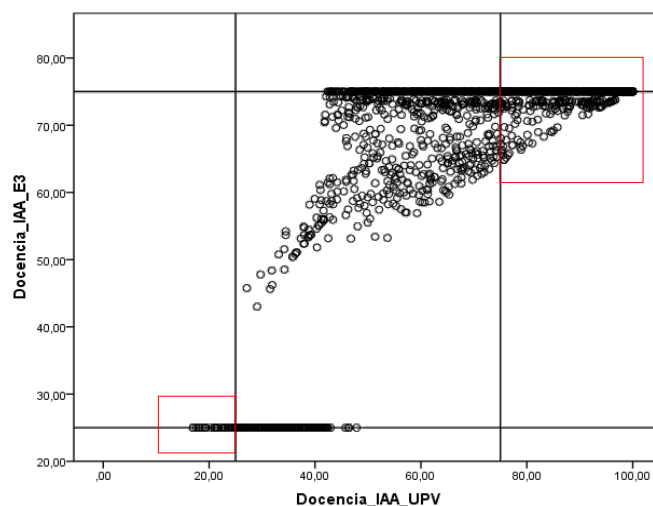


Figura 7.22. Comparación del IAA_E3 y IAA_UPV según la contribución de la docencia

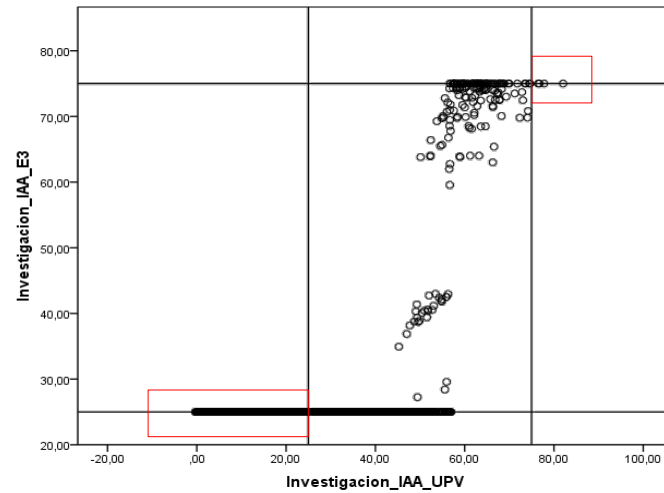


Figura 7.23. Comparación del IAA_E3 y IAA_UPV según la contribución de la investigación

En la figura 7.24 se comparan los rankings resultantes del índice IAA para los escenarios E3 y UPV.

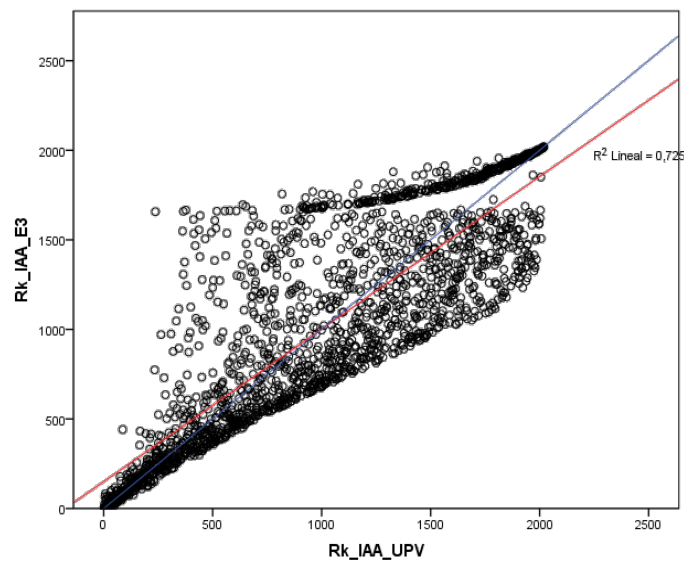


Figura 7.24. Comparación en posiciones de los escenario E3 y UPV

Obsérvese que a pesar de existir una alta correlación entre los rankings de los profesores en ambos escenarios, existen muchas diferencias en las posiciones del profesorado, ya que muchos de ellos se alejan de la recta azul que es la que iguala las posiciones en ambos escenarios.

Análisis descriptivo del escenario E3

Para estudiar el comportamiento del escenario E3 se ha realizado un análisis descriptivo obteniéndose los resultados que se muestran a continuación.

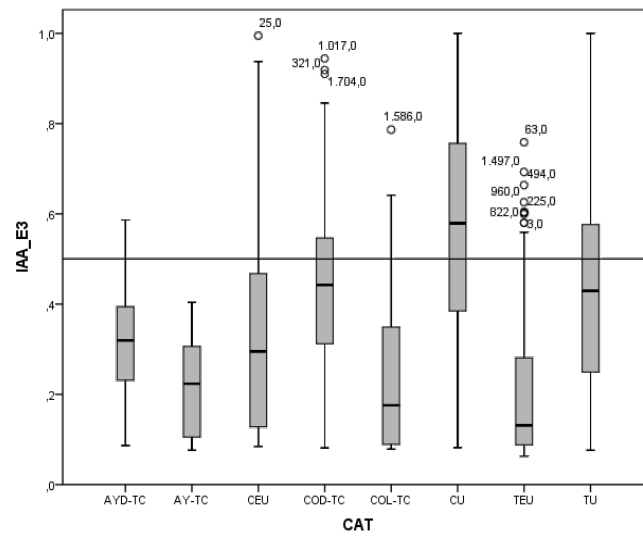


Figura 7.25. IAA_E3 por categoría laboral

Cabe destacar que la media de la puntuación del índice IAA_E3 es inferior a 0,5 para todas las categorías laborales excepto para los Catedráticos de Universidad. En particular, el escenario E3 asigna a los Titulares de Escuela Universitaria la menor puntuación, en promedio, de actividad académica. Asimismo, se observa que existe mucha variabilidad entre las puntuaciones del IAA_E3 para el profesorado CEU, TU y CU.

Al analizar más detalladamente el comportamiento del escenario E3 para cada una de las dimensiones en la figura 7.26 se concluye que la dimensión de docencia contribuye en gran medida a la formación del IAA_E3 para la mayoría de los profesores. Destacan algunos profesores CU, TU y CEU por presentar una puntuación para el IAA_E3 formada en mayor medida por la investigación y la gestión.

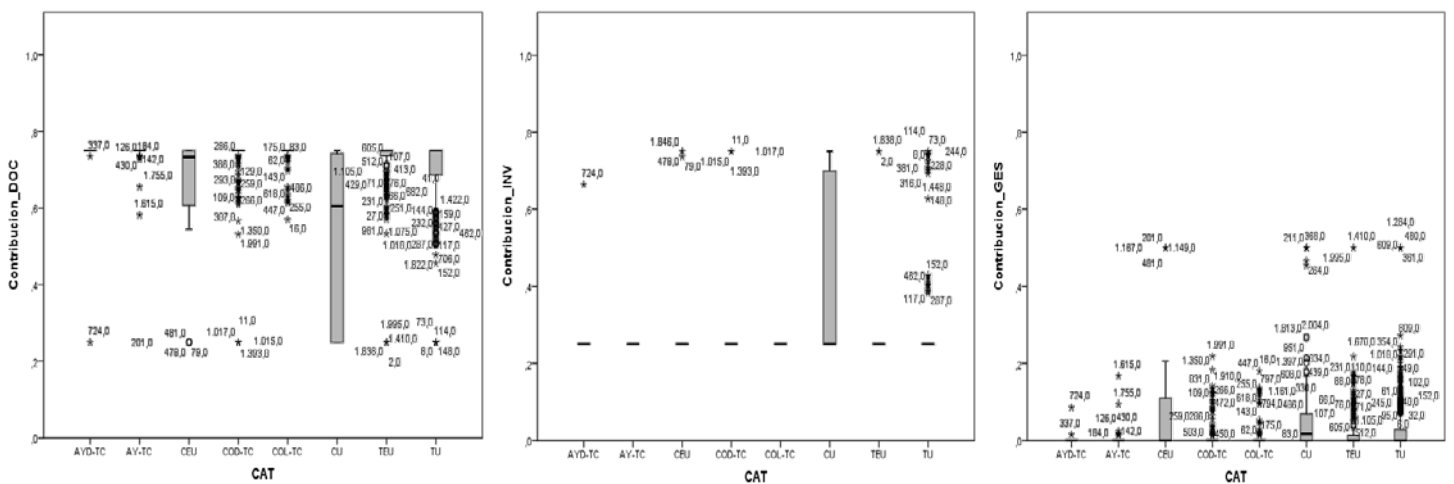


Figura 7.26. Contribución de las dimensiones al IAA_E3 por categoría laboral

Categorización de los perfiles de profesorado en coordenadas baricéntricas

De forma análoga al escenario IAA propuesto por la UPV se han representado las contribuciones de las dimensiones de docencia, investigación y gestión mediante coordenadas baricéntricas con el objetivo de categorizar los diferentes perfiles de profesorado.

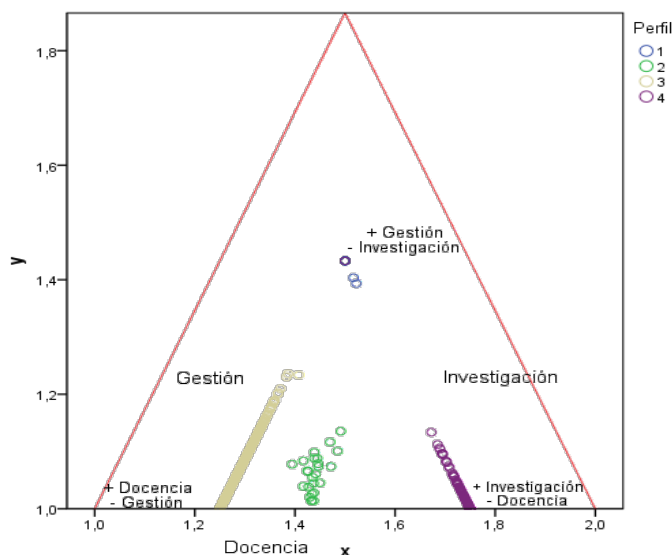


Figura 7.27. Contribuciones de las dimensiones al IAA_E3 en coordenadas baricéntricas

Como puede apreciarse en la figura 7.27, el escenario E3 agrupa a todo el profesorado en cuatro perfiles diferentes dependiendo de sus contribuciones de la docencia, investigación y gestión a la formación del índice con respecto al resto del profesorado. Los grupos se han identificado teniendo en cuenta las proyecciones de los puntos en los ejes correspondientes.

El perfil 1 se define como el *perfil Gestor (G)* y está formado por un grupo de profesores que presenta un IAA con una elevada contribución de la gestión. Este perfil lo forman el 1,5 % del total de profesores evaluados. El perfil 2 está compuesto por profesores en los que la dimensión docente e investigadora contribuye en gran medida a la formación de su índice, por lo que se podría definir como un *perfil Docente-Investigador (D+I)* y representan el 1,2% del total. Por otra parte, el grupo de profesores que presenta un IAA_E3 en el que contribuye mayoritariamente la docencia forman el perfil 3 o *perfil Docente (D)*. El perfil Docente está compuesto por la mayoría de profesores evaluados, 89,1%. Finalmente, el perfil 4 está compuesto por profesores en los que la investigación tiene una alta contribución a la formación de su IAA, por lo que forman el *perfil Investigador (I)* y constituyen el 8,1% del total del profesorado.

Las restricciones del modelo imponen que las dimensiones de docencia, investigación y gestión no contribuyan más de un cierto máximo establecido a la formación del IAA_E3. Esto implica que si, por ejemplo, un profesor tiene muy buena puntuación en el pie-share correspondiente a la dimensión de gestión, el escenario E3 le asigna una contribución de su

actividad gestora a la formación de su índice académico máxima de un 50%, asignando un 50% a las contribuciones de investigación y docencia. De la misma forma, si un profesor obtiene un IAA con una alta contribución de la docencia, el modelo asigna un 75% como máximo a la contribución de docencia y un 25% como mínimo a la contribución de la dimensión de investigación, puesto que el modelo exige un mínimo de contribución de investigación en la formación del índice.

CAT	% profesores de la CAT respecto al total	G	D+I	D	I
AY-TC	2,2%			¹ 100,0% ² 22,5%	
AYD-TC	1,2%			95,7% 12,2%	4,3% 0,6%
COL-TC	7,1%			100,0% 8,0%	
COD-TC	11,8%			98,3% 12,9%	2,5% 3,7%
TEU	21,2%	0,5% 6,4%		99,1% 23,6%	0,5% 1,2%
TU	38,7%	0,9% 22,5%	1,7% 52,0%	91,4% 38,7%	6,0% 28,7%
CEU	2,4%	8,5% 12,8%		87,2% 2,3%	6,4% 1,8%
CU	15,4%	5,8% 58,0%	3,9% 48,0%	56,6% 9,8%	6,0% 28,7%
TOTAL	100%	1,5%	1,2%	89,1%	8,1%

¹ Porcentaje de profesores de la categoría laboral que están en el perfil correspondiente con respecto al número total de profesores de esa categoría laboral

² Porcentaje de profesores de la categoría laboral que están en el perfil correspondiente con respecto al número total de profesores en dicho perfil

Tabla 7.7. Asociación de categorías laborales a los perfiles definidos en el escenario E3

En la tabla 7.7 se describe cada uno de los perfiles obtenidos en el modelo E3. Cabe destacar que el perfil Docente está compuesto por profesores de todas las categorías laborales, pero principalmente está formado por profesores TU, TEU y AY. Además, todo el profesorado AY y COL pertenece a este perfil. Por otra parte, el perfil Gestor y Docente-Investigador está formado por profesores TEU, TU, CEU y CU en el primer caso y TU y CU en el segundo. Finalmente, el perfil Investigador está compuesto en su mayoría por profesores TU y CU aunque también se pueden encontrar algunos profesores AYD, COD, TEU y CEU que presentan altas contribuciones de la investigación a la formación de sus índices.

El modelo IAA_E3 asigna un mayor valor promedio en la actividad académica al perfil Investigador y Docente-Investigador, siendo el perfil Docente el que posee en promedio una menor puntuación en el índice IAA_E3 (véase figura 7.28). No obstante, cabe destacar que este resultado no significa que no haya profesores con perfil docente o gestor que sean muy eficaces¹⁷ en su actividad académica con respecto al resto como se verá más adelante.

¹⁷ Recuérdese que en este contexto un profesor *c* se dice que es *eficaz* cuando obtiene un IAA próximo a 1. Esto significa que existen muy pocos profesores (o incluso ninguno) que con los pesos del profesor *c* obtienen un IAA mayor al del profesor *c*

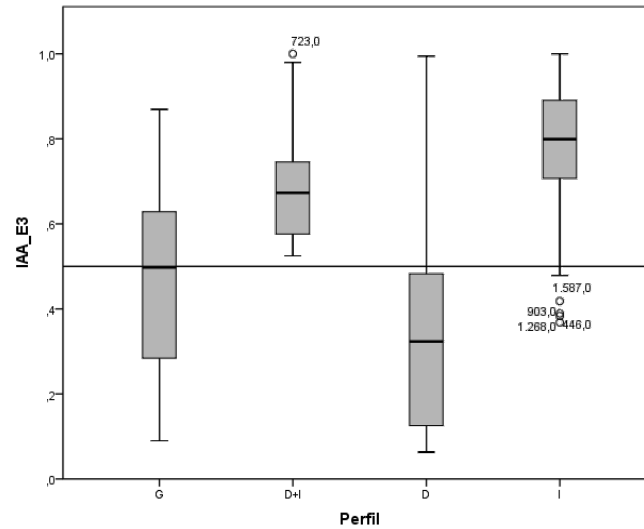


Figura 7.28. IAA_E3 por perfiles de profesorado

Los resultados obtenidos mediante el modelo E3 son totalmente coherentes con las características de las diferentes figuras de profesorado en la universidad. Asimismo, el modelo E3 asigna un conjunto de pesos a cada dimensión, y de forma específica a cada profesor, que tiene en cuenta el desarrollo de cada una de las dimensiones con respecto al resto de profesores y refleja de forma correcta lo que cada profesor realiza según su categoría laboral.

Cabe destacar que si dos profesores presentan una misma contribución de las dimensiones a su IAA_E3 no significa que sean igual de eficaces con respecto al resto de profesores. Es por eso que, a continuación, se presenta un estudio sobre la variación del índice de actividad académica para los distintos perfiles. Este estudio se ha realizado de forma general y de forma particular para cada uno de ellos.

PERFIL GESTOR

El perfil Gestor está formado únicamente por un 1,5% del total de profesores evaluados. Se trata de un perfil de profesorado cuya contribución de la gestión a su actividad académica supone aproximadamente un 50%, la contribución de la docencia un 25% y la contribución de la investigación a su actividad un 25%. Las contribuciones de los profesores que pertenecen a este perfil son muy similares, sin embargo, las puntuaciones que obtienen en las dimensiones de docencia, investigación y gestión hacen que sus IAA_E3 sean diferentes.

Si se analizan las distintas categorías de profesorado laboral que pertenecen al perfil Gestor, la figura 7.29 muestra que, en promedio, los profesores CU, CEU y TU son los que obtienen una puntuación superior para el IAA_E3, mientras que los profesores Titulares de Escuela Universitaria obtienen una baja valoración de su actividad académica.

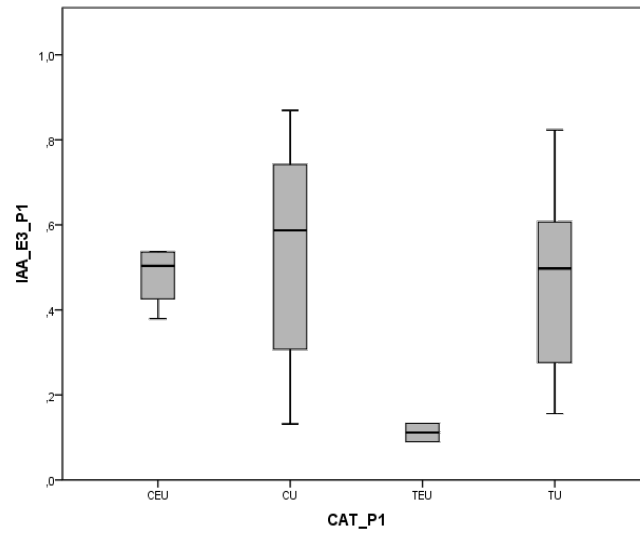


Figura 7.29. IAA_E3 por categoría laboral para el perfil Gestor

Analizando las puntuaciones de este grupo de profesores en las dimensiones de docencia, investigación y gestión, la figura 7.30 revela que el profesorado con perfil Gestor suele tener una puntuación cercana a los valores medios de la UPV en docencia, una puntuación menor a los valores medios de la UPV en investigación, exceptuando a un grupo de profesores CU en esta última dimensión, y una puntuación superior a los valores medios de la UPV en gestión.

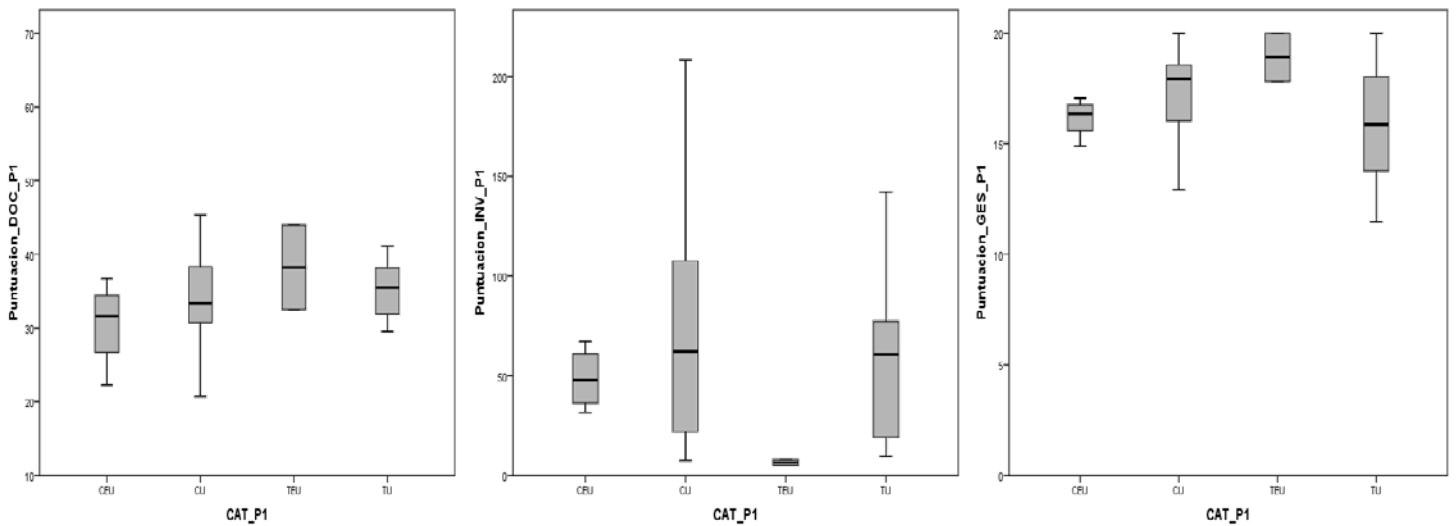


Figura 7.30. Puntuaciones de las dimensiones para el perfil Gestor en el escenario E3

PERFIL DOCENTE-INVESTIGADOR

En el caso del perfil Docente-Investigador, formado por un 1,2% del profesorado evaluado, tanto el profesorado TU como CU, que son los que forman este perfil, tienen una puntuación superior a 0,5 y próxima a la unidad. La figura 7.31 muestra claramente que dos profesores Catedráticos de Universidad que pertenecen a este perfil son muy eficaces en el desarrollo de su actividad académica.

Este perfil está formado por profesores con una contribución de la docencia a su actividad académica de aproximadamente un 50%, una contribución de la investigación de un 40% y una contribución de las actividades de gestión de un 10% aproximadamente.

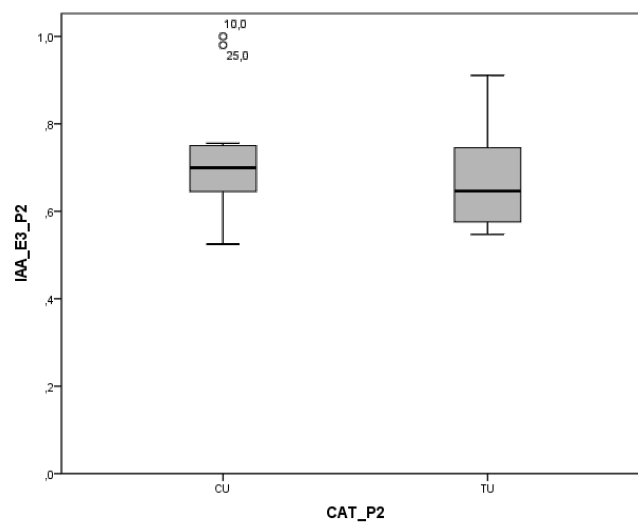


Figura 7.31. IAA_E3 por categoría laboral para el perfil Docente-Investigador

En la figura 7.32 se observa que los profesores que pertenecen al perfil Docente-Investigador tienen puntuaciones en las dimensiones de docencia e investigación cercanas a los valores medios de la UPV y una baja puntuación en las actividades de gestión.

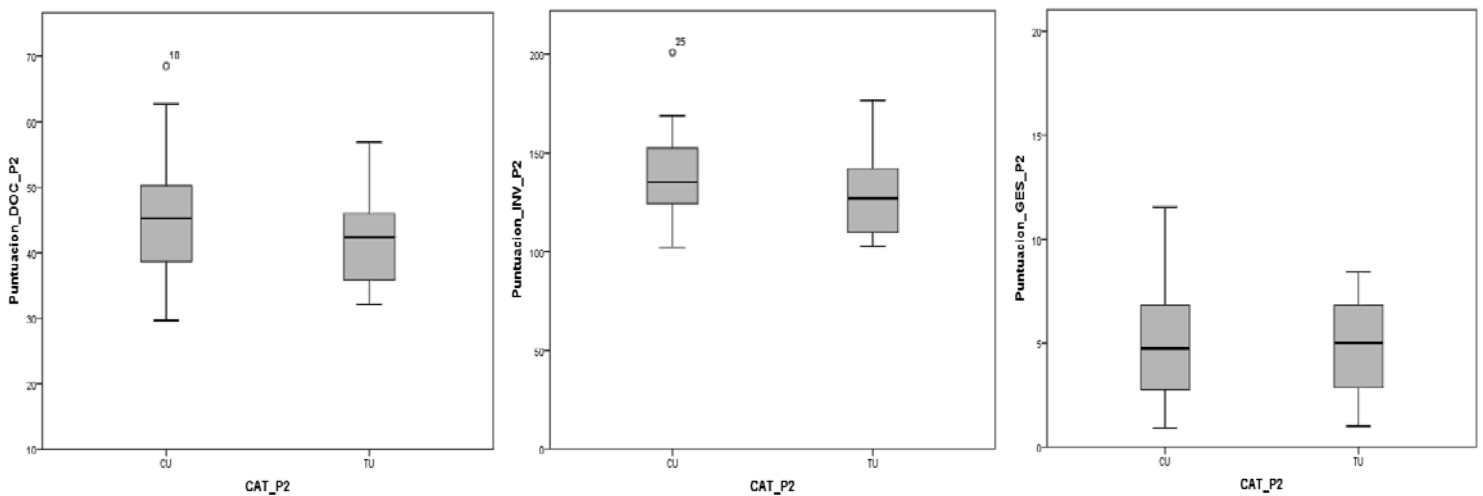


Figura 7.32. Puntuaciones de las dimensiones para el perfil Docente-Investigador en el escenario E3

PERFIL DOCENTE

El perfil Docente está formado por la mayoría del profesorado (89,1%) que pertenece a las distintas categorías laborales. Está compuesto por profesores que presenta un IAA_E3 con una contribución de la docencia de 75% y una contribución de las actividades de investigación de 25% aproximadamente. Por otra parte, los profesores de este perfil se caracterizan por no presentar una alta contribución de las actividades de gestión a la formación del IAA_E3.

En la figura 7.33 se puede observar que los profesores TEU, COL y AY son los que en promedio obtienen una valoración para su actividad académica, según el modelo E3, inferior al resto de categorías. De forma contraria, el profesorado CU, COD y TU obtiene en promedio una puntuación IAA mayor. En general, se observa que el profesorado correspondiente al perfil docente no obtiene una elevada puntuación en el indicador IAA_E3.

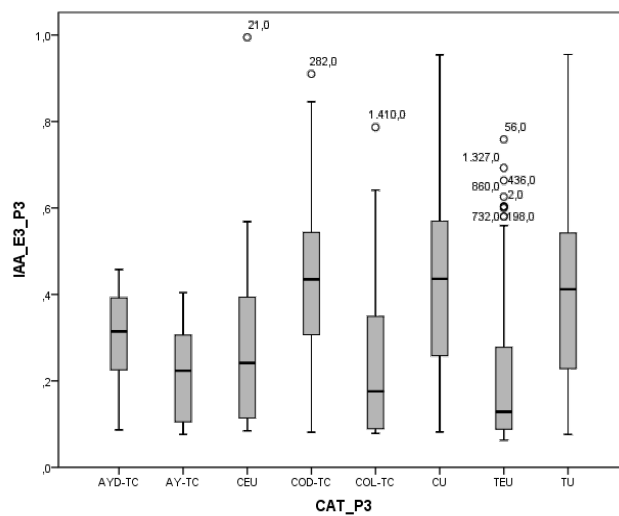


Figura 7.33. IAA_E3 por categoría laboral para el perfil Docente

Según la figura 7.34, el conjunto de profesores que pertenece al perfil Docente presenta una elevada puntuación en la dimensión de docencia. La mayoría del profesorado AY y AYD tiene una puntuación baja en las tres dimensiones, por lo que se trata de profesores con una baja puntuación en el modelo IAA como se verá más adelante en el análisis de los grupos de profesorado muy eficaces y poco eficaces en el desarrollo de la actividad académica.

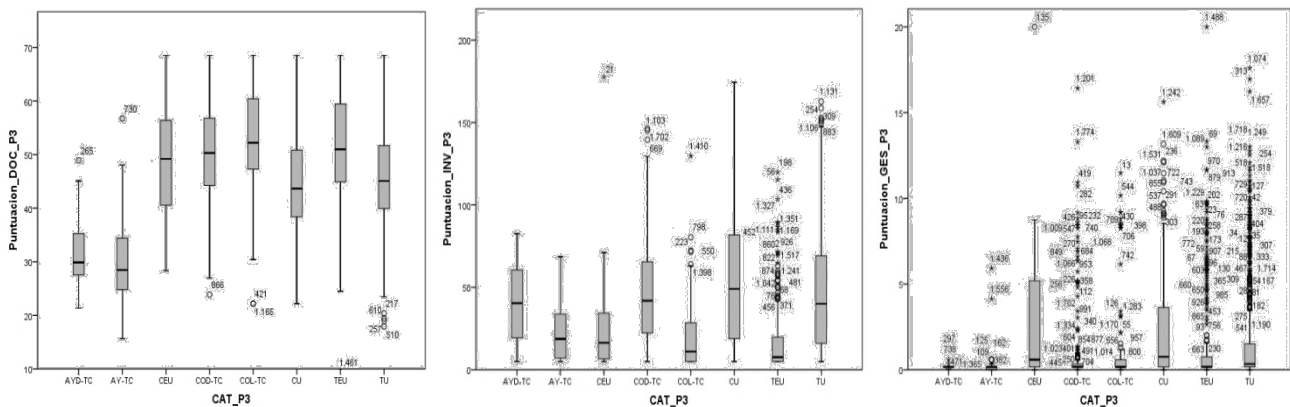


Figura 7.34. Puntuaciones de las dimensiones para el perfil Docente en el escenario E3

PERFIL INVESTIGADOR

El 8,1% del profesorado universitario pertenece a este perfil académico. Está formado en su mayoría por profesores CU, TU, COD y CEU, siendo su puntuación para el indicador IAA_E3 muy elevada según se observa en la figura 7.35. Se trata de profesores con una contribución de la docencia a su actividad académica de un 25% y una contribución de la investigación de un 75%, aproximadamente. Además, estos profesores presentan una baja contribución de la gestión a su actividad académica.

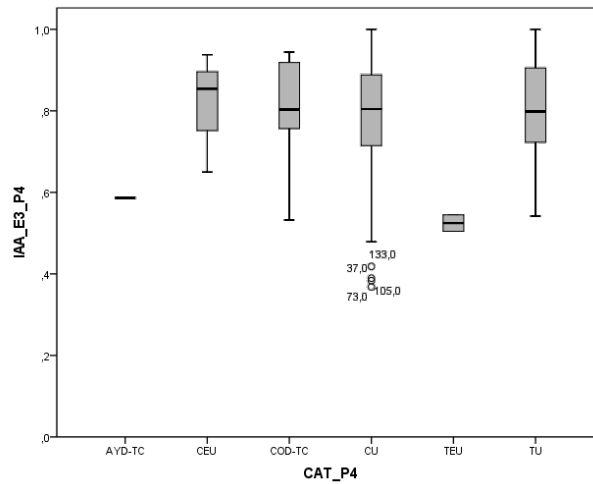


Figura 7.35. IAA_E3 por categoría laboral para el perfil Investigador

Tal y como se aprecia en la figura 7.36, el perfil Investigador, como su nombre indica, está formado por profesores con puntuaciones elevadas en la dimensión de investigación con respecto a los valores medios de la UPV. Este grupo de profesores, en general, obtiene una puntuación cercana a los valores medios de la UPV en docencia y una baja puntuación en la dimensión de gestión.

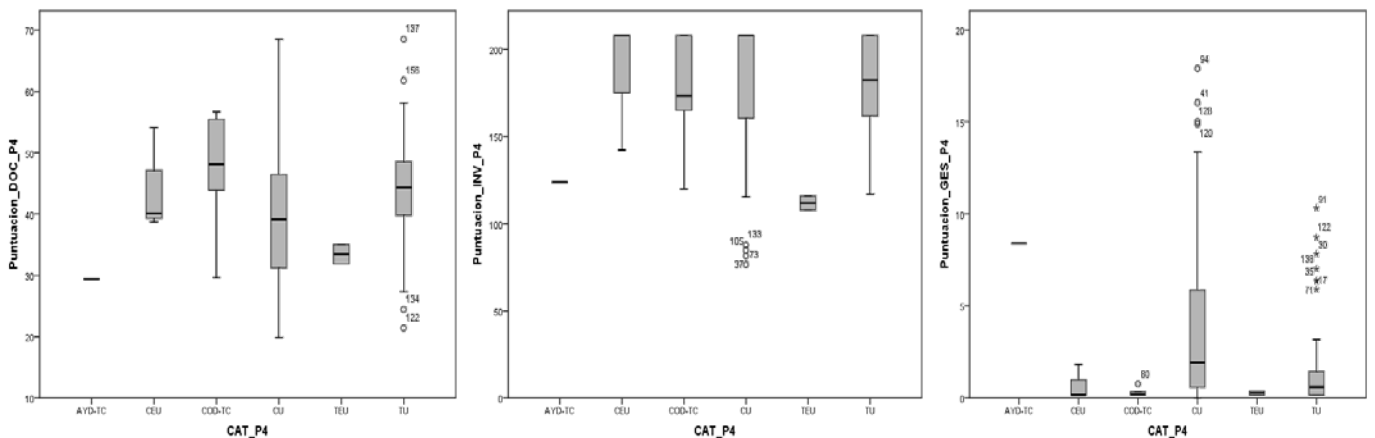


Figura 7.36. Puntuaciones de las dimensiones para el perfil Investigador en el escenario E3

Los resultados de este escenario muestran que la categorización de perfiles académicos realizada por el modelo del escenario E3 es adecuada según el conocimiento que se tiene sobre las características de las categorías laborales evaluadas.

Adicionalmente a los resultados anteriores, se ha realizado un estudio más profundo sobre las características del profesorado con una puntuación IAA_E3 muy alta y muy baja, es decir, un estudio del profesorado que ha obtenido un alto y bajo nivel de eficacia según el modelo del escenario E3.

En las figuras 7.37 y 7.38 se representa al profesorado con una puntuación en el IAA_E3 inferior a 0,2. Se trata de profesores que pertenecen, en su mayoría, al perfil Docente, y lo forman profesores de todas las categorías laborales. Este grupo de profesores representa el 30% del total de profesores evaluados, por lo que se trata de una cantidad elevada de profesores con un IAA_E3 inferior a 0,2.

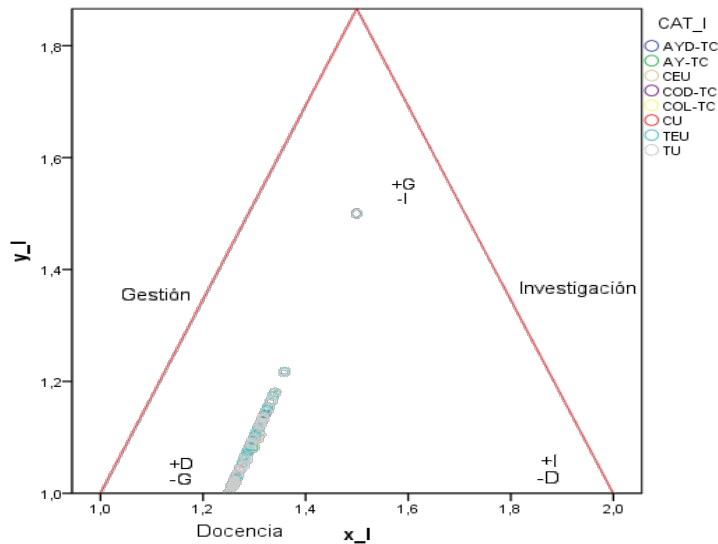


Figura 7.37. Contribuciones de las dimensiones para profesores con IAA_E3<0,2 en coordenadas baricéntricas

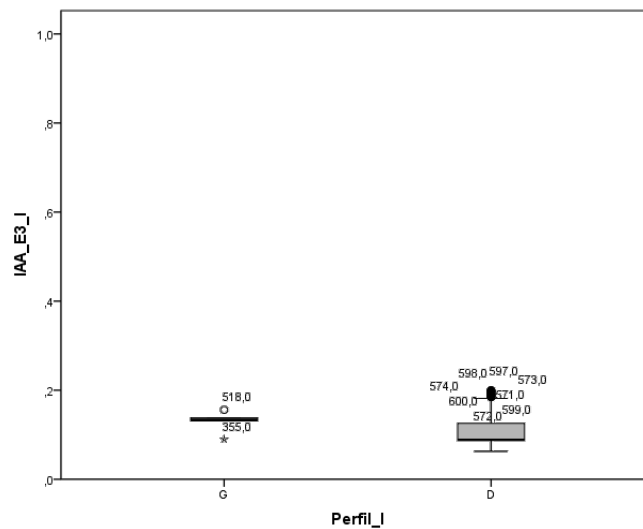


Figura 7.38. IAA_E3<0,2 por perfil de profesorado

Por el contrario, el grupo de profesores muy eficaces ($IAA_{E3} > 0,8$) está formado por profesores de todos los perfiles, pero únicamente de las categorías CEU, COD, CU y TU, tal y como muestra la figura 7.39 y 7.40. Este grupo de profesores corresponde al 5,6% del total de profesores evaluados.

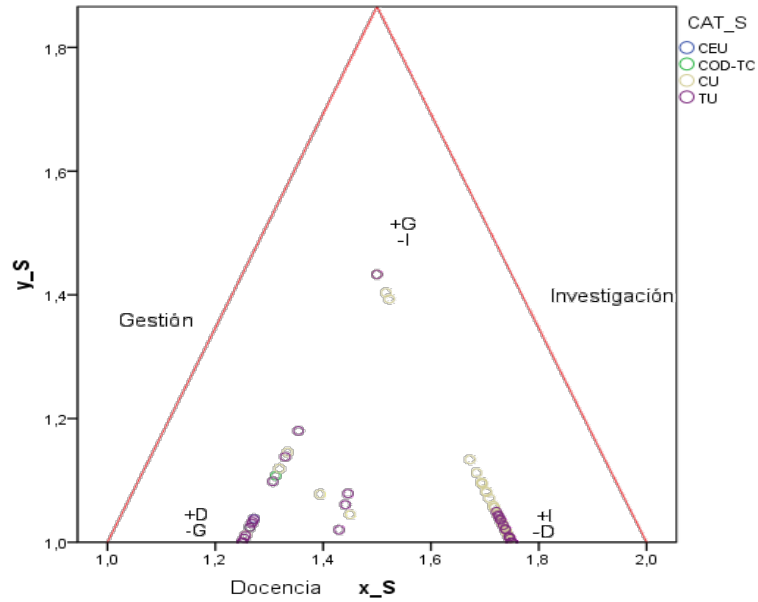


Figura 7.39. Contribuciones de las dimensiones para profesores con $IAA_{E3} > 0,8$ en coordenadas baricéntricas

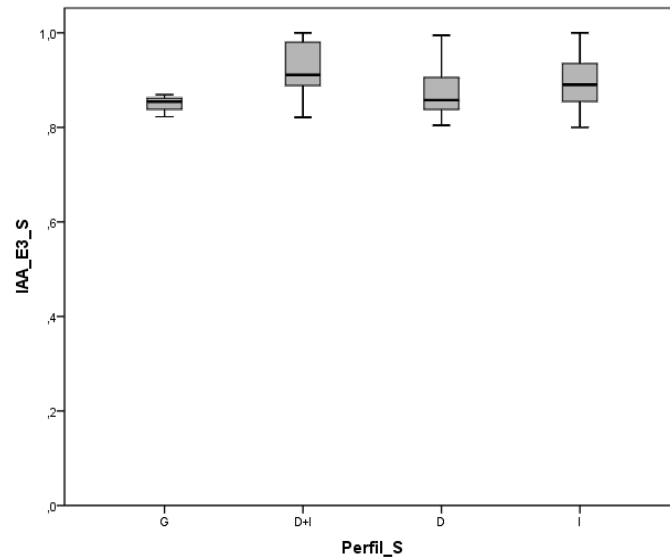


Figura 7.40. $IAA_{E3} > 0,8$ por perfil de profesorado

Escenario E4. Modelo BOD con restricciones de tipo II y acotación en la dimensión Investigación (IAA_E4)

Según los resultados expuestos anteriormente, el escenario E3 asigna una puntuación IAA_E3 menor que 0,2 al 30% de los profesores evaluados, lo que puede llevar a pensar que las restricciones impuestas en este escenario son excesivamente estrictas. Es por eso que el escenario que se presenta a continuación se define con unas restricciones más permisivas y los resultados obtenidos se adaptan mejor a los criterios y a las necesidades de la UPV para la evaluación de la actividad académica del profesorado.

Este último escenario E4 calcula una puntuación académica para cada profesor de la misma forma que el escenario E3, cambiando en este caso el valor de las restricciones de las contribuciones a cada una de las dimensiones. El modelo planteado en este escenario es el siguiente:

$$IAA_{E4_c}^* = \max_{(w_{c,1}, \dots, w_{c,q})} \sum_{i=1}^3 w_{c,i} I_{c,i}$$

s. a

$$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020$$

$$0,01 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,95$$

$$0,01 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,95$$

$$0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} I_{c,3}} \leq 0,8$$

$$w_{c,i} \geq 0 \quad i = 1, \dots, 3$$

$IAA_{E4_c}^*$: puntuación máxima del IAA para el profesor c con respecto al resto de profesores

$w_{c,i}$: pesos asignados por el modelo BOD para el profesor c

$I_{c,1}$: puntuación de la dimensión de docencia IAD para el profesor c

$I_{c,2}$: puntuación de la dimensión de investigación VAIP para el profesor c

$I_{c,3}$: puntuación de la dimensión de gestión G para el profesor c

Los valores de las restricciones se han fijado tomando el mínimo y el máximo de las opiniones de los expertos para cada dimensión. De esta forma, las restricciones de tipo II tienen en cuenta las opiniones del grupo de expertos y los distintos perfiles de profesorado.

Se ha realizado el mismo análisis comparativo con respecto al escenario de la UPV, obteniendo en este caso una puntuación para el profesorado superior con el modelo E4. Sólo existen 5 profesores que obtienen una mejor valoración con el escenario de la UPV y esto se

debe, al igual que en el escenario E3, a que el escenario de la UPV asigna a estos profesores unas contribuciones de las dimensiones que no cumplen las restricciones planteadas en el modelo E4.

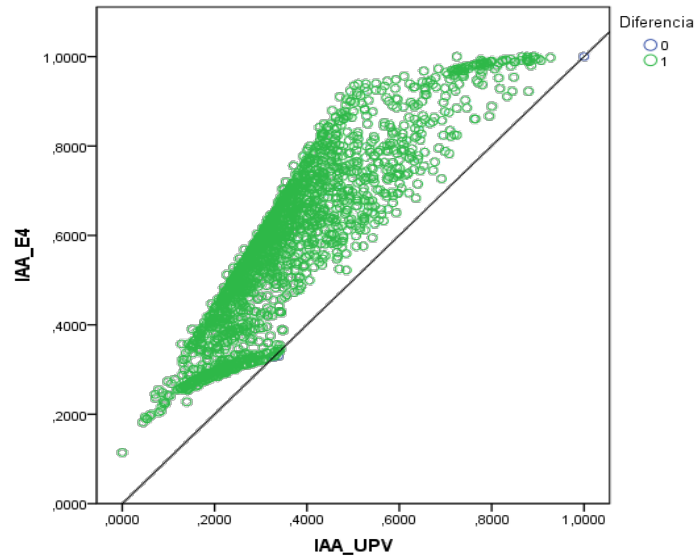


Figura 7.41. Comparación del IAA_E4 y IAA_UPV

Asimismo, se han comparado las posiciones del profesorado en el escenario E4 y las posiciones obtenidas en el escenario UPV. Tal y como se observa en la figura 7.42, aunque existe un grado de correlación elevado entre las posiciones que asignan ambos escenarios, se evidencia una ordenación muy distinta según el modelo de escenario utilizado.

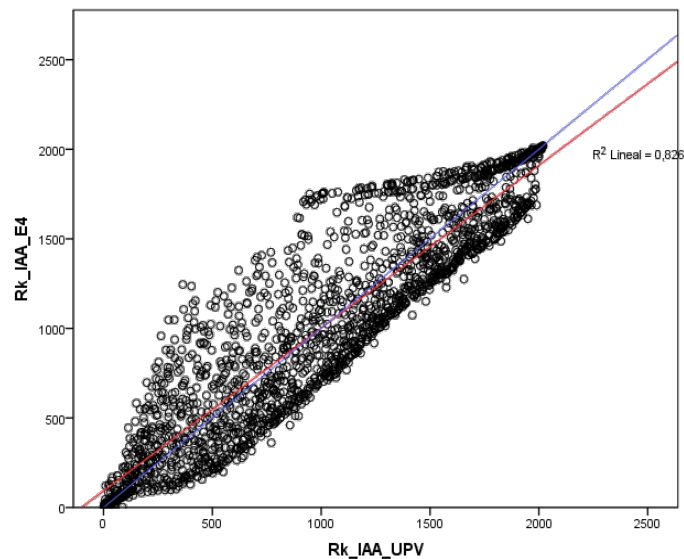


Figura 7.42. Comparación en posiciones de los escenarios E4 y UPV

Análisis descriptivo del escenario E4

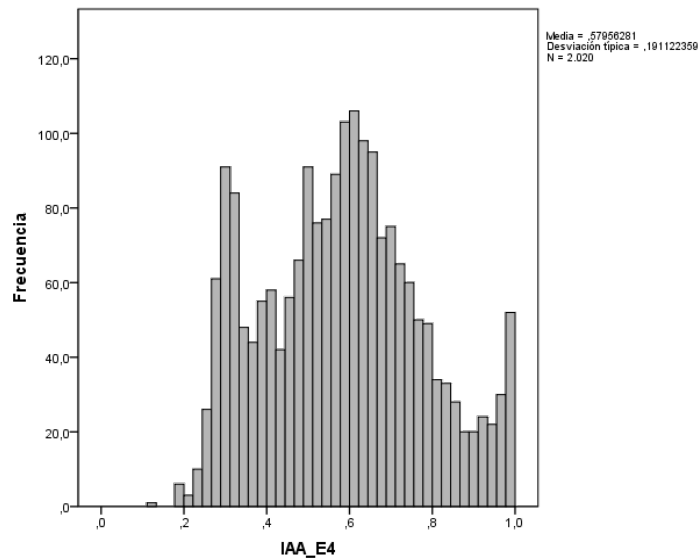


Figura 7.43. Histograma del escenario E4

A diferencia del escenario E3, que asigna una puntuación baja a una gran parte del profesorado, la distribución del escenario E4 muestra que este modelo asigna de forma equilibrada las puntuaciones del IAA en el rango [0,1], centrándose muchas de éstas alrededor de la puntuación 0,6.

La figura 7.44 muestra que, en promedio, todas las categorías laborales obtienen una puntuación superior a 0,5 exceptuando el profesorado AY, AYD y TEU. Por el contrario, el escenario E3 asigna puntuaciones inferiores a 0,5 para la mayoría de categorías laborales (véase figura 7.25).

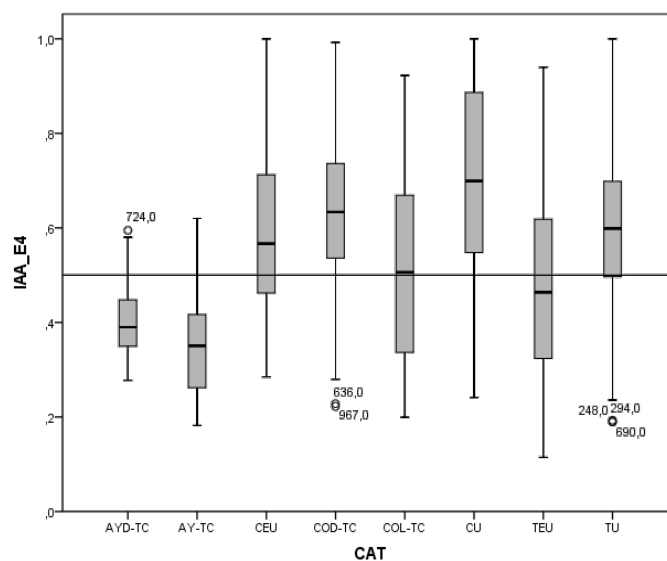


Figura 7.44. IAA_E4 por categoría laboral

Al analizar las dimensiones de docencia, investigación y gestión del IAA_E4 cabe destacar que, en general, el modelo E4 asigna una contribución de la docencia al índice IAA_E4 más elevada que las contribuciones de investigación y gestión. El único grupo de profesores que varía en gran medida este resultado son los CU. Este resultado se analiza de forma detallada más adelante.

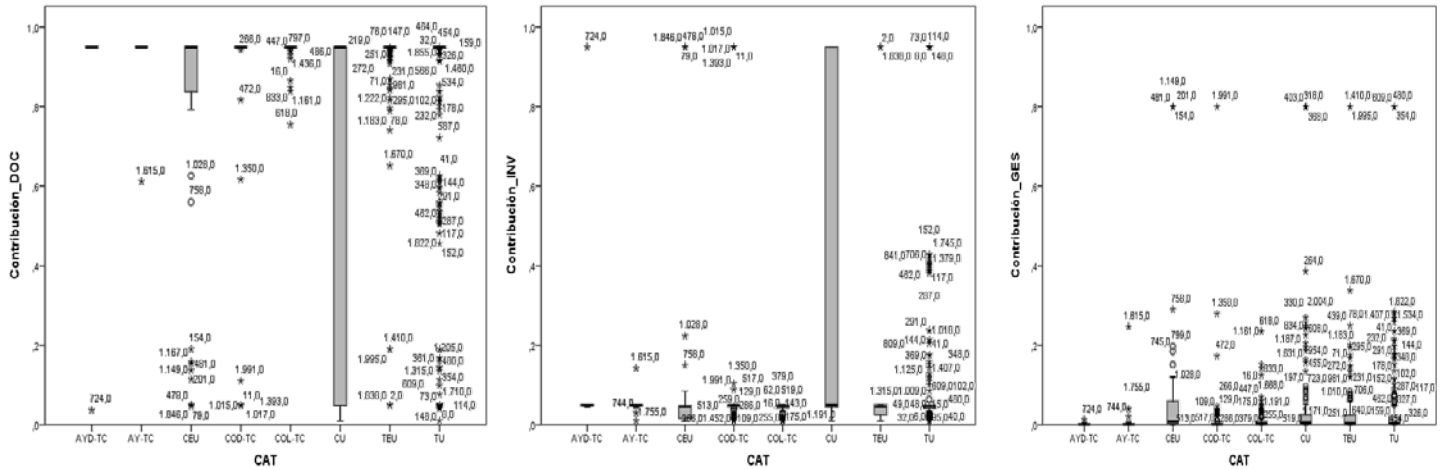


Figura 7.45. Contribución a las dimensiones por categoría laboral para el escenario E4

Categorización de los perfiles de profesorado en coordenadas baricéntricas

El cálculo de las contribuciones de las dimensiones asignadas por el modelo E4 en coordenadas baricéntricas ha permitido detectar diferentes comportamientos del profesorado según su grado de dedicación a la docencia, investigación y/o gestión.

Los perfiles obtenidos mediante este escenario coinciden con el escenario E3, aunque las restricciones impuestas en ambos modelos no son las mismas. El escenario E3 es más restrictivo sobre las dimensiones, mientras que el escenario E4 permite asignar como máximo una contribución del 95% para la dimensión de docencia e investigación y una contribución del 80% para las actividades de gestión. De esta forma, mediante el modelo E4 se han identificado más claramente los perfiles de profesores según la contribución de las dimensiones de docencia, investigación y gestión al índice.

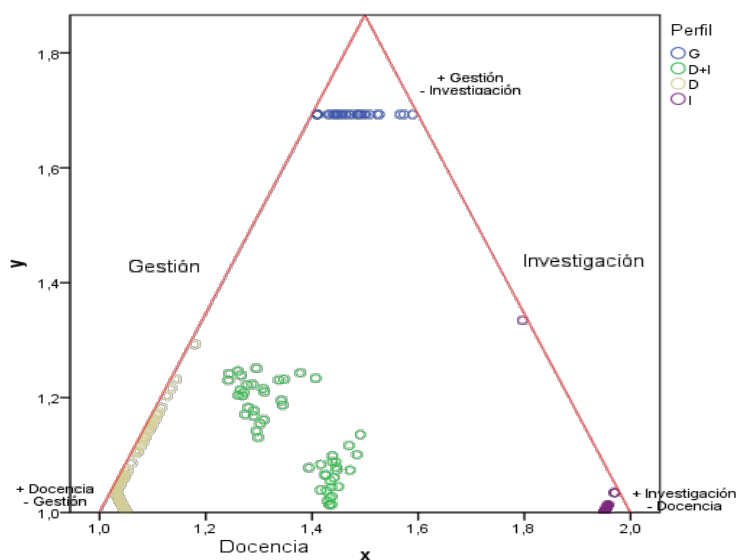


Figura 7.46. Contribuciones del IAA_E4 en coordenadas baricéntricas

A partir de la figura 7.46 se distinguen claramente cuatro perfiles diferentes de profesorado según las contribuciones de las tres dimensiones a la formación del IAA_E4. Los perfiles se definen de igual forma que en el modelo E3: *perfil Gestor* formado por el 2% del total del profesorado evaluado, *perfil Docente-Investigador* constituido por el 2,6% de profesores, *perfil Docente* formado por el 87,3% de profesores y *perfil Investigador* por el 8,1% de profesores.

A continuación se muestran con más detalle las características del profesorado que forma cada uno de los perfiles definidos.

CAT	% profesores de la CAT respecto al total	G	D+I	D	I
AY-TC	2,2%		¹ 2,2% ² 1,9%	97,8% 2,49%	
AYD-TC	1,2%			95,7% 1,24%	4,3% 0,6%
COL-TC	7,1%			100,0% 8,2%	
COD-TC	11,8%	0,4% 2,5%	0,4% 1,9%	96,6% 13,1%	2,5% 3,7%
TEU	21,2%	0,5% 5,0%		99,0% 24,0%	0,5% 1,2%
TU	38,7%	1,5% 30,0%	3,3% 50,0%	89,1% 39,5%	6,0% 29,7%
CEU	2,4%	10,4% 12,5%	4,1% 3,8%	79,2% 2,2%	6,3% 1,8%
CU	15,4%	6,4% 50,0%	7,0% 42,3%	52,7% 9,3%	33,8% 64,0%
TOTAL	100%	2,0%	2,6%	87,3%	8,1%

¹ Porcentaje de profesores de la categoría laboral que están en el perfil correspondiente con respecto al número total de profesores de esa categoría laboral

² Porcentaje de profesores de la categoría laboral que están en el perfil correspondiente con respecto al número total de profesores en dicho perfil

Tabla 7.8. Asociación de categorías laborales a los perfiles definidos en el escenario E4

Según el escenario E4, el profesorado COD, TEU, TU, CEU y CU se encuentra en la mayoría de los perfiles definidos. Sin embargo, el profesorado AY, AYD y COL sólo se clasifica en algunos de ellos. El profesorado AY se encuentra en su mayoría en el perfil Docente, únicamente poco profesores AY pertenecen al perfil Docente-Investigador. Lo mismo ocurre con el profesorado AYD que pertenece en su mayoría al perfil Docente, aunque se pueden encontrar algunos profesores AYD con un perfil Investigador. El profesorado COL se clasifica exclusivamente en el perfil Docente. Por otra parte, los perfiles Gestor, Docente-Investigador e Investigador están formados en su mayoría por TU y CU.

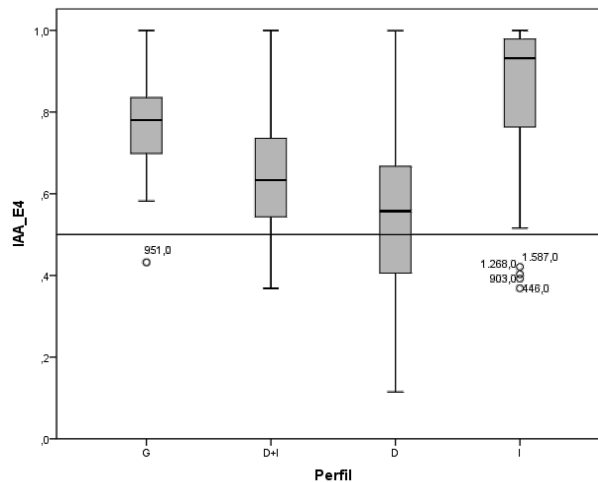


Figura 7.47. IAA_E4 por perfiles de profesorado

Tal y como se aprecia en la figura 7.47, el perfil Investigador es el que obtiene mayor puntuación media en el IAA_E4, seguido por el perfil Gestor. Por otra parte, el perfil Docente es el que menor puntuación media obtiene en el modelo. Si se comparan estos resultados con el IAA propuesto por la universidad se concluye que en ambos escenarios los profesores investigadores obtienen una alta puntuación en su valoración académica, siendo los profesores con un perfil docente los que menor puntuación obtienen. No obstante, en el escenario E4 también se pueden encontrar profesores de perfiles diferentes al Investigador con una muy buena puntuación en el IAA, tal y como se expone más adelante en el estudio de los profesores con un elevado nivel de eficacia.

Se ha realizado un estudio detallado de las características de cada uno de los perfiles obtenidos y del profesorado que lo forma.

PERFIL GESTOR

El perfil gestor está formado mayormente por profesores TU, CU y CEU con una puntuación en el IAA_E4 elevada tal y como se observa en la figura 7.48. Además, está formado únicamente por un profesor COD y dos profesores TEU.

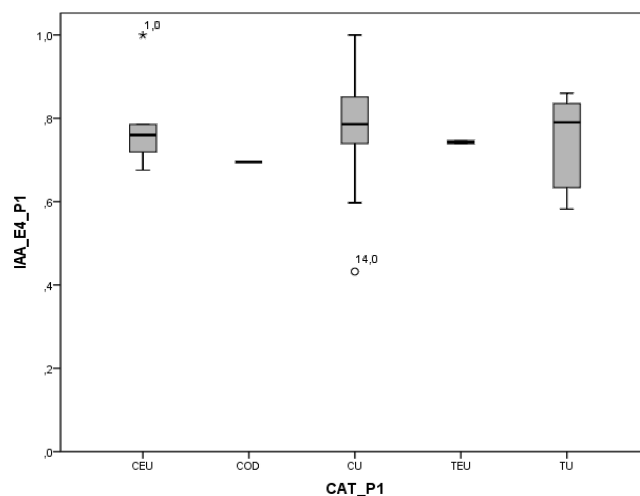


Figura 7.48. IAA_E4 por categoría laboral para el perfil Gestor

Se trata de profesores con una puntuación en la dimensión de docencia cercana a los valores medios de la UPV y una puntuación relativamente baja en la dimensión de investigación con respecto a los valores medios de la UPV. No obstante, en esta última dimensión se pueden encontrar profesores con una alta puntuación. Por último, este perfil está formado por profesores con una elevada puntuación en las actividades de gestión con respecto a la media de la UPV.

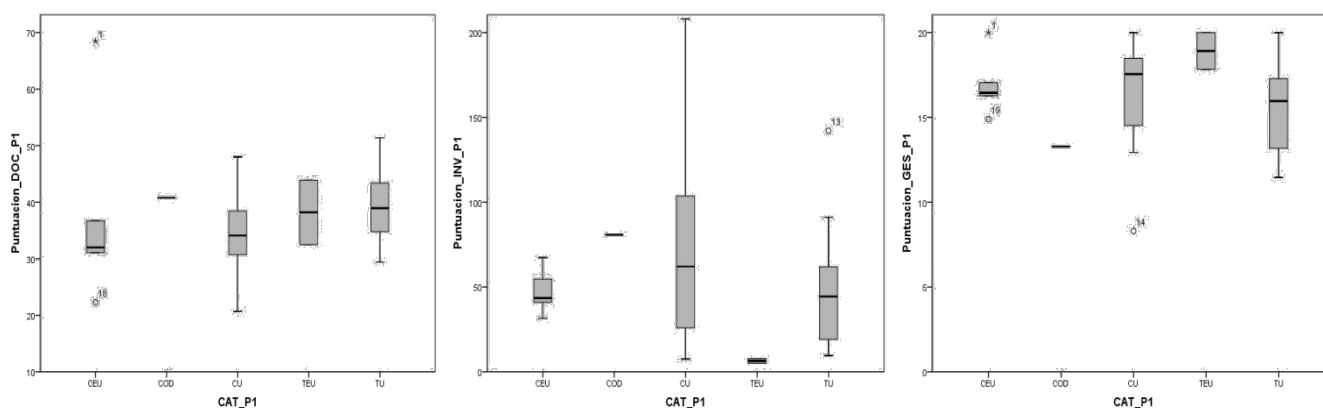


Figura 7.49. Puntuaciones de las dimensiones para el perfil Gestor en el escenario E4

PERFIL DOCENTE-INVESTIGADOR

El perfil Docente-Investigador está formado en su mayoría por profesores TU, CU (véase figura 7.50). También lo forman un profesor AY, un profesor COD y dos profesores CEU. La figura 7.51 muestra que los profesores TU y CU, que forman la mayoría de este perfil, tienen una puntuación para docencia e investigación bastante elevada con respecto a los valores medios de la UPV, siendo más bajas las puntuaciones obtenidas en las actividades de gestión.

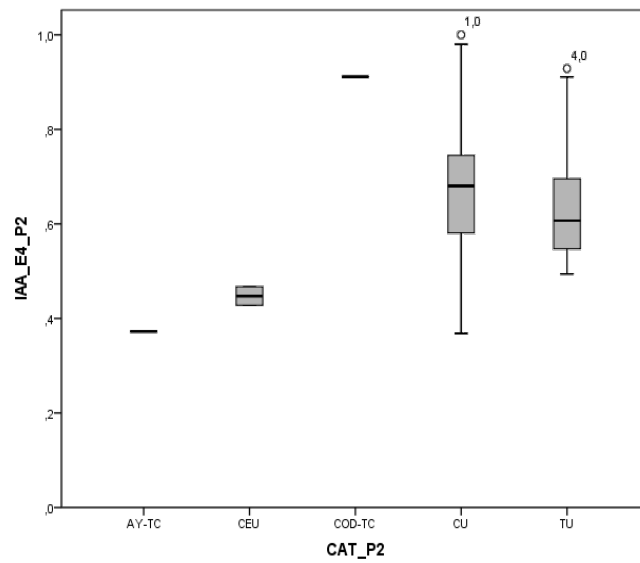


Figura 7.50. IAA_E4 por categoría laboral para el perfil Docente-Investigador

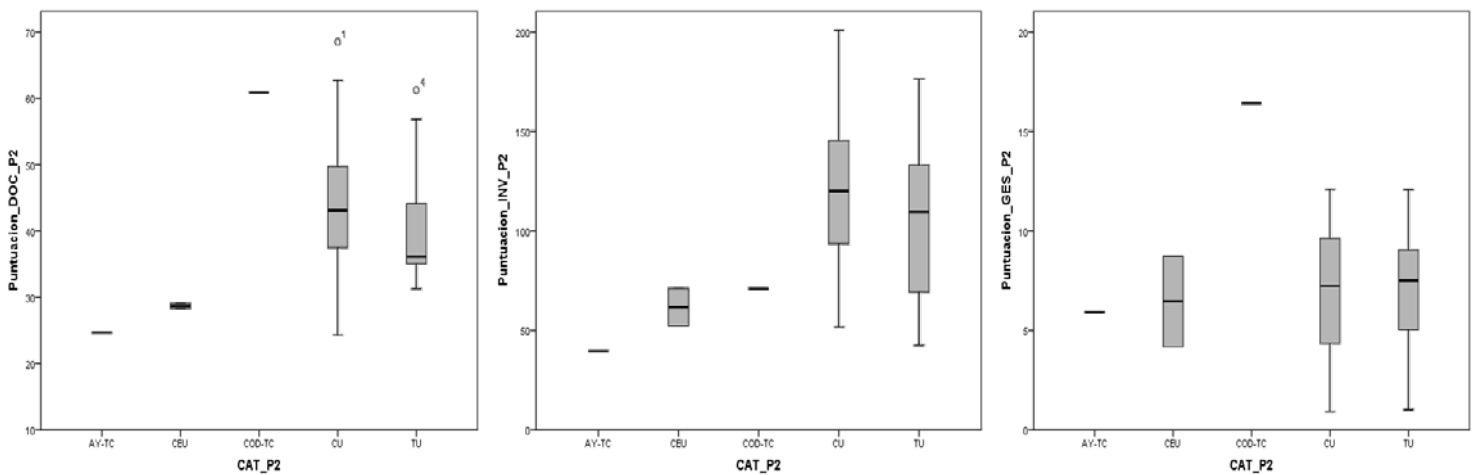


Figura 7.51. Puntuaciones de las dimensiones para el perfil Docente-Investigador en el escenario E4

PERFIL DOCENTE

El perfil Docente está formado por el 87,3% del profesorado evaluado, por lo que la contribución de la docencia a la formación del IAA_E4 predomina sobre la contribución de las demás dimensiones para la mayoría de profesores evaluados. Además, está formado por profesores de todas las categorías laborales, entre los que destaca el profesorado COD, TU, CU y CEU por tener una puntuación media en el IAA_E4 más elevada que el resto. Aun así, se trata del perfil con mayor variabilidad en las puntuaciones IAA de los profesores, según se observa en la figura 7.52.

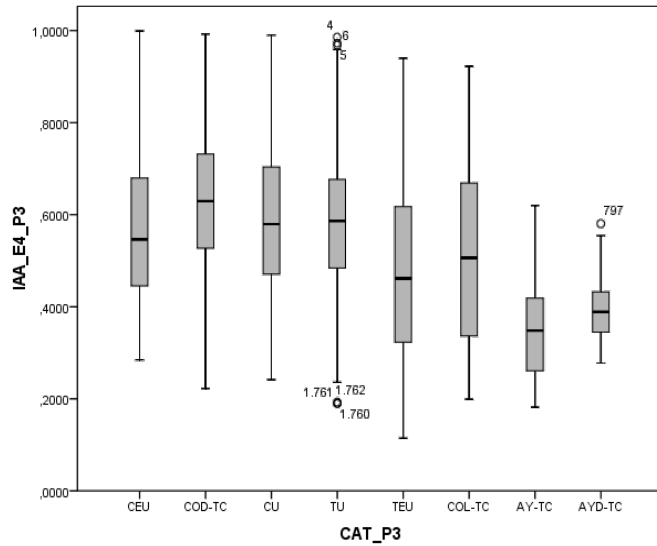


Figura 7.52. IAA_E4 por categoría laboral para el perfil Docente

La figura 7.53 muestra que también existe una gran variabilidad en las puntuaciones obtenidas en cada una de las tres dimensiones por el profesorado de este perfil. En general, cabe destacar que este grupo de profesores presenta una alta puntuación en la dimensión de docencia y una baja puntuación en la dimensión de investigación y gestión con respecto a los valores medios de la UPV.

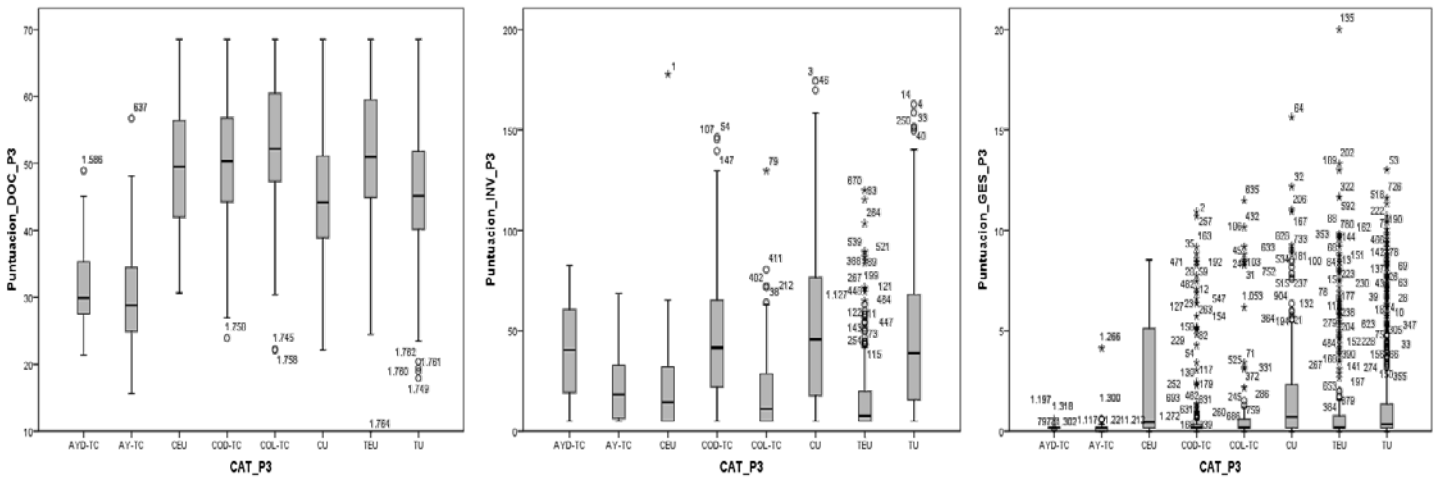


Figura 7.53. Puntuaciones de las dimensiones para el perfil Docente en el escenario E4

PERFIL INVESTIGADOR

A partir de la figura 7.54 se aprecia que, exceptuando al profesorado TEU y AYD, la mayoría de profesores de las categorías COD, TU, CU y CEU obtienen unas puntuaciones

superiores a 0,7. Por tanto, se considera que son profesores que realizan su actividad académica de forma eficaz con respecto al resto según el modelo E4.

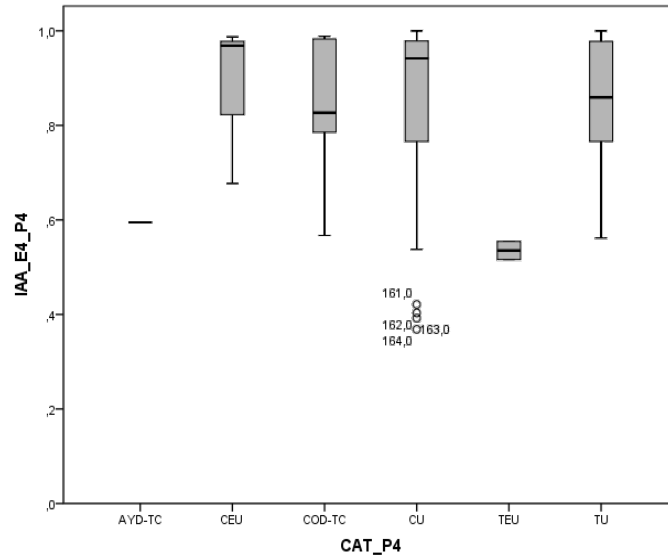


Figura 7.54. IAA_E4 por categoría laboral para el perfil Investigador

El perfil Investigador está formado por profesores con una elevada puntuación en la dimensión de investigación, una puntuación cercana a los valores medios de la UPV en docencia y una baja puntuación en gestión. No obstante, se pueden encontrar profesores en este perfil que tienen una máxima puntuación en la dimensión de docencia y muy buena gestión, situación que suele ocurrir con profesores CU.

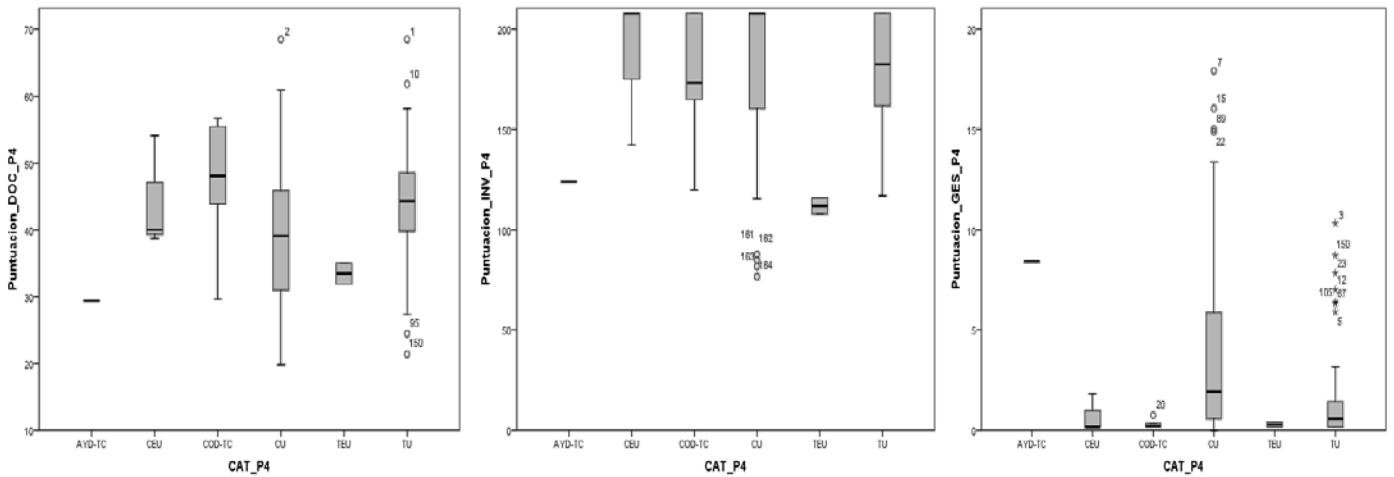


Figura 7.55. Puntuaciones de las dimensiones para el perfil Investigador en el escenario E4

El profesorado AYD y TEU forman la minoría de este grupo con puntuaciones en las dimensiones de docencia, investigación y gestión relativamente bajas con respecto al resto de categorías que forman este perfil.

De forma análoga al escenario E3 se ha realizado un estudio sobre el profesorado con menor y mayor puntuación obtenida en el escenario E4. En la figura 7.56 se han representado en coordenadas baricéntricas las contribuciones de las tres dimensiones para los profesores que han

obtenido una puntuación en el IAA_E4 inferior a 0,3. Nótese que este valor se ha fijado en 0,2 para el escenario de la UPV y el escenario E3. En este caso se ha considerado el valor 0,3, puesto que el escenario E4 asigna un IAA inferior a 0,2 a muy poco profesores.

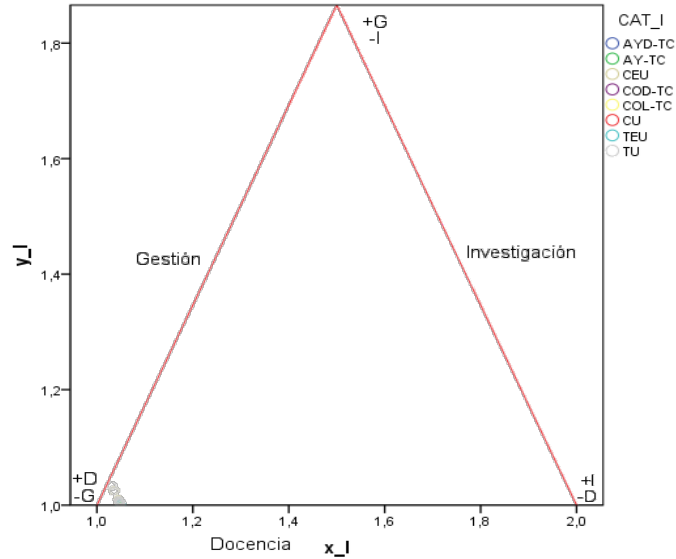


Figura 7.56. Contribuciones de las dimensiones para los profesores con IAA_E4<0,3 en coordenadas baricéntricas

Los resultados muestran que el profesorado que ha obtenido una valoración académica inferior a 0,3 pertenece al perfil Docente y está formado mayoritariamente por TU, aunque hay profesores de todas las categorías laborales como se puede apreciar en la figura 7.56. Este grupo de profesores representa un 7,2% del total de profesores evaluados.

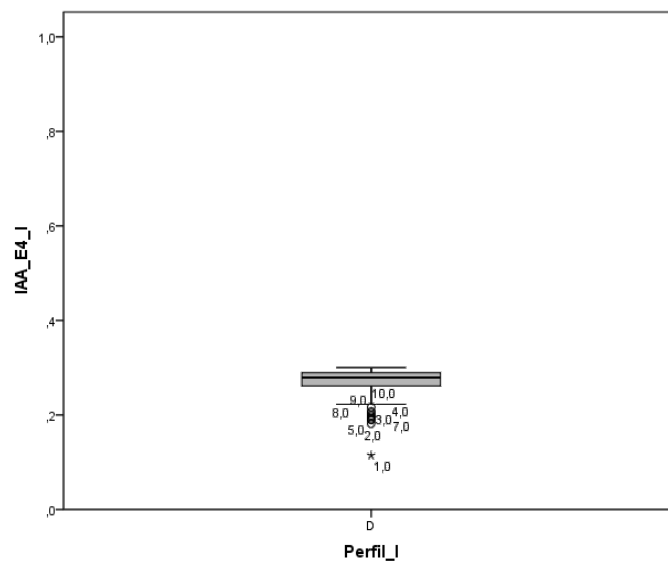


Figura 7.57. IAA_E4<0,3 por perfil de profesorado

De forma similar al escenario E3, se ha realizado el estudio de los profesores que han obtenido una alta puntuación en el escenario E4. Se trata, por tanto, de profesores eficaces con respecto al resto en el desarrollo de su actividad académica, ya que con sus pesos y su perfil definido según el modelo de evaluación, existen pocos o ningún profesor que superen su valoración académica. En la figura 7.58 se observa que, según el escenario E4, existen profesores muy eficaces en el desarrollo de la actividad académica que pertenecen a todos los perfiles, aunque ninguno de ellos pertenece a la categoría laboral de AY o AYD.

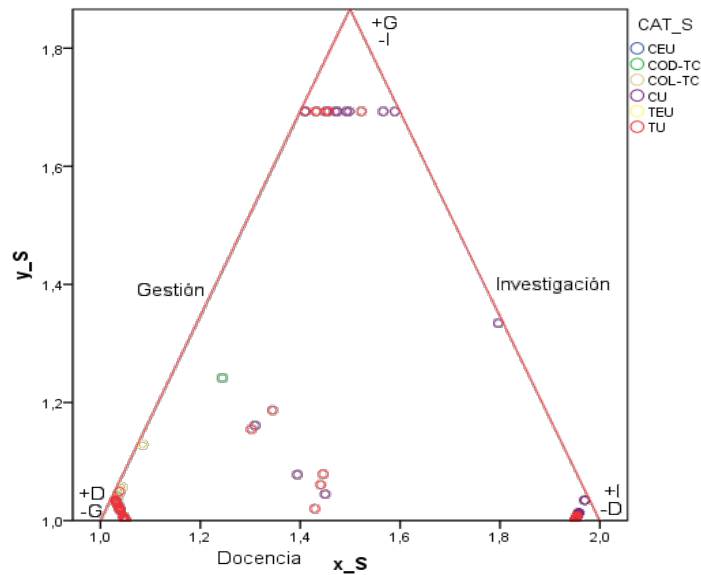


Figura 7.58. Contribuciones de las dimensiones para profesores con $IAA_{E4} > 0,8$ en coordenadas baricéntricas

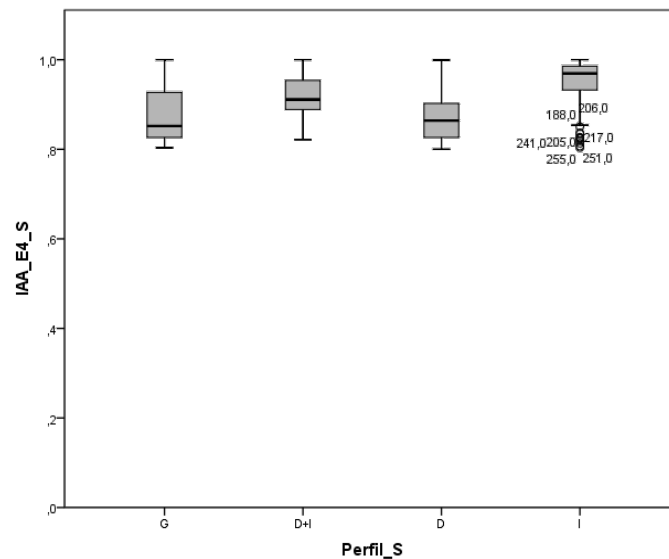


Figura 7.59. $IAA_{E4} > 0,8$ por perfil de profesorado

De la figura 7.59 cabe destacar que dentro del grupo de profesores con una valoración académica superior a 0,8, el perfil Investigador obtiene una puntuación media superior que el

resto en la puntuación del IAA_E4. Este grupo de profesores representa un 13% del total de profesores evaluados.

Los resultados de los diferentes escenarios propuestos han sido evaluados por el grupo de expertos de la UPV, resultando ser el escenario E4 el que más se ajusta a las directrices actuales de la universidad. Los expertos argumentan que las restricciones reflejan la situación real del PDI de la UPV y permiten definir distintos perfiles de profesorado según el desarrollo de las actividades de docencia, investigación y gestión con respecto al resto de profesores. Sin embargo, el escenario E4 no trata de ser único e inamovible, sino ajustable a las necesidades y cambios propuestos por la universidad. Por tanto, este capítulo proporciona una herramienta metodológica útil y adaptable para la construcción de un indicador compuesto en la gestión académica bajo ciertas directrices planteadas según la institución de estudio.

7.2.6. Visualización de los resultados de la metodología BOD

Las ilustraciones gráficas pueden ayudar en la interpretación final de los resultados de cada profesor. Existen muchas representaciones diferentes para interpretar los resultados individualizados de cada profesor. Para una revisión de las distintas metodologías que se pueden utilizar para representar los resultados de la técnica BOD véase (El-Mahgary y Lahdelma, 1995).

En este caso se ha realizado la representación aplicando la idea del gráfico de sectores. Este tipo de visualización de los resultados presenta la ventaja de combinar dos tipos de información en un único gráfico: el valor del IAA y las contribuciones de las dimensiones de docencia, investigación y gestión de cada profesor a su IAA ($\%w_1I_1$, $\%w_2I_2$ y $\%w_3I_3$, respectivamente). Además, mediante un gráfico de barras se representan las puntuaciones absolutas de las dimensiones de docencia, investigación y gestión. De esta manera, el profesor dispone de información para analizar su puntuación obtenida en cada una de las dimensiones con respecto al máximo que podría alcanzar.

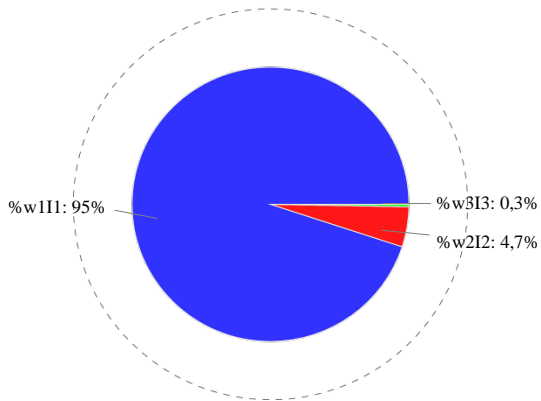
El valor del IAA viene representado por el tamaño del gráfico de sectores, mientras que las contribuciones corresponden a los distintos sectores del gráfico. La circunferencia externa dibujada en línea discontinua representa la máxima valoración que puede obtener un profesor en la metodología BOD, es decir, 100%. El objetivo de esta representación en sectores es proporcionar una idea al profesor evaluado de cómo de similar o diferente es su nivel de eficacia relativa al nivel de eficacia del profesor o profesores benchmark. Si el círculo del profesor evaluado cubre prácticamente la totalidad de la circunferencia indica que el nivel de eficacia relativa en el desarrollo de su actividad académica es similar al nivel de eficacia de los profesores benchmark¹⁸. De forma contraria, si el círculo tiene un área mucho menor que el área de la circunferencia, el profesor evaluado deberá mejorar su actividad académica con el fin de

¹⁸ Recuérdese que el(los) profesor(es) benchmark son aquel(llos) que con los pesos óptimos del profesor evaluado obtiene(n) un IAA mayor que el IAA del profesor evaluado.

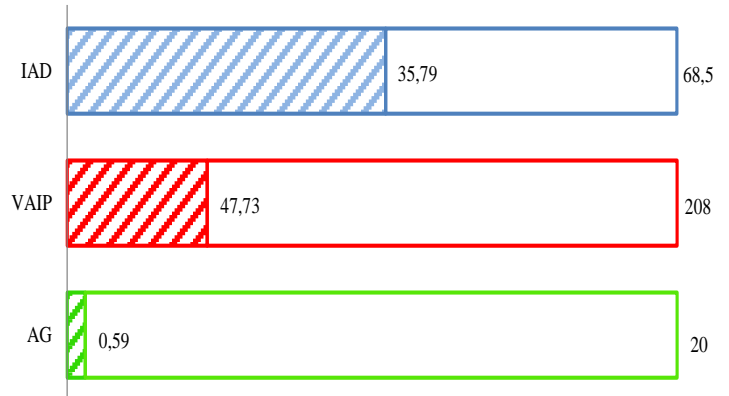
obtener un nivel de eficacia relativa similar o igual al profesorado de referencia o benchmark. De esta forma, el gráfico propuesto ofrece al profesorado un feedback de sus resultados muy fácil de interpretar que puede ayudar a mejorar su actividad académica.

En las figuras 7.60 y 7.61 se muestra dicha representación para seis profesores del escenario E4. Nótese que los gráficos de sectores ilustran la idea de la metodología BOD, puesto que las contribuciones de las dimensiones son diferentes en los profesores evaluados, pero a su vez todas cumplen las restricciones del modelo IAA_E4. Por ejemplo, el profesor con un ID 2659 obtiene una puntuación en el IAA_E4 de 49,30%. Este valor ofrece al profesor una medida de su eficacia relativa en cuanto a la actividad académica desarrollada con respecto al resto. La contribución de la docencia a la formación de su IAA_E4 es de 95%, la contribución de la investigación 4,7% y la contribución de la gestión 0,3%, por lo que se considera que es un profesor que desarrolla mucha más docencia que investigación o gestión. La categorización por perfiles realizada en el apartado anterior clasifica a este profesor en el perfil Docente. De forma análoga se pueden describir los resultados del resto del profesorado.

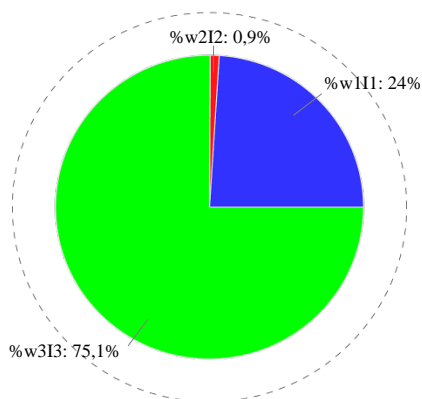
ID 2659, Perfil 3, COD-TC: 49,30%



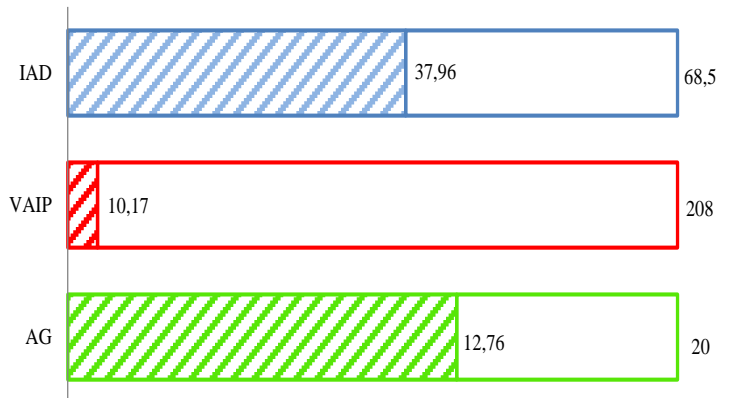
ID 2659



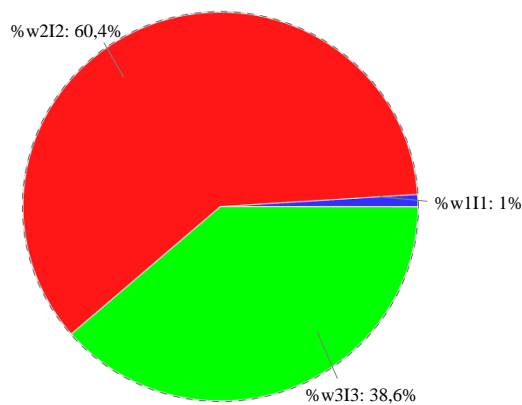
ID 1947, Perfil 1, TU: 64,99%



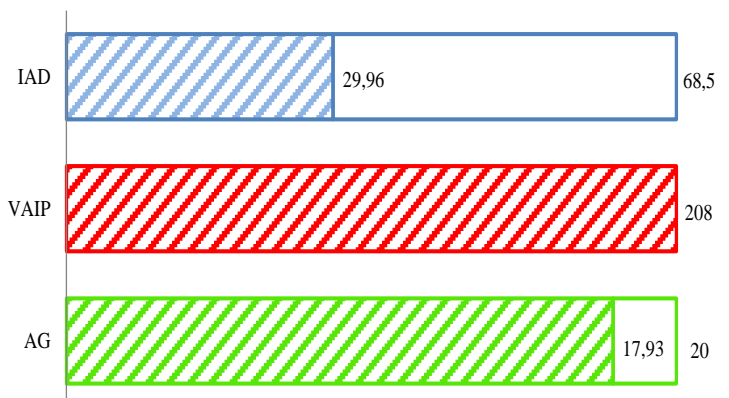
ID 1947



ID 359, Perfil 4, CU: 99,44%



ID 359

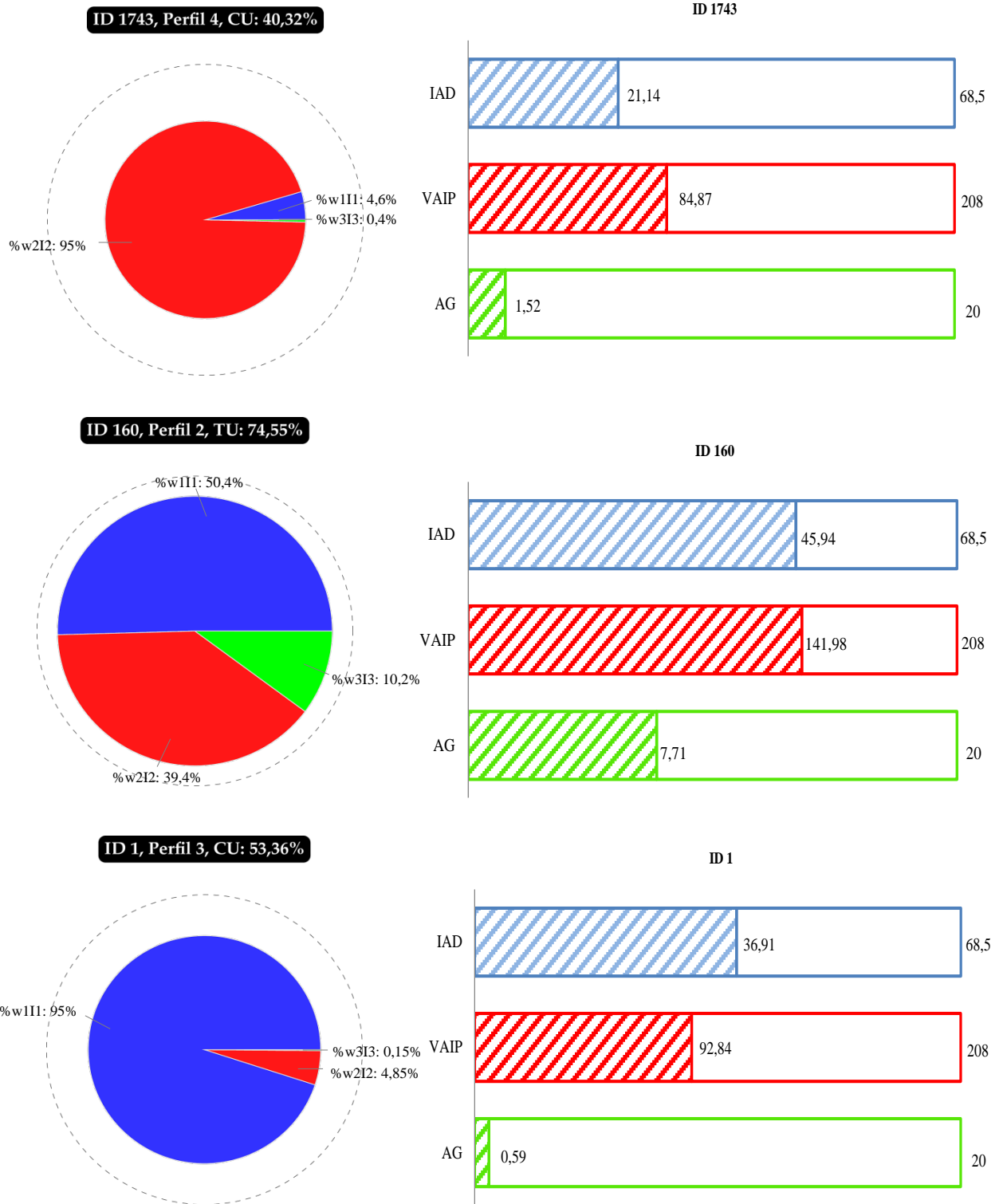


Leyenda:

- Contribución de la docencia al IAA_E4
- Contribución de la investigación al IAA_E4
- Contribución de la gestión al IAA_E4
- Puntuación del profesor para la dimensión de docencia
- Puntuación del profesor para la dimensión de investigación
- Puntuación del profesor para la dimensión de gestión

- Perfil 1:** Gestor
- Perfil 2:** Docente-Investigador
- Perfil 3:** Docente
- Perfil 4:** Investigador

Figura 7.60. Visualización de los resultados del IAA_E4 para ID 2659, ID 359, ID 1947



Leyenda:

- Contribución de la docencia al IAA_E4
- Contribución de la investigación al IAA_E4
- Contribución de la gestión al IAA_E4
- Puntuación del profesor para la dimensión de docencia
- Puntuación del profesor para la dimensión de investigación
- Puntuación del profesor para la dimensión de gestión

- Perfil 1:** Gestor
- Perfil 2:** Docente-Investigador
- Perfil 3:** Docente
- Perfil 4:** Investigador

Figura 7.61 Visualización de los resultados del IAA_E4 para ID 1743, ID 160, ID1

7.3. Resumen y conclusiones

En este capítulo se propone una metodología de construcción de un IC para la evaluación de la actividad académica del profesorado universitario. Para ello, se ha utilizado una técnica no paramétrica conocida como *Benefit of the Doubt* para la medición de la eficacia relativa del PDI de la UPV mediante la evaluación de las actividades realizadas en las dimensiones de docencia, investigación y gestión. Se trata de una técnica empleada cada vez más para la construcción de ICs en la evaluación de políticas de mejora y en la medición de la eficiencia de las IES.

La metodología empleada para la construcción del IAA presenta dos grandes ventajas. Por una parte, el IAA se puede interpretar como una medida global constituyendo una herramienta de ayuda para la universidad en el reconocimiento de resultados, mediante diferentes instrumentos: incentivos, carrera profesional, asignación de personal, etc. Por otra parte, el IAA proporciona información al profesorado sobre las fortalezas y debilidades de su actividad académica en cada dimensión evaluada con respecto al resto del profesorado, y permite planificar estrategias de mejora. Aunque la metodología presenta ciertas limitaciones, por el uso de valoraciones subjetivas en su desarrollo, se considera una herramienta de medición adecuada para la evaluación de la actividad académica del profesorado de la UPV.

En primer lugar, un estudio metodológico del escenario propuesto por la UPV para la evaluación de la actividad académica muestra que el tipo de normalización aplicado no es adecuado debido a la distribución asimétrica de las dimensiones de investigación y gestión. Por otra parte, se ha aplicado un análisis de sensibilidad global para evaluar la calidad y validez del IAA_UPV. Los resultados evidencian que el conjunto de pesos efectivos que miden la importancia de cada dimensión en términos de varianza no corresponde con la importancia deseada por la universidad. Por último, la idea de fijar un peso a cada una de las dimensiones del modelo, cuando se desean agregar dimensiones que a su vez son complejas y multidimensionales, es criticada por muchos autores en la literatura. No se conoce una definición precisa del concepto de profesor eficaz, por lo que en general, los expertos y tomadores de decisiones discrepan sobre la importancia que debe tener cada una de las dimensiones del estudio. Por tanto, la dificultad de modelizar los procesos de las actividades principales de un docente universitario justifican el uso de la técnica BOD.

Posteriormente, se han propuesto diversos escenarios de trabajo que utilizan la técnica BOD para la evaluación de la actividad académica del profesorado. En el primer escenario E1 se aplica la metodología BOD sin restricciones sobre los pesos. La principal limitación de este modelo es que asigna, en la mayoría de los casos, el máximo peso a aquella dimensión con una puntuación elevada y un peso nulo al resto de dimensiones. En consecuencia, el modelo evalúa a cada profesor centrándose en una de las dimensiones e ignorando el resto y, además, se obtienen resultados que no son justificables desde el punto de vista teórico. Es por eso que este escenario no se ha considerado factible para la evaluación de la actividad académica del profesorado.

El segundo escenario planteado, E2, introduce un tipo de restricciones sobre los pesos conocido como restricciones de proporción. Durante el proceso de construcción del IAA no se obtiene un consenso sobre el peso fijo que debe tener cada una de las dimensiones, pero sí se

tiene información de la opinión de los expertos sobre cada una de éstas. Por tanto, para construir el escenario E2 los expertos han llegado a un consenso sobre las restricciones que deben cumplir las contribuciones de cada una de las dimensiones al IAA. Las restricciones introducidas en este escenario permiten definir perfiles de profesorado según la distinta contribución de la docencia, investigación y gestión al IAA. Sin embargo, el modelo del escenario E2 es inconsistente cuando dos de las dimensiones de un profesor presentan puntuaciones nulas, por lo que tampoco se ha considerado válido para medir la actividad académica del profesorado.

Para resolver la presencia de ceros en la base de datos se ha fijado una cota inferior en la dimensión de investigación para todo el profesorado. La idea de incluir esta cota inferior se ha consultado con el grupo de expertos y se ha justificado de forma favorable. Las restricciones definidas en el segundo escenario junto con la inclusión de esta cota constituyen el tercer escenario definido, E3. Sin embargo, la distribución del IAA_E3 muestra que el modelo asigna a la mayoría de profesores una puntuación para la actividad académica inferior a 0,5. Por tanto, el escenario E3 se ha considerado excesivamente restrictivo para la evaluación académica del profesorado y no se adapta a las necesidades actuales de la UPV.

Finalmente, se ha planteado un último escenario E4 en el que se ha tomado el máximo y el mínimo de las opiniones de los expertos para cada una de las dimensiones como restricciones del modelo. Este último escenario introduce en el modelo toda la variabilidad existente en las opiniones de los expertos. En este caso, el modelo IAA_E4 asigna a los distintos profesores una puntuación para la actividad académica más elevada que el escenario E3. Además, se han estudiado los distintos perfiles de profesorado según el grado de contribución de las dimensiones a la formación de su IAA_E4 con respecto al resto del profesorado, identificando cuatro perfiles diferenciados: perfil Docente, perfil Gestor, perfil Docente-Investigador y perfil Investigador. Asimismo, se ha estudiado la tipología de profesores que pertenece a cada perfil y se han identificado los profesores con un IAA inferior a 0,3 y superior a 0,8. El escenario E4 presenta una mayor flexibilidad en las restricciones de las contribuciones y esto hace que permita identificar a profesores de todas las categorías laborales con una elevada puntuación en su actividad académica. Por tanto, se trata del escenario que más se ajusta a las actuales orientaciones de la UPV. No obstante, la metodología aplicada se puede adaptar a cualquier cambio en las directrices de la UPV en cuanto a la evaluación de la actividad académica del profesorado.

Por último, se ha propuesto una forma de visualización de los resultados obtenidos en el escenario E4. Esta representación proporciona un feedback al profesorado de su actividad académica desarrollada con respecto al resto. Se trata de un método de visualización sencillo de interpretar que permite detectar carencias del profesor y diseñar políticas de mejora en el desarrollo de su actividad académica.

Capítulo 8

Conclusiones y líneas futuras de investigación

El uso de ICs como herramienta de evaluación en el ámbito de la educación superior se ha incrementado notablemente en los últimos años. Se trata de un instrumento de evaluación simple de interpretar que permite ilustrar y comprender el funcionamiento de las actividades de una universidad, recibir un feedback sobre las actuaciones evaluadas con el fin de establecer políticas de mejora y rendir cuentas a la institución para la toma de decisiones en el contexto de evaluación. No obstante, hay que tener especial cuidado en su metodología de construcción, ya que ICs mal contruidos pueden llevar a interpretaciones y toma de decisiones erróneas. Tal y como se afirma en un comunicado sobre ICs realizado por la Comisión Europea:

“[...] Composite indicators would have the advantage of providing a broader coverage of information than can be included in the current list of structural indicators and they would also allow for a reduction in the number of indicators presented in the list. However, because composite indicators invite strong policy messages to be concluded they need to be robust and based on a sound methodology. More work is therefore necessary to develop such indicators, to examine how they could be integrated into the list of structural indicators and to assess the consistency of the policy messages they send” (European Commission, 2001).

La investigación abordada en esta Tesis Doctoral se fundamenta en la necesidad por parte de las instituciones de educación superior de desarrollar herramientas eficientes para el apoyo en la toma de decisiones educativas. El estudio se centra, en particular, en el diseño de instrumentos de evaluación para la gestión interna de la calidad de la actividad del profesorado universitario perteneciente a la Universitat Politècnica de València, para favorecer así su desarrollo y reconocimiento, definir planes de mejora e innovación docente y mejorar y garantizar la calidad de la enseñanza.

En el proceso de construcción de ICs intervienen etapas en las que se emiten juicios de valor que son objeto de debate entre las partes interesadas. Las etapas más criticadas corresponden a las fases de ponderación y agregación de los indicadores simples. Según Sharpe (2004):

“The aggregators believe there are two major reasons that there is value in combining indicators in some manner to produce a bottom line. They believe that such a summary statistic can indeed capture reality and is meaningful, and that stressing the bottom line is extremely useful in garnering media interest and hence the attention of policy makers. The second school, the non-aggregators, believe one should stop once an appropriate set of indicators has been created and not to the further step of producing a composite index. Their key objection to aggregation is what they see as the arbitrary nature of the weighting process by which the variables are combined”.

Así pues, existe un debate continuo sobre la construcción de ICs, a pesar de que cada vez es más frecuente su uso en el ámbito de la educación superior para la evaluación interna de las universidades y para las comparaciones entre éstas mediante los conocidos rankings de universidades. Es por eso que hay que cuidar su metodología de construcción tanto desde el punto de vista teórico como estadístico para una mayor aceptación de éstos en la sociedad.

La técnica de agregación lineal empleada para la combinación de los indicadores simples en un único índice suele ser la más utilizada debido a su simplicidad y transparencia (Nardo et al., 2008). Sin embargo, esta técnica es criticada por su naturaleza compensatoria, de tal forma que un rendimiento bajo en algunos de los indicadores se compensa por altos valores en otros. La técnica de agregación geométrica es una solución intermedia entre la compensación total y los métodos de agregación no compensatorios, no obstante, también presenta un comportamiento compensatorio entre sus indicadores. La principal crítica que reciben estas técnicas de agregación es la interpretación dada a los pesos nominales como medidas de importancia de cada uno de los indicadores definidos, cuando en realidad son factores de escala que informan de la compensación entre indicadores (Munda, 2008; Decancq y Lugo, 2013). Esta confusión puede poner en duda la calidad del IC construido, por lo que es necesario aplicar un análisis de sensibilidad global para estudiar la validez del IC.

El procedimiento de modelización *State Dependent Parameter* aplicado a los escenarios propuestos para la evaluación de la actividad docente del profesorado de la UPV ha permitido evaluar la calidad de los escenarios mediante el cálculo de los efectos principales de los indicadores definidos en el análisis de sensibilidad global. Dichos efectos principales ofrecen una definición precisa de la importancia, en términos de varianza, que tiene cada indicador simple en la formación del IC. Una discrepancia elevada entre el conjunto de efectos principales y la importancia asignada a cada indicador por el grupo de expertos revela que el IC desarrollado no mide lo que pretende medir. Como consecuencia, se evidencia la necesidad de reflexionar sobre la metodología aplicada para la construcción de los ICs y sobre las implicaciones de las elecciones metodológicas.

En muchas ocasiones resulta complicado encontrar pesos nominales que proporcionen la importancia deseada de los indicadores simples por el grupo de expertos, por lo que se aconseja evitar la asociación de pesos nominales con coeficientes de importancia en las técnicas de

agregación compensatoria. Una posible solución es considerar el uso de un método de agregación no compensatorio en el que los pesos asignados a los indicadores toman el papel de medidas de importancia (Munda, 2008).

El modelo propuesto denotado como E15 se ha construido con el objetivo de obtener un efecto principal de los indicadores simples lo más ajustado posible a la opinión del grupo de expertos de la CED de la UPV, por lo que se trata de un escenario válido según las orientaciones marcadas por la universidad. Este escenario propuesto no trata de ser el único escenario factible para la medición de la actividad docente, sino una herramienta metodológica para la construcción de ICs ajustable a las necesidades que presente la institución.

El análisis de sensibilidad global debe ir acompañado de un análisis de incertidumbre que permita estudiar la robustez de los ICs frente a pequeñas variaciones en la estructura de los pesos nominales asignados a los indicadores en el proceso de construcción. La aplicación del análisis de incertidumbre muestra, mediante la aplicación de la metodología Copeland, que el escenario E15 ha resultado ser el más robusto. Este resultado proporciona información interesante a la universidad en el caso en que se decidiese realizar una reestructuración del esquema de ponderación del IC por decisiones estratégicas de la institución.

A menudo, los expertos no logran llegar a un consenso sobre la importancia que debe tener cada uno de los indicadores simples en el desarrollo del IC. Resulta mucho más sencillo llegar a un acuerdo en la asignación de restricciones a las dimensiones que reflejen la contribución permitida de cada una éstas en la formación del IC. Este es el caso de la construcción del IAA que engloba las actividades de docencia, investigación y gestión realizadas por el profesorado de la UPV, propuesto en el último capítulo de la tesis. La asignación de un peso fijo a cada una de las dimensiones resulta difícil de consensuar debido a los diferentes perfiles de profesorado y a su diferente dedicación a la docencia, investigación y gestión.

La técnica de ponderación y agregación no paramétrica *Benefit of the Doubt* se considera adecuada para la construcción del IAA, ya que presenta la ventaja de asignar un conjunto de pesos personalizado a cada profesor según su eficacia relativa en las tres dimensiones con respecto al resto, de forma que su puntuación en el índice sea máxima y se cumplan las restricciones de las dimensiones del modelo asignadas por el grupo de expertos de la CED de la UPV. La idea de asignar un conjunto de pesos individualizado a cada profesor con el fin de obtener una IAA máximo es mucho más aceptada por el profesorado evaluado que el uso de un conjunto de pesos fijo e igual para todo el profesorado.

El escenario propuesto E4 ha resultado ser adecuado según las orientaciones y necesidades de la universidad. Este escenario introduce en su modelo unas restricciones mucho más permisivas que el escenario E3 sobre las contribuciones de la docencia, investigación y gestión a la formación del índice. Mediante este escenario se asigna un esquema de ponderación a cada profesor según su actividad académica con respecto al resto, de forma que se cumplan las

restricciones del modelo. Las contribuciones de las dimensiones calculadas para cada profesor proporcionan una idea de su perfil docente. Así pues, los perfiles identificados mediante el uso de coordenadas baricéntricas han sido cuatro: perfil Gestor, Perfil Docente-Investigador, Perfil Docente y perfil Investigador. Cada uno de estos perfiles indica que la contribución de la actividad correspondiente es muy elevada en la formación del IAA del profesor.

Una de las etapas más importantes en la construcción de ICs corresponde a la visualización de los resultados. Es muy importante transmitir de la manera más sencilla y clara posible los resultados obtenidos a los tomadores de decisiones. La técnica de visualización propuesta para el IAA del escenario E4 proporciona un feedback de los resultados obtenidos para cada profesor que ofrece una idea clara del nivel de eficacia relativa de la actividad académica desarrollada por el profesor con respecto al resto. Además, esta técnica informa del grado de contribución de cada una de las dimensiones a la formación de su IAA. De esta forma, el docente dispone de una imagen general del desarrollo de su actividad académica que le permite establecer mejoras en su actuación docente.

Sobre estas conclusiones debe señalarse la principal limitación del estudio que corresponde al uso de opiniones subjetivas en el proceso de construcción de ICs. Son muchas las etapas que se han construido a partir de valoraciones subjetivas de expertos: elección del conjunto de indicadores, ponderación, análisis de sensibilidad, consideración de una cota mínima para el uso de la técnica BOD. Por ello, el procedimiento de evaluación utilizado debe ser válido, robusto y justificado por el grupo de expertos.

A partir de los resultados obtenidos en este trabajo y de la principal limitación planteada, se propone como línea de investigación futura estudiar en profundidad las diferentes variantes del método BOD que tratan de resolver problemas analíticos que no cubre el modelo empleado. Algunos autores proponen el uso de modelos que cumplen la propiedad de invarianza traslacional para solucionar la presencia de puntuaciones nulas o negativas en los datos (Ali y Seiford, 1990; Pastor y Lovell, 1995; Pastor, 1996).

Por otra parte, se plantea la posibilidad de aplicar técnicas de análisis de sensibilidad alternativas al método no paramétrico *State Dependent Parameter* que permitan evaluar la calidad del IC construido. Xu y Gernter (2008); Kucherenko, Tarantola y Annoni (2012); Paruolo, Saltelli y Saisana (2013), han desarrollado métodos de cálculo de los efectos principales aplicados a datos que presentan un cierto grado de correlación entre las variables. Se propone aplicar éstas técnicas y comparar los resultados obtenidos.

Asimismo, es de interés reconsiderar el uso de técnicas de agregación multi-criterio no compensatorias para poder tratar a los pesos nominales como coeficientes de importancia (Munda, 2008). De esta forma, se evita uno de los principales problemas en la construcción de ICs con un esquema de ponderación compensatorio.

Igualmente, se sugiere hacer un estudio sobre las diferentes técnicas de visualización e interpretación de los resultados del análisis de incertidumbre cuando el tamaño de la muestra utilizado es elevado. Los ICs construidos hasta el momento consideran la evaluación de pocas unidades de análisis que suelen ser países, universidades, facultades, etc., por lo que los métodos de visualización de gráficos o tablas permiten analizar los resultados del análisis de incertidumbre (Saisana y Tarantola, 2002; Nardo et al. 2008). Sin embargo, tal y como se ha visto en este estudio, éstos métodos de visualización no aportan resultados relevantes cuando se evalúan miles de observaciones, por lo que se sugiere estudiar como trabajo futuro diversos métodos de visualización para la interpretación de los resultados del análisis de incertidumbre cuando se trabaja con un número elevado de observaciones.

Y, por último, puede resultar de interés aplicar las diferentes metodologías de construcción de ICs propuestas en la Tesis Doctoral para evaluar actividades del ámbito de la educación superior alternativas a la evaluación de la actividad académica del profesorado.

Capítulo 9

Referencias

Abbott, M. y Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review*, 22(1):89-97.

Adriaanse, A. (1993). *Environmental policy performance indicators. A study on the development of indicators for environmental policy in the Netherlands*. The Hague: Sdu Uitgeverij Koninginnegracht.

AENOR (2003). *Sistemas de gestión de la calidad. Guía para la implantación de sistemas de indicadores*, UNE 66175, Comité técnico AEN/CTN 66, Madrid.

Agasisti, T. y Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. *Higher Education*, 59(1):85-103.

Agencia Nacional de Evaluación de la Calidad y Acreditación (2006). *DOCENTIA. Modelo de evaluación*, V. 1.0 – 14/11/2006, Unidad de proyectos de ANECA, Madrid.

Agencia Nacional de Evaluación de la Calidad y Acreditación (2012). *Actividades de evaluación*. Consultado en <http://www.aneca.es/Programas>. [Consulta: 3-10-2012].

Aguillo, I., Ortega, J. y Fernández, M. (2008). Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments. *Higher Education in Europe*, 33(2/3):234-244.

Ahn, T., Charnes, A. y Cooper, W. W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. *Socio-Economic Planning Sciences*, 22(6):259-269.

Ali, A. I. y Seiford, L. (1990). Translation invariance in data envelopment analysis. *Operations Research Letters*, 9(6):403-405.

Allen, R., Athanassopoulos, A., Dyson, R. G. y Thanassoulis, E. (1997). Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. *Annals of Operations Research*, 73(0):13-34.

Altbach, P. (2006). The Dilemmas of Rankings. *International Higher Education*, 42(Winter):2-3.

Annoni, P. (2010). *Transformations*. En: Constructing composite indicators: From theory to practise. 20-21 de mayo. Ispra-Italia.

- Annoni, P., Bruggemann, R. y Saltelli, A. (2012). Random and quasi-random designs in variance-based sensitivity analysis for partially ordered sets. *Reliability Engineering and System Safety*, 107(-):184-189.
- Annoni, P. y Kozovska, K. (2010). *EU Regional Competitiveness Index 2010*, EUR 24346 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Archer, G. y Saltelli, I. (1997). Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(1):99-120.
- Arneson, R. J. (1990). Primary goods reconsidered. *Noûs*, 24(3):429-454.
- ARWU (2003). *Academic Ranking of World Universities*. Consultado en www.shanghairanking.com. [Consulta: 20-11-2011].
- Astin, A. W. (1985). *Achieving Educational Excellence*. San Francisco: Jossey-Bass.
- Athanassopoulos, A. D. y Shale, E. (1997). Assessing the Comparative Efficiency of Higher Education Institutions in the UK by the Means of Data Envelopment Analysis. *Education Economics*, 5(2):117-134.
- Avital, M. y Collopy, F. (2001). Assessing research performance: implications for selection and motivation. *Sprouts: Working Papers on Information Environments*, 1(3):40-61.
- Badasyan, N., Shideler, D. y Silva, S. (2011). Broadband achievement index: Moving beyond availability. *Telecommunications Policy*, 35(11):933-950.
- Bana e Costa, C. y Oliveira, M. (2011). A multicriteria decision analysis model for faculty evaluation. *Omega - The International Journal of Management Science*, 40(4):424-436.
- Bandura, R. y Martin, C. (2006). *A Survey of Composite Indices Measuring Country Performance: 2006 Update*, UNDP/ODS Working Paper, Office of Development Studies, New York.
- Bauer, R. (1996). *Social Indicators*. MIT Press, Cambridge, MA.
- Beasley, J. E. (1995). Determining Teaching and Research Efficiencies. *Journal of the Operational Research Society*, 46(4):441-452.
- Billaut, J. C., Bouyssou, D. y Vincke, P. (2010). Should you believe in the shanghai ranking? *Scientometrics*, 84(1):237-263.
- Borden, V. M. H. y Bottrill, K. V. (1994). Performance indicators : History, definitions and methods. *New Directions for Institutional Research*, 1994(82):5-21.
- Boscarino, J., Figley, C. y Adams, R. (2004). Compassion Fatigue Following the September 11 Terrorist Attacks: A Study of Secondary Trauma among New York City Social Workers. *International Journal of Mental Health Systems*, 6(2):57-66.
- Bousofiane, A., Dyson, R. G. y Thanassoulis, E. (1991). Applied Data Envelopment Analysis. *European Journal of Operational Research*, 52(1):1-15.

- Box, G. E. P. (1979). Robustness in the strategy of scientific model building. En R. L. Launer y G. N. Wilkinson (Eds.) *Robustness in statistics* (pp. 201-236). New York: Academic Press.
- Box, G. E. P. y Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26(2):211-252.
- Boyssou, D., Marchant, T., Pirlot, M. y Tsoukiàs, A. (2006). *Evaluation and Decision Models with Multiple Criteria: Stepping Stones for the Analyst*. New York: Springer.
- Bricall, J. M. (2000). *Informe Universidad 2000*. Consultado en <http://www.oei.es/oeivirt/cap7.pdf>. [Consulta: 9-10-2012].
- Burnham, K. y Anderson, R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag.
- Byrnes, P., Färe, R. y Grosskopf, S. (1984). Measuring productive efficiency: an application to Illinois strip mines. *Management Science*, 30(6):671-681.
- Calderón, C. y Escalera, G. (2008). La evaluación de la docencia ante el reto del EEES. *Educación XXI*, 11(3):237-256.
- Carot, J. M., Bas, M. C., Carrión, A., y García-Gutiérrez, V. R. (2010). *Creating a ranking methodology for measuring university third mission activities*. En: 32nd Annual EAIR Forum - Linking Society and universities. New missions for universities. 1-4 de septiembre. Valencia-España.
- Carot, J. M., Bas, M. C. y García-Gutiérrez, V. R. (2009). Los rankings de universidades. En M. Parellada (Ed) *La contribución de las universidades españolas al desarrollo* (pp. 251-253). Barcelona: Fundación CyD.
- Carot, J.M., Henríquez, P., Haug, G., Mora, J.G., Ristoff, D., Vidal, J., Vila, L., y González, E. (2012). *Sistema Básico de Indicadores para la Educación Superior de América Latina*, Editorial Universitat Politècnica de València, Valencia.
- Carrington, R., Coelli, T. y Prasada, D. S. (2005). The performance of Australian universities: conceptual issues and preliminary results. *Economic Papers*, 24(2):145-163.
- Cerdán, J., Sauquillo, L., Luján, J. y Puente, J. (1992). La evaluación externa de centros. *Aula. De innovación educativa*, 6(-):65-68.
- Charnes, A., Cooper, W., Lewin, A. y Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston: Kluwer Academic Publishers.
- Charnes, A., Cooper, W. y Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429-444.
- Charnes, A., Cooper, W. y Rhodes, E. (1981). Evaluating Program and Managerial Efficiency: An application of Data Envelopment Analysis to Program Follow Through. *Management Science*, 27(6):668-697.
- Cherchye L., Moesen W., Rogge N., Van Puyenbroeck T., Saisana M., Saltelli A., Liska R., y Tarantola S. (2006). *Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index*,

Cherchye, L. (2001). Using Data Envelopment Analysis to Assess Macroeconomic Policy Performance. *Applied Economics*, 33(3):407-416.

Cherchye, L., Lovell, C. A. K., Moesen, W. y Van Puyenbroeck, T. (2005). *One Market, One Number? A Composite Indicator Assessment of EU Internal Market Dynamics*, Discussions Paper Series (DPS) 05.13, Center for Economic Studies, Katholieke Universiteit Leuven.

Cherchye, L., Lovell, C. A. K., Moesen, W. y Puyenbroeck, T. (2007a). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, 51(3):749-779.

Cherchye, L., Moesen, W., Rogge, N. y Van Puyenbroeck, T. (2007b). An introduction to 'Benefit of the Doubt' Composite Indicators. *Social Indicators Research*, 82(1):111-145.

Cherchye, L., Moesen, W., Rogge, N. y Puyenbroeck, T. (2011). Constructing composite indicators with imprecise data: A proposal. *Expert Systems with Applications*, 38(9):10940-10949.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R. y Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society*, 59:239-251.

Cherchye, L., Moesen, W. y Van Puyenbroeck, T. (2004). Legitimately Diverse, Yet Comparable: On Synthesizing Social Inclusion Performance in the EU. *Journal of Common Market Studies*, 42(5):919-955.

Chevalier, S., Choiniere, R. y Bernier, L. (1992). *User guide to 40 Community Health Indicators*. Ottawa: Community Health Division, Health and Welfare Canada.

Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C. (2001). *A robust and scalable clustering algorithm for mixed type attributes in large database environment*. En: KDD '01 Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. 26-29 Agosto. [DOI: 10.1145/502512.502549].

Cobban, A. B. (1988). *The medieval universities: Oxford and Cambridge*. University of California Press, Berkeley, CA.

Cooper, W., Seiford, L. y Tone, K. (2000). *Data Envelopment Analysis: a comprehensive text with models, applications, references and DEA-solver software*. Boston: Kluwer Academic Publishers.

Cooper, W., Seiford, L. y Tone, K. (2004). *Handbook on Data Envelopment Analysis*. Dordrecht: Kluwer Academic Publishers.

Copeland, A. H. (1951). *A reasonable social welfare function*, Mimeographed notes from a seminar on applications of mathematics to the social sciences, Ann Arbor, MI, University of Michigan.

Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. y Jones, D. R. (1992). Quality-of-life Assessment: Can We Keep It simple? *Journal of the Royal Statistical Society A*, 155(3):353-393.

- Cribari-Neto, F., Jensen, M. J. y Novo, A. A. (1999). Research in Econometric Theory: Quantitative and Qualitative Productivity Rankings. *Econometric Theory*, 15(5):719-752.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297-334.
- Cuadras, C. M. (2006). *Modelos estadísticos multivariantes*. Barcelona: Promociones y Publicaciones Universitarias.
- Cukier, R. I., Fortuin, C., Schuler, K. E., Petschek, A. G. y Schaibly, J (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8):3873-3878.
- De la Orden, A., Asensio, I., Carballo, R., Fernández, J., Fuentes A., García, J. M. y Guardia, S. (1997). *Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación*. Consultado en http://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_2.htm. [Consulta: 9-10-2012].
- De Witte, K. y Rogge, N. (2011). Accounting for exogenous influence in performance evaluations of teachers. *Economics of Education Review*, 30(4):641-653.
- De Witte, K., Rogge, N., Cherchye, L. y Puyenbroeck, T. (2013). Economies of scope in research and teaching: A non-parametric investigation. *Omega*, 41(2):305-314.
- Decancq, K. y Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An overview. *Econometric Reviews*, 32(1):7-34.
- Deming, W. E. (1989). *Calidad, productividad y competitividad: la salida de la crisis*. Madrid: Díaz de Santos.
- Dempster, A. P., Laird, N. M. y Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1-38.
- Dempster, A. P. y Rubin, D. B. (1983). Introduction. En W. G. Madow, I. Olkin y D. B. Rubin (Eds.) *Incomplete Data in Sample Surveys: Theory and Bibliographies* (pp. 3-10). New York: Academic Press.
- Despotis, D. K. (2005). A reassessment of the Human Development Index via Data Envelopment Analysis. *Journal of the Operational Research Society*, 56(8):969-980.
- Doksum, K. y Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Annals of Statistics*, 23(5):1443-1473.
- Doménech, R. (1992). Medidas no paramétricas de eficiencia en el sector bancario español. *Revista Española de Economía*, 9(2):171-196.
- Dunteman, G. (1989). *Principal components analysis (Quantitative Applications in the Social Sciences)*. Newbury Park, CA.: SAGE Publications, Inc.
- Dyson, R. G. y Thanassoulis, E. (1988). Reducing weights flexibility in Data Envelopment Analysis. *Journal of the Operational Research Society*, 39(6):563-576.

El Emam, K., Goldenson, D., McCurley, J., y Herbsleb, J. (1998). *Success or Failure? Modeling the Likelihood of Software Process Improvement*, Technical report ISERN-98-15, International Software Engineering Research Network.

El-Mahgary, S. y Lahdelma, R. (1995). Data Envelopment Analysis: Visualizing the results. *European Journal of Operational Research*, 83(3):700-710.

Emery, C. R., Kramer, T. R. y Tian, R. G. (2003). Return to academic standards: A critique of student evaluation of teaching effectiveness. *Quality Assurance in Education*, 11(1):37-46.

European Commission (2001). *Communication from the Commission. Structural Indicators*, Commission of the European Communities, 619 final, 30.10.2001, Brussels.

European Commission Education and Training (2009). *The Bologna Process-Towards the European Higher Education Area*. Consultado en http://ec.europa.eu/education/higher-education/doc1290_en.htm. [Consulta: 4-6-2011].

European Commission-DG ECFIN (2000). *Business Climate Indicator for the Euro Area*. Consultado en http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/studies/bci_presentation_paper.pdf. [Consulta: 15-3-2010].

European Commission-DG ECFIN (2004). *Economic Sentiment Indicator*. Consultado en http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm. [Consulta: 8-2-2010].

European Commission-DG ENTR (2001a). *European Innovation Scoreboard 2001*, CORDIS focus Supplement, 18, European Commission, Luxembourg.

European Commission-DG ENTR (2001b). *Summary Innovation Index*. Consultado en <http://www.proinno-europe.eu/page/summary-innovation-index-0>. [Consulta: 12-4-2010b].

European Commission-DG ENTR (2004). *Composite Indicator on e-business readiness*. Consultado en <http://ec.europa.eu/enterprise/sectors/ict/competitiveness/ebi/>. [Consulta: 5-2-2010].

European Commission-DG MARKT (2001). *Internal Market Scoreboard*. Consultado en http://ec.europa.eu/internal_market/score/index_en.htm. [Consulta: 18-3-2010].

Fagerberg, J. (2001). Europe at the Crossroads: The Challenge from Innovation-based Growth. En D. Archibugi y B.-Å. Lundvall (Eds.) *Globalising Learning Economy* (pp. 45-60). New York: Oxford University Press.

Färe, R., Grosskopf, S., Norris, M. y Zhang, Z. (1994). Productivity Growth, Technical Progress and Efficiency Change in Industrialised Countries. *American Economic Review*, 84(1):66-83.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A*, 120(3):253-290.

Ferrier, G.D., Lovell, C.A.K. (1990). Measuring cost efficiency in banking: econometric and linear programming evidence. *Journal of Econometrics*, 46(1-2):229-245.

- Fleurbaey, M. (1992). Social choice and the indexing dilemma. *Social Choice Welfare*, 29(4):633-648.
- Forni, M., Hallin, M., Lippi, M. y Reichlin, L. (2001). Coincident and Leading Indicators for the Euro Area. *The Economic Journal*, 111(471):C62-C85.
- Foster, J. y Sen, A. (1997). *On Economic Inequality*. Oxford: Clarendon Press.
- Freudenberg, M. (2003). *Composite Indicators of Country Performance: A Critical Assessment*, OECD STI Working Paper, DSTI/DOC(2003)16, OECD Publishing, Paris.
- Gallopín, G. C. (1996). Environmental and sustainability indicators and the concept of situational indicators. A system approach. *Environmental Modeling and Assessment*, 1(3):101-117.
- Gallopín, G. C. (1997). Indicators and their use: Information for decision making. En B. Moldan, S. Billharz y R. Matravers (Eds.) *Sustainability Indicators: A Report on the project on Indicators of Sustainable Development* (pp. 13-27). Chichester: John Wiley & Sons, Inc.
- George, M. D. (1982). Assessing Program Quality. *Designing Academic Program Review. New Directions for Higher Education* (pp. 45-54). San Francisco: Jossey-Bass.
- Ghanima, A. (2010). Ranking using the Copeland score: A comparison with the Hasse Diagram. *Journal of Chemical Information and Modeling*, 50(5):785-791.
- Giménez, V. y Martínez, J. L. (2006). Cost efficiency in the university: A departmental evaluation model. *Economics of Education Review*, 25(5):543-553.
- Golany, B. (1988). A Note on Including Ordinal Relations among Multipliers in Data Envelopment Analysis. *Management Science*, 34(8):1029-1033.
- Gómez, J. M. y Mancebón, M. J. (2012). La evaluación de la eficiencia de las universidades públicas españolas: En busca de una evaluación neutral entre áreas de conocimiento. *Presupuesto y gasto público*, 67(2):43-70.
- Granger, C. W. J. (2001). Macroeconometrics - Past and future. *Journal of Econometrics*, 100(1):17-19.
- Grao, J. and Winter, R. (1999). *Indicadores para la Calidad y Calidad de los indicadores*. En: Seminario Indicadores en la Universidad: información y decisiones. 9-11 de junio.
- Griliches, Z. (1990). Patent Statistics as Economic Indicators: A survey. *Journal of Economic Literature*, 28(4):1661-1707.
- Grupp, H. y Mogege, M. E. (2004). Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*, 33(9):1373-1384.
- Hair, J. F., Anderson, R., Tatham, R. y Black, W. (2007). *Análisis Multivariante*. Madrid: Prentice Hall.
- Hammond, A., Adriaanse, A., Rodenburg, E., Bryant, D., y Woodward, R. (1995). *Environmental Indicators: A Systematic Approach to Measuring and Reporting on*

Environmental Policy Performance in the Context of Sustainable Development, World Resources Institute, Washington.

Harvey, L. y Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education*, 18(1):9-34.

Haug, G. (2011). Introducción. En J. G. Mora, J. M. Vilalta y G. Haug (Eds.) *Garantía de calidad y rendición de cuentas en las Universidades españolas* (pp. 25-35). Madrid: Documento de Trabajo 5.

Henard, F. (2010). *Learning our lesson: Review of Quality Teaching in Higher Education*, doi: 10.1787/9789264079281-en, OECD Publishing,

Hendrik, W., Howard, C., y Maximilian, A. (2008). *Consequences of data error in aggregate indicators: Evidence from the Human Development Index*, Technical report of the Department of Agricultural and Resource economics, UCB, UC Berkeley.

Higher Education Quality Council (1996). *Indicators of programme quality*. London: Higher Education Quality Council.

Homma, T. y Saltelli, A. (1996). Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, 52(1):1-17.

Horta, I. M., Camanho, A. S. y Moreira da Costa, J. (2012). Performance assessment of construction companies: A study of factors promoting financial soundness and innovation in the industry. *International Journal of Production Economics*, 137(1):84-93.

Huggins, R. (2003). Creating a UK Competitiveness Index: Regional and Local Benchmarking. *Regional Studies*, 37(1):89-96.

Institute for the Protection and Security of Citizen (2013). *SS-ANOVA-R*. Consultado en <http://ipsc.jrc.ec.europa.eu/?id=790>. [Consulta: 1-2-2012].

Instituto de Evaluación (2011a). *Panorama de la educación: Indicadores de la OCDE 2011, informe español*. Ministerio de Educación. Consultado en <http://www.educacion.gob.es/dctm/evaluacion/indicadores-educativos/panorama-2011.pdf?documentId=0901e72b80eb8ea5>. [Consulta: 15-12-2011a].

Instituto de Evaluación (2011b). *Sistema estatal de indicadores de la educación. Edición 2011*. Ministerio de Educación. Consultado en <http://www.educacion.gob.es/dctm/evaluacion/indicadores-educativos/seie-2011.pdf?documentId=0901e72b810b3cc3>. [Consulta: 12-12-2011b].

Ishikawa, K. (1986). *¿Qué es control total de la calidad?* Bogotá: Norma.

Jacobs, R., Smith, P., y Goddard, M. (2004). *Measuring Performance: An Examination of Composite Performance Indicators*, CHE Technical Paper Series 29, Centre for Health Economics, University of York.

Jansen, R. C. (1994). Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, 138(3):871-881.

- Jencks, S. F., Huff, E. D. y Cuerton, T. (2003). Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998-1999 to 2000-2001. *The Journal of the American Medical Association*, 289(3):305-312.
- John, J. A. y Draper, N. R. (1980). An Alternative Family of Transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):190-197.
- Johnes, G. y Johnes, J. (2009). Higher education institutions' cost and efficiency: Taking the decomposition a further step. *Economics of Education Review*, 28(1):107-113.
- Johnes, J. (2006). Data Envelopment Analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3):273-288.
- Johnes, J. y Johnes, G. (1995). Research Funding and Performance in U.K. University Departments of Economics: A Frontier Analysis. *Economics of Education Review*, 14(3):301-314.
- Joint Research Centre (2010). *An information server on composite indicators and ranking systems*. Consultado en <http://composite-indicators.jrc.ec.europa.eu/>. [Consulta: 15-5-2010].
- Juran, J. M. (1990). *Juran y la planificación de la calidad*. Madrid: Díaz de Santos.
- Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39(1):31-36.
- Kells, H. R. (1993). *The Development of Performance Indicators for Higher Education*. (2nd ed.). Paris: OECD Publishing.
- Kerr, R. (1982). *Psychomotor learning*. New York: Saunders College.
- Kleinknecht, A., Van Montfort, K. y Brouwer, E. (2002). The non-trivial choice between Innovation Indicators. *Economic Innovation and New Technologies*, 11(2):109-121.
- Koopmans, T. C. (1947). Measurement Without Theory. *The Review of Economic Statistics*, 29(3):161-172.
- Kucherenko, S., Tarantola, S., Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937-946.
- Lawrence, B., Weathersby, G., and Patterson, V. (1970). *The Outputs of Higher Education: Their Identification, Measurement and Evaluation*. En: Boulder, Colo.: Western Interstate Commission for Higher Education.
- Leamer, E. (1983). Let's take the con out of econometrics. *The American Economic Review*, 73(1):31-43.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404):1198-1202.
- Little, R. J. A. y Rubin, D. B. (1987). *Statistical Analysis with Missing Data, Second Edition*. New Jersey: John Wiley & Sons, Inc.

- Lovell, C. A. K., Pastor, J. T. y Turner, J. A. (1995). Measuring Macroeconomic Performance in the OECD: A comparison of European and non-European countries. *European Journal of Operational Research*, 87(3):507-518.
- Maassen, G. H. (1997). An application of structural equation modeling in public opinion research: Conceptualizing public and opinions. *International Journal of Public Opinion Research*, 9(2):146-169.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En L. M. Le Cam y J. Neyman (Eds.) *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (pp. 281-297). Berkeley: University of California Press.
- Manly, B. F. J. (1976). Exponential data transformations. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 25(1):37-42.
- Marsh, H.W. y Roche, L. (1994). *The Use of Students' Evaluation of University Teaching to Improve Teaching Effectiveness*, Australian Department of Employment, Education, and Training, Canberra.
- Martín, R. (2008). La medición de la eficiencia universitaria: Una aplicación del Análisis Envolverte de Datos. *Formación Universitaria*, 1(2):17-26.
- Martínez, M. (2000). Análisis de la eficiencia productiva de las instituciones de educación superior. *Papeles de Economía Española*, 86(-):179-191.
- McMillan, M. y Datta, D. (1998). The relative efficiencies of Canadian universities: A DEA perspective. *Canadian Public Policy*, 24(4):485-511.
- Medina, F. y Galván, M. (2007). *Imputación de datos: teoría y práctica*, CEPAL - Serie Estudios estadísticos y prospectivos N° 54, División de Estadística y Proyecciones Económicas, Santiago de Chile.
- Medina, R. (2005). Misiones y funciones de la Universidad en el Espacio Europeo de Educación Superior. *Revista Española de Pedagogía*, 63(230):17-42.
- Melyn, W. y Moesen, W. (1991). *Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available*, Public Economic research Paper 17, Center for Economic Studies, KU Leuven.
- Miller, R. I. (1979). *The Assessment of College Performance*. San Francisco: Jossey-Bass.
- Mora, J. G. (1991a). La evaluación institucional: una perspectiva general. En M. De Miguel, J. G. Mora y S. Rodríguez (Eds.) *La Evaluación de las Instituciones Universitarias* (pp. 73-92). Madrid: Consejo de Universidades. Secretaria General.
- Mora, J.G. (1991b). *Calidad y rendimiento de las instituciones universitarias*, Consejo de Universidades. Secretaria General, Madrid.
- Mora, J. G. (1998). La evaluación institucional de la Universidad. *Revista de Educación*, 315(1):29-44.

- Mora, J. G. (1999a). *Indicadores y decisiones de las universidades*. En: Seminario Indicadores en la Universidad: información y decisiones. 9-11 de junio.
- Mora, J. G. (1999b). La Universidad: una empresa al servicio de la sociedad del conocimiento. *Cuadernos IRC*, 1(-):41-55.
- Mora, J. G. (2011). Garantía de calidad como responsabilidad social de universidades y gobiernos. En J. G. Mora, J. M. Vilalta y G. Haug (Eds.) *Garantía de calidad y rendición de cuentas en las Universidades españolas* (pp. 25-35). Madrid: Documento de Trabajo 5.
- Mora, J. G. y Vilalta, J. M. (2011). Summing up. En J. G. Mora, J. M. Vilalta y G. Haug (Eds.) *Garantía de calidad y rendición de cuentas en las Universidades españolas* (pp. 25-35). Madrid: Documento de Trabajo 5.
- Munda, G. (1995). *Multicriteria evaluation in a fuzzy environment: theory and applications in ecological economics*. Heidelberg: Physica-Verlag.
- Munda, G. (2005). Multiple Criteria Decision Analysis and Sustainable Development. En J. Figueira, S. Greco y M. Ehrgott (Eds.) *Multiple Criteria Decision Analysis. State of the art surveys* (pp. 953-986). New York: Springer International Series in Operations Research and Management Science.
- Munda, G. (2008). *Social Multi-Criteria Evaluation for a Sustainable Economy*. Berlin Heidelberg: Springer-Verlag.
- Munda, G. y Nardo, M. (2005). *Constructing Consistent Composite Indicators: the Issue of Weights*, EUR 21834 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Munda, G. y Nardo, M. (2009). Non-compensatory/non-linear composite indicators for ranking countries: a defensible setting. *Applied Economics*, 41(12):1513-1523.
- Nardo, M., Loi, M., Rosati, R., y Manca, A. (2011). *The Consumer Empowerment Index*, EUR 24791 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Nardo, M., Saisana, M., Saltelli, A., y Tarantola, S. (2005). *Tools for Composite Indicators Building*, EUR 21682 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., y Giovannini, E. (2008). *Handbook on constructing composite indicators: methodology and user guide*, OECD Statistics Working Paper, STD/DOC (2005)3, OECD Publishing, Paris.
- National Institute of Science and Technology Policy (2004). *Science and Technology Indicators: 2004. A Systematic Analysis of Science and Technology Activities in Japan*, REPORT No. 73, NISTEP, Japan.
- Neave, G. (1991). Sobre el desarrollo de la calidad, la eficiencia y la iniciativa: Un resumen de las tendencias recientes de la enseñanza superior en europa occidental. En M. De Miguel, J. G. Mora y S. Rodríguez (Eds.) *La Evaluación de las Instituciones Universitarias* (pp. 17-38). Madrid: Consejo de Universidades. Secretaría General.
- Neggers, J. y Kim, H. (1998). *Basic posets*. Singapore: World Scientific.

- Ng, Y. y Li, S. (2000). Measuring the research performance of chinese higher education institutions: An application of Data Envelopment Analysis. *Education Economics*, 8(2):139-156.
- Nicoletti, G., Scarpetta, S., y Boylaud, O. (2000). *Summary indicators of product market regulation with an extension to employment protection legislation*, OECD Economics Department Working Papers ECO/WKP(99)18, OECD Publishing, Paris.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Olivera, A. (1997). Indicadores en Geografía Social. *Estudios geográficos*, 58(229):689-710.
- Ott, W. (1978). *Environmental Indices: Theory and Practise*. Michigan: Ann Arbor Science.
- Parker, J. (1991). *Environmental reporting and environmental indices*. Cambridge: Churchill College.
- Paruolo, P., Saltelli, A. y Saisana, M. (2013). Ratings and rankings: Voodoo or Science? *Journal of the Royal Statistical Society, Series A*, 176(3):609-634.
- Pastor, J.T. (1996). Translation invariance in data envelopment analysis: a generalization. *Annals of Operations Research*, 66(2):91-102.
- Pastor, J.T. y Lovell, C.A.K. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters*, 18(3):147-151.
- Pearson, K. (1905). *On the General Theory of Skew Correlation and Non-linear Regression*, Volume XIV of Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs, Dulau & Co, Ltd., London.
- Pedraja-Chaparro, F., Salinas-Jimenez, J. y Smith, P. (1997). On the role of weight restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis*, 8(2):215-230.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw Hill.
- Pérez, C. (2005). *Métodos estadísticos avanzados con SPSS*. Madrid: Thomson Paraninfo.
- Pérez, R. (2000). La calidad de la educación. En R. Pérez, F. López, M. D. Peralta y P. Muncio (Eds.) *Hacia una educación de calidad. Gestión, instrumentos y evaluación* (pp. 13-44). Madrid: Narcea.
- Pina, V. y Torres, L. (1995). Evaluación del rendimiento de los departamentos de contabilidad de las universidades españolas. *Hacienda Pública Española*, 135(-):183-190.
- Pirsing, R. M. (1974). *Zen and the art of motorcycle maintenance*. New York: William Morrow & Co.
- Podinovski, V. V. (1999). Side effects of absolute weight bounds in DEA models. *European Journal of Operational Research*, 115(3):583-595.
- Pozo, C. (2010). *Análisis de indicadores de evaluación de la calidad docente en la Universidad Pública Española*, Diseño de una guía de buenas prácticas docentes, Programa de Estudios y Análisis, EA2009-0125, Universidad de Almería.

- Pritchard, R. D., Watson, M. D., Kelly, K. y Paquin, A. R. (1998). *Helping Teachers Teach Well: A New System for Measuring and Improving Teaching Effectiveness in Higher Education*. San Francisco: The New Lexington Press.
- Programa de las Naciones Unidas para el Desarrollo (2010). *Informe sobre Desarrollo Humano 2010*, La verdadera Riqueza de las naciones: caminos al desarrollo humano, Mundi-Prensa, Madrid.
- Pulido, A. y Pérez, J. (2003). *Propuesta metodológica para la evaluación de la calidad docente e investigadora: planteamiento y experimentación*, Cuadernos del Fondo de Investigación R. Stone nº 8, Instituto L.R. Klein de la UAM, Madrid.
- QS (2005). *QS World University Rankings*. Consultado en <http://www.topuniversities.com/>. [Consulta: 21-9-2011].
- Quintanilla, M. A. (1998). En pos de la calidad: notas sobre una nueva frontera para el sistema universitario español. *Revista de Educación*, 315(1):85-95.
- Rabitz, H., Alis, Ö. F., Shorter, J. y Shim, K. (1999). Efficient input-output model representations. *Computer Physics Communications*, 117(1-2):11-20.
- Ramón, N., Ruiz, J. L. y Sirvent, I. (2010). Using Data Envelopment Analysis to assess effectiveness of the process at the university with performance indicators of quality. *International Journal of Operations and Quantitative Management*, 16(4):87-103.
- Ratto, M., Pagano, A. y Young, P. (2007). State Dependent Parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177(11):863-876.
- Rauhvargers, A. (2011). *Global university rankings and their impact*, EUA Report on rankings 2011, European University Association, Brussels.
- Rawls, J. (1999). *A Theory of Justice*. Oxford: Oxford University Press.
- Robins, J. M., Rotnitzky, A. y Zhao, L. P. (1994). Estimation of Regression Coefficients when some Regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846-866.
- Rogge, N. (2010). Granting teachers the "benefit of the doubt" in performance evaluations". *International Journal of Education Management*, 25(6):590-614.
- Rogge, N. (2011). *Evaluating teaching and research performances: a non-parametric 'Benefit of the Doubt' approach*, Tesis Doctoral, Faculty of Business and Economics, Katholieke Universiteit Leuven.
- Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. *Ecological Indicators*, 23(-):143-154.
- Roll, Y., Cook, W. D. y Golany, B. (1991). Controlling factors weights in Data Envelopment Analysis. *IIE Transactions*, 23(1):2-9.
- Roll, Y. y Golany, B. (1993). Alternate methods of treating factor weights in DEA. *Omega*, 21(1):99-109.

- Roy, B. (1996). *Multicriteria methodology for decision analysis*. Dordrecht: Kluwer Academic Publishers.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New Jersey: John Wiley & Sons, Inc.
- Saaty, T. L. (1980). *The Analytic Hierarchy Process: planning, priority setting, resource allocation*. New York: McGraw-Hill.
- Sacks, J., Welch, W., Mitchell, T. y Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409-435.
- Saisana, M. (2008). *2007 Composite Learning Index: Robustness Issues and Critical Assessment*, EUR 23274 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Saisana, M. (2010). *Imputation of missing data*. En: Constructing composite indicators: From theory to practise. 20-21 Mayo.
- Saisana, M., Annoni, P., y Nardo, M. (2009). *A robust model to measure governance in African countries*, EUR 23773 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Saisana, M. y D'Hombres, B. (2008). *Higher Education Rankings: Robustness Issues and Critical Assessment*, EUR 23487 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Saisana, M. y Saltelli, A. (2008). *Sensitivity analysis of the 2008 Environmental Performance Index*, EUR 23485 EN, JRC Scientific and Technical Reports (EUR collection), Italy.
- Saisana, M., Saltelli, A. y Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A*, 168(2):1-17.
- Saisana, M. y Tarantola, S. (2002). *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*, EUR Report 20408 EN, European Commission, JRC, Institute for the Protection and Security of the Citizen, Ispra, Italy.
- Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280-297.
- Saltelli, A. (2007). Composite Indicators between analysis and advocacy. *Social Indicators Research*, 81(1):65-77.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. y Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259-270.
- Saltelli, A., Chan, K. P. S. y Scott, M. (2000). *Sensitivity analysis*. Chichester: John Wiley & Sons, Ltd.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. y Tarantola, S. (2008). *Global Sensitivity Analysis. The primer*. Chichester: John Wiley & Sons, Ltd.

- Saltelli, A. y Tarantola, S. (2002). On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association*, 97(459):702-709.
- Saltelli, A., Tarantola, S., Campolongo, F. y Ratto, M. (2004). *Sensitivity Analysis in practice: a guide to assessing scientific models*. New York: John Wiley & Sons, Ltd.
- Saltelli, A., Tarantola, S. y Chan, K. P. S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1):39-56.
- Schmidt, P. (1986). Frontier Production Function. *Econometric Reviews*, 4(2):289-328.
- Schuschny, A.R. y Soto, H. (2009). *Guía metodológica: Diseño de indicadores compuestos de desarrollo sostenible*, CEPAL-Colección Documentos de proyectos, Sede de las Naciones Unidas, Nueva York.
- Schwartzman, S. (1988). *La calidad de la Educación Superior en América Latina. Texto preparado para el Seminario sobre la Eficiencia y la Calidad de la Educación Superior en América Latina*, Basílica. Consultado en <http://www.schwartzman.org.br/simon/calidad.htm>. [Consulta: 9-7-2012].
- SCImago (2007). *SJR - SCImago Journal & Country Rank*. Consultado en <http://www.scimagojr.com>. [Consulta: 18-1-2012].
- Seiford, L. (1994). A DEA Bibliography (1978-1992). En A. Charnes, W. Cooper, A. Lewin y L. Seiford (Eds.) *Data Envelopment Analysis: Theory, Methodology and Applications* Boston: Kluwer Academic Publishers.
- Sharpe, A. (2004). *Literature Review of Frameworks for Macro-indicators*, CSLS Research Report 2004-03, Centre for the Study of Living Standards, Ottawa.
- Simlab (2011). *Software package for uncertainty and sensitivity analysis*. Ver. 3.2.10.
- Sizer, J., Spee, A. y Bormans, R. (1992). The role of performance indicators in higher education. *Higher Education*, 24(2):133-155.
- Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1(4):407-414.
- Sobol', I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S. y Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering and System Safety*, 92(7):957-960.
- Späth, H. (1980). *Cluster Analysis Algorithms*. Chichester, England: Ellis Horwood.
- Stiglitz, J. E., Sen, A. y Fitoussi, J. P. (2010). *Mismeasuring Our Lives: Why GDP Doesn't Add Up*. New York: The New Press.
- Tarantola, S., Gatelli, D. y Mara, T. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering and System Safety*, 91(6):717-727.
- Tarantola, S., Liska, R., Saltelli, A., Leapman, N., y Grant, C. (2004). *The Internal Market Index 2004*, EUR 21274 EN, JRC Scientific and Technical Reports (EUR collection), Italy.

Tarantola, S., Nardo, M., Saisana, M. y Gatelli, D. (2006). A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator. *Reliability Engineering and System Safety*, 91(10-11):1135-1141.

Tarantola, S., Saisana, M., Saltelli, A., Schmiedel, F., y Leapman, N. (2002). *Statistical Techniques and Participatory Approaches for the composition of the European Internal Market Index 1992-2001*, EUR 20547 EN, JRC Scientific and Technical Reports (EUR collection), Italy.

Thanassoulis, E. y Allen, R. (1998). Simulating weights restrictions in Data Envelopment Analysis by means of unobserved DMUs. *Management Science*, 44(4):586-594.

Thanassoulis, E., Portela, M. C. y Allen, R. (2004). Incorporating value judgments in DEA. En W. Cooper, L. Seiford y J. Zhu (Eds.) *Handbook on Data Envelopment Analysis* (pp. 99-138). Dordrecht: Kluwer Academic Publishers.

The Bologna Declaration (1999). *Joint declaration of the European Ministers of Education*. Consultado en http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/MDC/BOLOGNA_DECLARATION1.pdf. [Consulta: 9-10-2012].

Thompson, R., Dharmapala, P. y Thrall, R. (1993). Importance for DEA of Zeros in Data, Multipliers, and Solutions. *Journal of Productivity Analysis*, 4(4):379-390.

Thompson, R., Langemeier, L., Lee, C.-T., Lee, E. y Thrall, R. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, 46(1-2):93-108.

Tomkins, C. y Green, R. (1988). An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. *Financial Accountability and Management*, 4(2):147-164.

Torres-Salinas, D., Moreno-Torres, J., Robinson, N., Delgado, E. y Herrera, F. (2011). *Rankings ISI de las Universidades Españolas según Campos y Disciplinas Científicas (2º ed)*. Consultado en <http://rankinguniversidades.es>. [Consulta: 18-1-2012].

United Nations Development Programme (2001). *Human Development Report 2001*, Making new technologies work for human development, Oxford University Press, New York.

Universitat Politècnica de València (2007). *Plan Estratégico UPV (2007-2014): Resumen Ejecutivo*. Consultado en <http://www.upv.es/noticias-upv/documentos/2714-es.pdf>. [Consulta: 4-5-2010].

Universitat Politècnica de València (2008). *Reglamento para la evaluación de la actividad de investigación, desarrollo e innovación en la UPV*. Consultado en <http://www.upv.es/entidades/VI/infoweb/vi/info/U0451850.pdf>. [Consulta: 7-2-2013].

Universitat Politècnica de València (2012). *Propuesta Índice de Actividad Académica del Profesorado de la UPV*. Consultado en <http://www.upv.es/entidades/VCEAA/info/U0576427.pdf>. [Consulta: 7-2-2013].

Uriel, E. (1995). *Análisis de datos. Series temporales y Análisis multivariante*. Madrid: Editorial AC.

- Van Vught, F. A., Kaiser, F., File, J. M., Gaethgens, C., Peter, R. y Westerheijden, D. F. (2012). *U-Map. The European Classification of Higher Education Institutions*. Consultado en http://www.u-map.org/U-MAP_report.pdf. [Consulta: 16-7-2012].
- Van Vught, F. A. y Ziegele, F. (2011). *Multi-dimensional Global ranking of Universities; a feasibility project*. Consultado en http://ec.europa.eu/education/higher-education/doc/multirank_en.pdf. [Consulta: 9-7-2012].
- Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV (2009). *Informe de implantación del modelo de evaluación de la actividad docente de la UPV*. Consultado en <http://www.upv.es/entidades/VCEAA/info/U0576429.pdf>. [Consulta: 2-2-2011].
- Vicerrectorado de Calidad y Evaluación de la Actividad Académica de la UPV (2010). *Manual de evaluación de la actividad docente del profesorado de la UPV*. Consultado en <http://www.upv.es/upl/U0515160.pdf>. [Consulta: 6-12-2010].
- Vichi, M. y Kiers, H. (2001). Factorial k-means analysis for two-way data. *Computational Statistics and Data Analysis*, 37(1):49-64.
- Vidal, J. (1999). *Indicadores de Rendimiento para las Universidades Españolas: necesidad y disponibilidad*. En: Seminario Indicadores en la Universidad: información y decisiones. 9-11 de junio.
- Vincke, P. (1992). *Multicriteria Decision-Aid*. New York: Wiley.
- Vroeiijstijn, A. I. (1995). *Improvement and accountability: Navigating between Scylla and Charybdis*. London: Jessica Kingsley.
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236-244.
- Warning, S. (2004). Performance Differences in German Higher Education: Empirical Analysis of Strategic Groups. *Review of Industrial Organization*, 24(4):393-408.
- Westerheijden, D. F. (1990). Peers, performance and power. En F. Goedegeebure, P. Maasen y D. F. Westerheijden (Eds.) *Peer review and performance indicators* (pp. 183-207). Utrecht: Lema B.V.
- Wilson, J. W. y Jones, C. P. (2002). An Analysis of the S&P 500 Index and Cowles's Extensions: Price Indexes and Stock Returns, 1870-1999. *Journal of Business*, 75(3):505-533.
- Wong, Y.-H. B. y Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 41(9):829-835.
- World Economic Forum (2001). *2001 Environmental Sustainability index*, Annual Meeting 2001, Yale Center for Environmental Law and Policy, Davos, Switzerland.
- Xu, C. y Gertner, G. Z. (2008). A general first-order global sensitivity analysis method. *Reliability Engineering and System Safety*, 93(7):1060-1071.
- Yorke, M. (1995). Siamese Twins? Performance indicators in the service of accountability and enhancement. *Quality in Higher Education*, 1(1):13-30.

Young P. (2001). The identification and estimation of nonlinear stochastic systems. En A. I. Mees (Ed) *Nonlinear Dynamics and Statistics* (pp. 127-163). Boston: Birkhauser.

Young, P. (2011). *Recursive Estimation and Time-Series Analysis: An introduction for the Student and Practitioner*. New York: Springer.

Zahedi, F. (1986). The Analytic Hierarchy Process: A Survey of the Method and Its Applications. *Interfaces*, 16(4):96-108.

Capítulo 10

Apéndices

Anexo A. Análisis Envolvente de Datos. Modelo DEA-CCR (Charnes, Cooper y Rhodes)

El Análisis Envolvente de Datos (DEA) es una técnica de programación lineal no paramétrica que evalúa la eficiencia relativa de un conjunto de unidades productivas homogéneas¹⁹ (referidas en la literatura DEA como unidades de decisión – *Decision Making Units* or DMUs²⁰). Este procedimiento se basa en la obtención de una frontera eficiente a partir de un conjunto de unidades de decisión, sin necesidad de asumir ninguna “forma” funcional entre los factores input y output del modelo.

En este sentido, el modelo DEA trata de optimizar la medida de eficiencia de cada unidad analizada mediante la maximización del ratio entre la suma ponderada de los inputs y la suma ponderada de los factores output de cada unidad de decisión y utiliza modelos de optimización lineal para calcular las ponderaciones o precios. De esta forma, el DEA compara la eficiencia relativa de un conjunto de unidades que producen factores output similares a partir de una serie de factores input comunes. Además, DEA determina las ponderaciones de forma endógena para cada input y output que son, para cada unidad de análisis, aquellas que maximizan su ratio de productividad y, por tanto, las que más la benefician en el análisis de su eficiencia.

La formulación del modelo básico DEA-CCR se describe como sigue:

$$\max_{(w_{c,1}, \dots, w_{c,q}, v_{c,1}, \dots, v_{c,s})} h_0(w, v) = \frac{\sum_{i=1}^q w_{c,i} y_{c,i}}{\sum_{u=1}^s v_{c,u} x_{c,u}} \quad (1.1)$$

s.a

$$\frac{\sum_{i=1}^q w_{c,i} y_{j,i}}{\sum_{u=1}^s v_{c,u} x_{j,u}} \leq 1 \quad j = 1, \dots, c, \dots, n \quad (1.1a)$$

$$w_{c,i}, v_{c,u} \geq \varepsilon > 0 \quad \forall i = 1, \dots, q; \forall u = 1, \dots, s \quad (1.1b)$$

donde:

¹⁹ Homogéneas en el sentido de producir el mismo tipo de outputs y tener los mismos objetivos

²⁰ En el contexto de la educación superior, las unidades de decisión DMUs pueden ser, según el estudio, universidades, facultades, grupos de investigación, profesores, estudiantes, etc.

h_0 es la función objetivo

q es el número de factores output

s es el número de factores input

$w_{c,i}$ es el peso del factor output i de la unidad c

$v_{c,u}$ es el peso del factor input u de la unidad c

$y_{c,i}$ es la cantidad del factor output i de la unidad c

$x_{c,u}$ es la cantidad del factor input u de la unidad c

En el modelo básico DEA los pesos únicamente tienen que satisfacer dos restricciones:

- **Normalización.** Los índices o ratios de eficiencia de todas las unidades de decisión calculados mediante el conjunto de pesos óptimo de la unidad bajo evaluación no pueden superar la unidad (1.1a)
- **No negatividad.** Los pesos estimados no pueden ser negativos (1.1b)

En particular, para cada DMU_j ($j = 1, \dots, n$), se tienen observaciones en las que se consume una cantidad $x_{j,u}$ ($u = 1, \dots, s$) y se produce una cantidad $y_{j,u}$ ($u = 1, \dots, s$), asumiendo que $x_{j,u} \geq 0$ y $y_{j,u} \geq 0$ y que cada DMU tiene al menos un valor para el factor input y output positivo. La formulación del modelo básico muestra el objetivo del método DEA: estimar para cada DMU el conjunto de pesos de los factores input $v_{c,u}^*$ ($u = 1, \dots, s$), y de los factores output $w_{c,i}^*$ ($i = 1, \dots, q$) de tal forma que maximice su *score* de eficiencia relativo a los *score* del resto de DMUs del conjunto de datos.

Nótese que la “forma-ratio” del modelo básico de la técnica DEA produce infinitas soluciones. Por ejemplo si $(v_{c,u}^*, w_{c,i}^*)$ es una solución del modelo, entonces $(\alpha v_{c,u}^*, \alpha w_{c,i}^*)$ también son soluciones óptimas del modelo $\alpha > 0$. Para seleccionar una solución óptima al problema, Charnes, Cooper y Rhodes (1978) propusieron una transformación del modelo básico en el que se incluyen las restricciones $\sum_{i=1}^q w_{c,i} y_{c,i} = 1$ y $\sum_{u=1}^s v_{c,u} x_{c,u} = 1$ en las dos orientaciones diferenciadas para la estimación de la eficiencia:

Orientación input

$$z_c = \max_{(w_{c,1}, \dots, w_{c,q}, v_{c,1}, \dots, v_{c,s})} \sum_{i=1}^q w_{c,i} y_{c,i} \quad (1.2a)$$

s.a

$$\sum_{i=1}^q w_{c,i} y_{c,i} - \sum_{u=1}^s v_{c,u} x_{j,u} \leq 0 \quad j = 1, \dots, c, \dots, n$$

$$\sum_{u=1}^s v_{c,u} x_{c,u} = 1$$

$$w_{c,i}, v_{c,u} \geq \varepsilon > 0 \quad \forall i = 1, \dots, q; \forall u = 1, \dots, s$$

Orientación output

$$h_c = \min_{(w_{c,1}, \dots, w_{c,q}, v_{c,1}, \dots, v_{c,s})} \sum_{u=1}^s v_{c,u} x_{c,u} \quad (1.2b)$$

s.a

$$\sum_{u=1}^s v_{c,u} x_{j,u} - \sum_{i=1}^q w_{c,i} y_{j,i} \geq 0 \quad j = 1, \dots, c, \dots, n$$

$$\sum_{i=1}^q w_{c,i} y_{c,i} = 1$$

$$w_{c,i}, v_{c,u} \geq \varepsilon > 0 \quad \forall i = 1, \dots, q; \forall u = 1, \dots, s$$

El primer enfoque consiste en minimizar el input manteniendo constante el output y se trata de la versión de orientación input del modelo CCR (Charnes, Cooper y Rhodes, 1978), mientras que el segundo enfoque tiene como objetivo obtener el máximo output manteniendo los factores

input constantes y es la versión de orientación output de ese mismo modelo. En la literatura existen otros modelos DEA que no se van a desarrollar en este trabajo. Una discusión sobre los distintos modelos se puede encontrar en (Charnes et al., 1994).

Teniendo en cuenta las orientaciones definidas, una unidad será considerada eficiente si y sólo si, no es posible incrementar las cantidades de output manteniendo fijas las cantidades de inputs utilizadas, ni es posible disminuir las cantidades de inputs empleadas sin alterar las cantidades de outputs obtenidas (Charnes, Cooper y Rhodes, 1981).

El objetivo de los dos modelos anteriores, independientemente de la orientación, es estimar el conjunto de pesos que optimiza el score de cada DMU_c con respecto al resto de DMUs del conjunto de datos. El modelo DEA asigna pesos elevados a los inputs que tienen un nivel bajo de consumo y a los outputs que tienen un nivel alto de producción ya que dichos niveles aumentan la eficiencia de la DMU. El caso contrario se produce para inputs con un alto nivel de consumo y outputs con bajo nivel de producción ya que en este caso, estos niveles perjudican la eficiencia de la DMU.

Aplicando el teorema de dualidad de programación lineal se resuelven los dos enfoques anteriores y se estiman los conjuntos de pesos correspondientes para cada observación. Los valores de las ponderaciones pueden ser distintos para cada observación, ya que se seleccionan de tal forma que sean lo más favorable posible para cada una de ellas.

En la literatura se han desarrollado diferentes enfoques de la metodología DEA basados en el enfoque DEA-CCR en el que se tratan inputs y outputs categóricos o se investigan los cambios de eficiencia de las DMU en diferentes periodos de tiempo, etc. Una revisión completa de los métodos existentes se puede encontrar en el libro titulado "Handbook on Data Envelopment Analysis" (Cooper, Seiford y Tone, 2004).

Anexo B. Resultados del análisis de sensibilidad global

Escenario E2

$$IC_{E2} = \frac{1}{3} * \left(\frac{1}{4} * IAD_{FORMA} + \frac{1}{4} * IAD_{GUIA} + \frac{1}{4} * IAD_{INFORMA} + \frac{1}{4} * IAD_{MAT} \right) + \frac{1}{3} * \left(\frac{1}{4} * IAD_{DOC} + \frac{1}{4} * IAD_{TUT} + \frac{1}{4} * IAD_{EVA} + \frac{1}{4} * IAD_{OTRAS} \right) + \frac{1}{3} * \left(\frac{1}{3} * K_1 + \frac{1}{3} * K_2 + \frac{1}{3} * K_4 \right)$$

E2	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,159	15,9	12,0	-3,9
IAD_{GUIA}	0,076	7,6	8,0	0,4
$IAD_{INFORMA}$	0,106	10,6	5,0	-5,6
IAD_{MAT}	0,145	14,5	10,0	-4,5
IAD_{DOC}	0,057	5,7	15,0	9,3
IAD_{TUT}	0,091	9,1	7,0	-2,1
IAD_{EVA}	0,022	2,2	8,0	5,8
IAD_{OTRAS}	0,094	9,4	10,0	0,6
K_1	0,171	17,1	18,0	0,9
K_2	0,037	3,7	5,0	1,3
K_3	0,000	0,0	1,0	1,0
K_4	0,041	4,1	1,0	-3,1
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.1. Análisis de sensibilidad global aplicado al escenario E2

Escenario E3

$$IC_{E3} = (IAD_{FORMA}^{0,343} * IAD_{GUIA}^{0,228} * IAD_{INFORMA}^{0,143} * IAD_{MAT}^{0,286})^{0,35} * (IAD_{DOC}^{0,375} * IAD_{TUT}^{0,175} * IAD_{EVA}^{0,2} * IAD_{OTRAS}^{0,25})^{0,4} * (K_1^{0,76} * K_2^{0,2} * K_4^{0,04})^{0,25}$$

E3	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,160	16,0	12,0	-4,0
IAD_{GUIA}	0,099	9,9	8,0	-1,9
$IAD_{INFORMA}$	0,045	4,5	5,0	0,5
IAD_{MAT}	0,177	17,7	10,0	-7,7
IAD_{DOC}	0,075	7,5	15,0	7,5
IAD_{TUT}	0,080	8,0	7,0	-1,0
IAD_{EVA}	0,027	2,7	8,0	5,3
IAD_{OTRAS}	0,125	12,5	10,0	-2,5
K_1	0,193	19,3	18,0	-1,3
K_2	0,007	0,7	5,0	4,3
K_3	0,000	0,0	1,0	1,0
K_4	0,013	1,3	1,0	-0,3
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.2. Análisis de sensibilidad global aplicado al escenario E3

Escenario E4

$$IC_{E4} = \frac{1}{11} * IAD_{FORMA} + \frac{1}{11} * IAD_{GUIA} + \frac{1}{11} * IAD_{INFORMA} + \frac{1}{11} * IAD_{MAT} + \frac{1}{11} * IAD_{DOC} + \frac{1}{11} * IAD_{TUT} + \frac{1}{11} * IAD_{EVA} + \frac{1}{11} * IAD_{OTRAS} + \frac{1}{11} * K_1 + \frac{1}{11} * K_2 + \frac{1}{11} * K_4$$

E4	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,111	11,1	12,0	0,9
IAD_{GUIA}	0,097	9,7	8,0	-1,7
$IAD_{INFORMA}$	0,066	6,6	5,0	-1,6
IAD_{MAT}	0,090	9,0	10,0	1,0
IAD_{DOC}	0,124	12,4	15,0	2,6
IAD_{TUT}	0,141	14,1	7,0	-7,1
IAD_{EVA}	0,065	6,5	8,0	1,5
IAD_{OTRAS}	0,103	10,3	10,0	-0,3
K_1	0,118	11,8	18,0	6,2
K_2	0,029	2,9	5,0	2,1
K_3	0,000	0,0	1,0	1,0
K_4	0,056	5,6	1,0	-4,6
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.3. Análisis de sensibilidad global aplicado al escenario E4

Escenario E5

$$IC_{E5} = \frac{1}{11} * IAD_{FORMA} + \frac{1}{11} * IAD_{GUIA} + \frac{1}{11} * IAD_{INFORMA} + \frac{1}{11} * IAD_{MAT} + \frac{1}{11} * IAD_{DOC} + \frac{1}{11} * IAD_{TUT} + \frac{1}{11} * IAD_{EVA} + \frac{1}{11} * IAD_{OTRAS} + \frac{1}{11} * K_1 + \frac{1}{11} * K_2 + \frac{1}{11} * K_4$$

E5	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,149	14,9	12,0	-2,9
IAD_{GUIA}	0,081	8,1	8,0	-0,1
$IAD_{INFORMA}$	0,116	11,6	5,0	-6,6
IAD_{MAT}	0,157	15,7	10,0	-5,7
IAD_{DOC}	0,063	6,3	15,0	8,7
IAD_{TUT}	0,098	9,8	7,0	-2,8
IAD_{EVA}	0,025	2,5	8,0	5,5
IAD_{OTRAS}	0,101	10,1	10,0	-0,1
K_1	0,150	15,0	18,0	3,0
K_2	0,025	2,5	5,0	2,5
K_3	0,000	0,0	1,0	1,0
K_4	0,034	3,4	1,0	-2,4
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.4. Análisis de sensibilidad global aplicado al escenario E5

Escenario E6

$$IC_{E6} = IAD_{FORMA}^{\frac{1}{11}} * IAD_{GUIA}^{\frac{1}{11}} * IAD_{INFORMA}^{\frac{1}{11}} * IAD_{MAT}^{\frac{1}{11}} * IAD_{DOC}^{\frac{1}{11}} * IAD_{TUT}^{\frac{1}{11}} * IAD_{EVA}^{\frac{1}{11}} * IAD_{OTRAS}^{\frac{1}{11}} * K_1^{\frac{1}{11}} * K_2^{\frac{1}{11}} * K_4^{\frac{1}{11}}$$

E6	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,113	11,3	12,0	0,7
IAD_{GUIA}	0,113	11,3	8,0	-3,3
$IAD_{INFORMA}$	0,097	9,7	5,0	-4,7
IAD_{MAT}	0,189	18,9	10,0	-8,9
IAD_{DOC}	0,076	7,6	15,0	7,4
IAD_{TUT}	0,095	9,5	7,0	-2,5
IAD_{EVA}	0,033	3,3	8,0	4,7
IAD_{OTRAS}	0,132	13,2	10,0	-3,2
K_1	0,120	12,0	18,0	6,0
K_2	0,013	1,3	5,0	3,7
K_3	0,000	0,0	1,0	1,0
K_4	0,019	1,9	1,0	-0,9
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.5. Análisis de sensibilidad global aplicado al escenario E6

Escenario E8

$$IC_{E8} = \frac{12}{100} * IAD_{FORMA} + \frac{8}{100} * IAD_{GUIA} + \frac{5}{100} * IAD_{INFORMA} + \frac{10}{100} * IAD_{MAT} + \frac{15}{100} * IAD_{DOC} + \frac{7}{100} * IAD_{TUT} + \frac{8}{100} * IAD_{EVA} + \frac{10}{100} * IAD_{OTRAS} + \frac{19}{100} * K_1 + \frac{5}{100} * K_2 + \frac{1}{100} * K_4$$

E8	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,220	22,0	12,0	-10,0
IAD_{GUIA}	0,072	7,2	8,0	0,8
$IAD_{INFORMA}$	0,047	4,7	5,0	0,3
IAD_{MAT}	0,143	14,3	10,0	-4,3
IAD_{DOC}	0,054	5,4	15,0	9,6
IAD_{TUT}	0,075	7,5	7,0	-0,5
IAD_{EVA}	0,017	1,7	8,0	6,3
IAD_{OTRAS}	0,096	9,6	10,0	0,4
K_1	0,251	25,1	18,0	-7,1
K_2	0,011	1,1	5,0	3,9
K_3	0,000	0,0	1,0	1,0
K_4	0,014	1,4	1,0	-0,4
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.6. Análisis de sensibilidad global aplicado al escenario E8

Escenario E9

$$IC_{E9} = IAD_{FORMA}^{\frac{12}{100}} * IAD_{GUIA}^{\frac{8}{100}} * IAD_{INFORMA}^{\frac{5}{100}} * IAD_{MAT}^{\frac{10}{100}} * IAD_{DOC}^{\frac{15}{100}} * IAD_{TUT}^{\frac{7}{100}} * IAD_{EVA}^{\frac{8}{100}} * IAD_{OTRAS}^{\frac{10}{100}} * K_1^{\frac{19}{100}} * K_2^{\frac{5}{100}} * K_4^{\frac{1}{100}}$$

E9	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,160	16,0	12,0	-4,0
IAD_{GUIA}	0,100	10,0	8,0	-2,0
$IAD_{INFORMA}$	0,045	4,5	5,0	0,5
IAD_{MAT}	0,176	17,6	10,0	-7,6
IAD_{DOC}	0,075	7,5	15,0	7,5
IAD_{TUT}	0,080	8,0	7,0	-1,0
IAD_{EVA}	0,027	2,7	8,0	5,3
IAD_{OTRAS}	0,125	12,5	10,0	-2,5
K_1	0,193	19,3	18,0	-1,3
K_2	0,007	0,7	5,0	4,3
K_3	0,000	0,0	1,0	1,0
K_4	0,013	1,3	1,0	-0,3
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.7. Análisis de sensibilidad global aplicado al escenario E9

Escenario E10

$$IC_{E10} = (\frac{1}{4} * IAD_{FORMA} + \frac{1}{4} * IAD_{GUIA} + \frac{1}{4} * IAD_{INFORMA} + \frac{1}{4} * IAD_{MAT})^{\frac{1}{3}} * (\frac{1}{4} * IAD_{DOC} + \frac{1}{4} * IAD_{TUT} + \frac{1}{4} * IAD_{EVA} + \frac{1}{4} * IAD_{OTRAS})^{\frac{1}{3}} * (\frac{1}{3} * K_1 + \frac{1}{3} * K_2 + \frac{1}{3} * K_4)^{\frac{1}{3}}$$

E10	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,164	16,4	12,0	-4,4
IAD_{GUIA}	0,060	6,0	8,0	2,0
$IAD_{INFORMA}$	0,079	7,9	5,0	-2,9
IAD_{MAT}	0,098	9,8	10,0	0,2
IAD_{DOC}	0,056	5,6	15,0	9,4
IAD_{TUT}	0,089	8,9	7,0	-1,9
IAD_{EVA}	0,022	2,2	8,0	5,8
IAD_{OTRAS}	0,082	8,2	10,0	1,8
K_1	0,216	21,6	18,0	-3,6
K_2	0,076	7,6	5,0	-2,6
K_3	0,000	0,0	1,0	1,0
K_4	0,059	5,9	1,0	-4,9
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.8. Análisis de sensibilidad global aplicado al escenario E10

Escenario E12

$$IC_{E12} = (0,3 * IAD_{FORMA} + 0,3 * IAD_{GUIA} + 0,3 * IAD_{INFORMA} + 0,3 * IAD_{MAT} + 0,7 * IAD_{DOC} + 0,7 * IAD_{TUT} + 0,7 * IAD_{EVA} + 0,7 * IAD_{OTRAS}) * (0,65 * K_1 + 0,3 * K_2 + 0,05 * K_4)$$

E12	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,171	17,1	12,0	-5,1
IAD_{GUIA}	0,057	5,7	8,0	2,3
$IAD_{INFORMA}$	0,050	5,0	5,0	0,0
IAD_{MAT}	0,090	9,0	10,0	1,0
IAD_{DOC}	0,060	6,0	15,0	9,0
IAD_{TUT}	0,093	9,3	7,0	-2,3
IAD_{EVA}	0,024	2,4	8,0	5,6
IAD_{OTRAS}	0,101	10,1	10,0	-0,1
K_1	0,297	29,7	18,0	-11,7
K_2	0,043	4,3	5,0	0,7
K_3	0,000	0,0	1,0	1,0
K_4	0,014	1,4	1,0	-0,4
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.9. Análisis de sensibilidad global aplicado al escenario E12

Escenario E13

$$IC_{E13} = (0,3 * IAD_{FORMA} + 0,3 * IAD_{GUIA} + 0,3 * IAD_{INFORMA} + 0,3 * IAD_{MAT} + 0,7 * IAD_{DOC} + 0,7 * IAD_{TUT} + 0,7 * IAD_{EVA} + 0,7 * IAD_{OTRAS}) * (0,65 * K_1 + 0,3 * K_2 + 0,05 * K_4)$$

E13	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,332	33,2	12,0	-21,2
IAD_{GUIA}	0,028	2,8	8,0	5,2
$IAD_{INFORMA}$	0,005	0,5	5,0	4,5
IAD_{MAT}	0,005	0,5	10,0	9,5
IAD_{DOC}	0,023	2,3	15,0	12,7
IAD_{TUT}	0,085	8,5	7,0	-1,5
IAD_{EVA}	0,028	2,8	8,0	5,2
IAD_{OTRAS}	0,070	7,0	10,0	3,0
K_1	0,397	39,7	18,0	-21,7
K_2	0,004	0,4	5,0	4,6
K_3	0,003	0,3	1,0	0,7
K_4	0,020	2,0	1,0	-1,0
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.10. Análisis de sensibilidad global aplicado al escenario E13

Escenario E14

$$IC_{E14} = (IAD_{FORMA}^{\frac{1}{4}} * IAD_{GUIA}^{\frac{1}{4}} * IAD_{INFORMA}^{\frac{1}{4}} * IAD_{MAT}^{\frac{1}{4}})^{\frac{1}{3}} * (IAD_{DOC}^{\frac{1}{4}} * IAD_{TUT}^{\frac{1}{4}} * IAD_{EVA}^{\frac{1}{4}} * IAD_{OTRAS}^{\frac{1}{4}})^{\frac{1}{3}} * (K_1^{\frac{1}{3}} * K_2^{\frac{1}{3}} * K_3^{\frac{1}{3}} * K_4^{\frac{1}{3}})^{\frac{1}{3}}$$

E14	S_i normalizado	Efectos principales	Pesos objetivo	Diferencia
IAD_{FORMA}	0,122	12,2	12,0	-0,2
IAD_{GUIA}	0,088	8,8	8,0	-0,8
$IAD_{INFORMA}$	0,064	6,4	5,0	-1,4
IAD_{MAT}	0,087	8,7	10,0	1,3
IAD_{DOC}	0,111	11,1	15,0	3,9
IAD_{TUT}	0,129	12,9	7,0	-5,9
IAD_{EVA}	0,056	5,6	8,0	2,4
IAD_{OTRAS}	0,097	9,7	10,0	0,3
K_1	0,138	13,8	18,0	4,2
K_2	0,042	4,2	5,0	0,8
K_3	0,000	0,0	1,0	1,0
K_4	0,067	6,7	1,0	-5,7
$ \Sigma $	1,000	100,0	100,0	

Tabla 10.11. Análisis de sensibilidad global aplicado al escenario E14

Anexo C. Resultados del análisis de incertidumbre

Escenario E_1

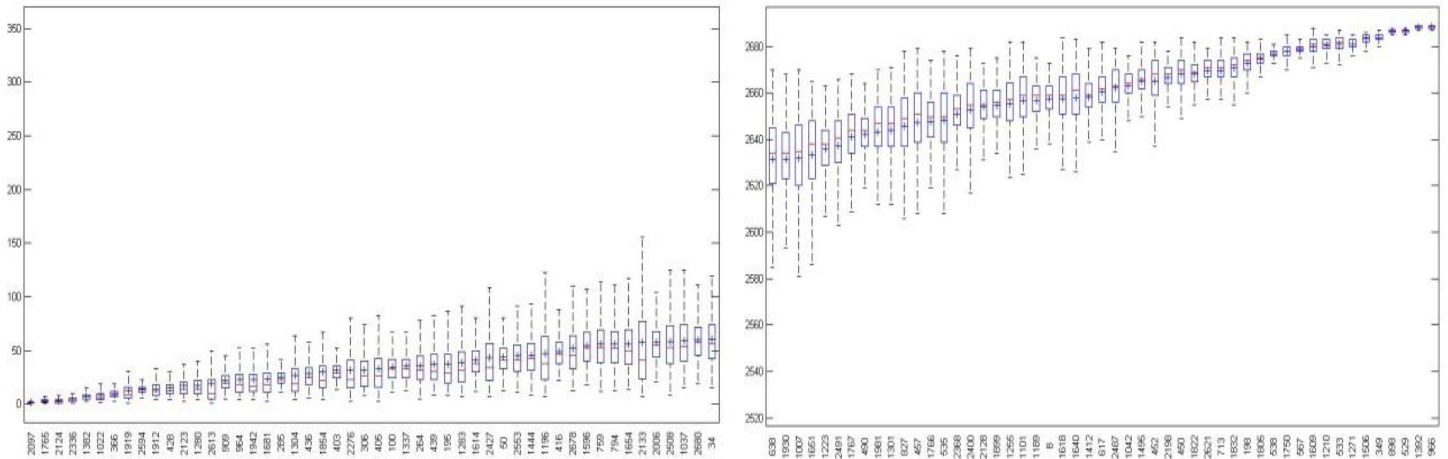


Figura 10.1. Resultados del AI donde se muestra el E_1 referencial (cruces azules), la media (marcas rojas) y el correspondiente P_5 y P_{95} de la distribución del AI para los 50 “primeros” y “últimos” profesores

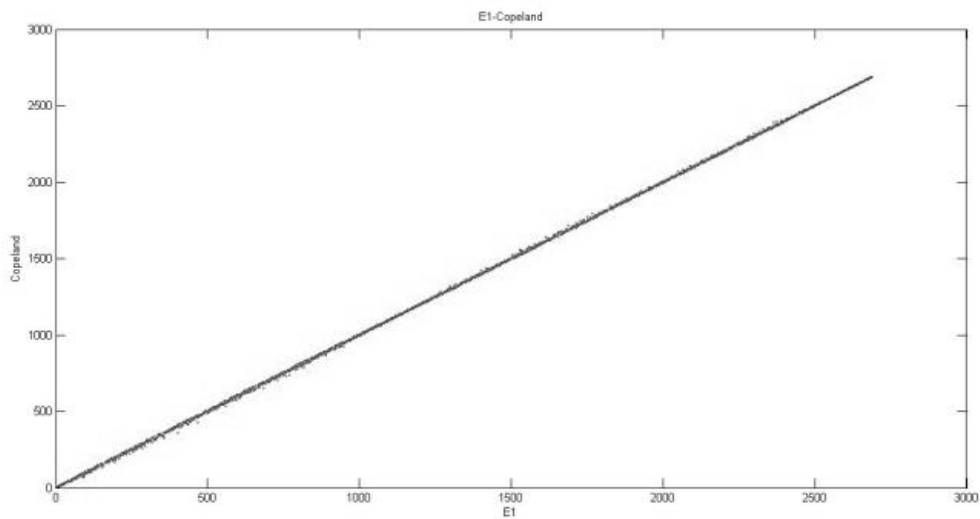


Figura 10.2. Comparación de las posiciones del escenario referencial E_1 y las posiciones aplicando Copeland

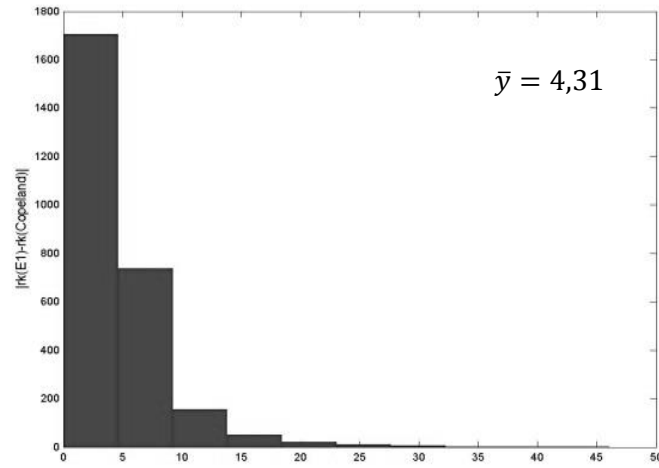


Figura 10.3. Histograma de las diferencias en posiciones del escenario referencial E1 y las posiciones aplicando Copeland

Escenario E₇

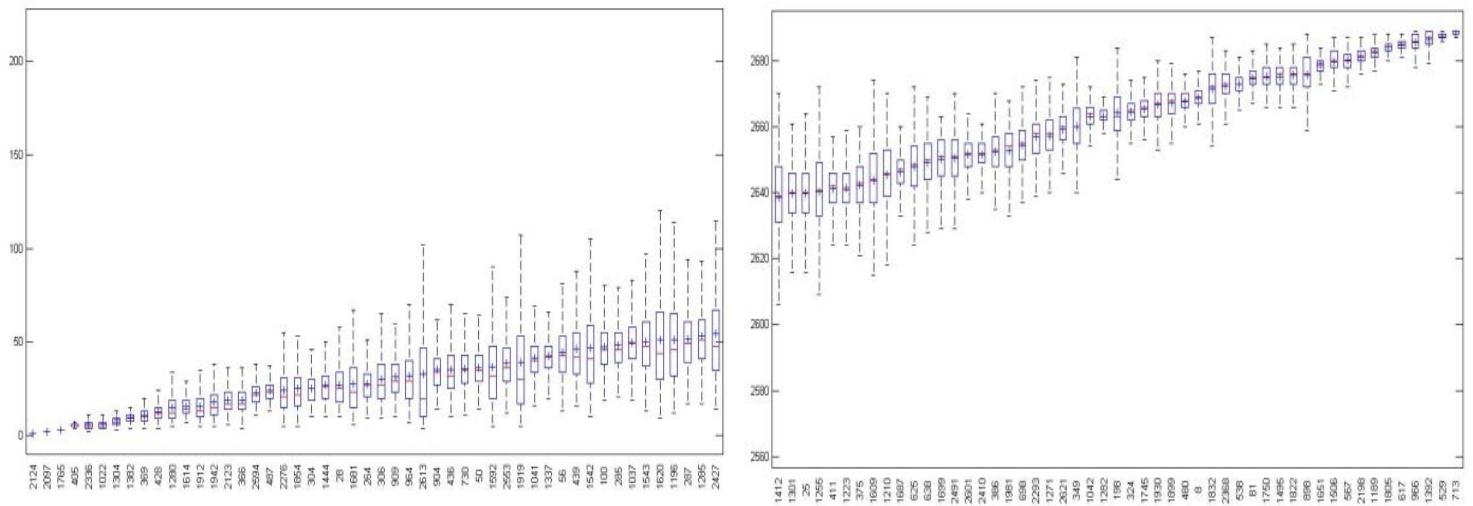


Figura 10.4. Resultados del AI donde se muestra el E7 referencial (cruces azules), la media (marcas rojas) y el correspondiente P₅ y P₉₅ de la distribución del AI para los 50 “primeros” y “últimos” profesores

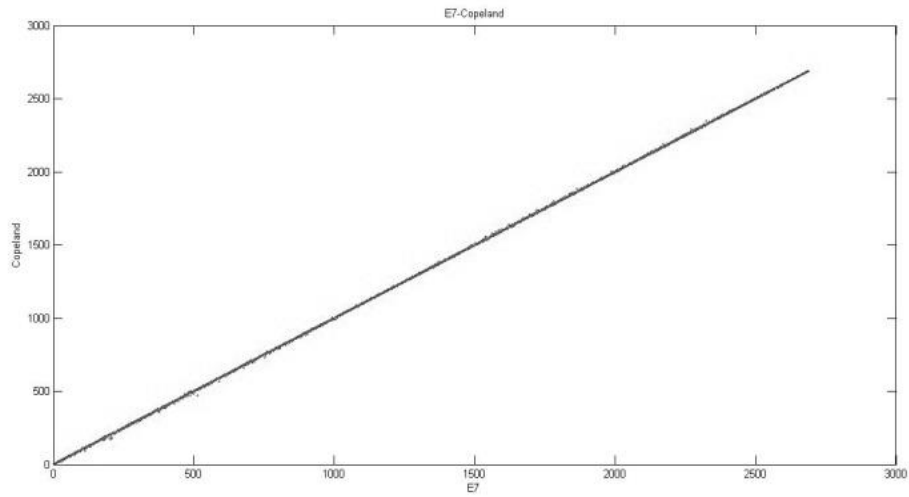


Figura 10.5. Comparación de las posiciones del escenario referencial E7 y las posiciones aplicando Copeland

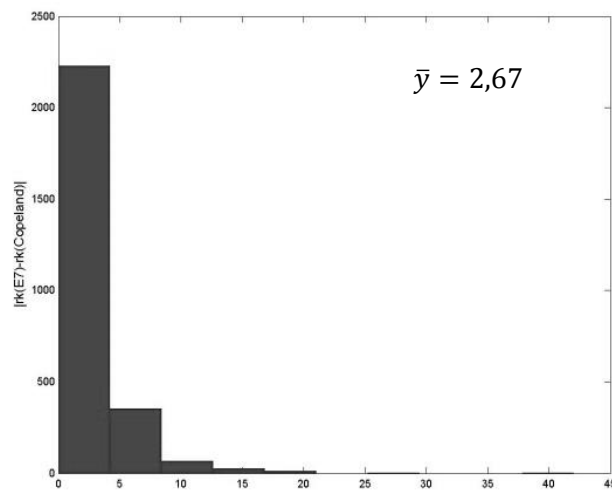


Figura 10.6. Histograma de las diferencias en posiciones del escenario referencial E7 y las posiciones aplicando Copeland

Escenario E_{11}

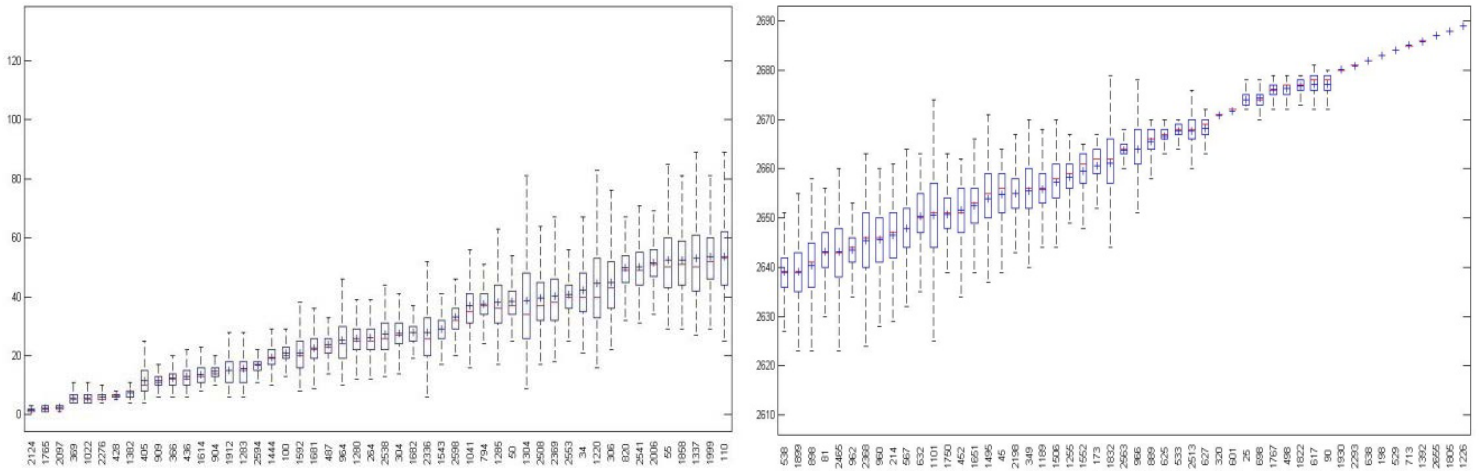


Figura 10.7. Resultados del AI donde se muestra el E11 referencial (cruces azules), la media (marcas rojas) y el correspondiente P_5 y P_{95} de la distribución del AI para los 50 “primeros” y “últimos” profesores

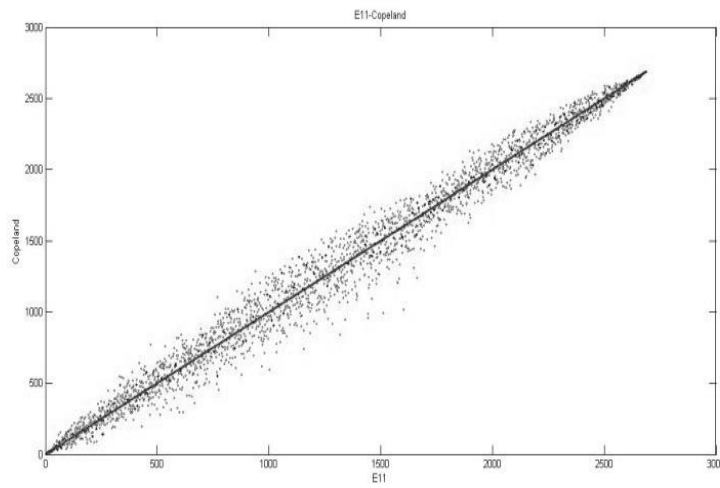


Figura 10.8. Comparación de las posiciones del escenario referencial E11 y las posiciones aplicando Copeland

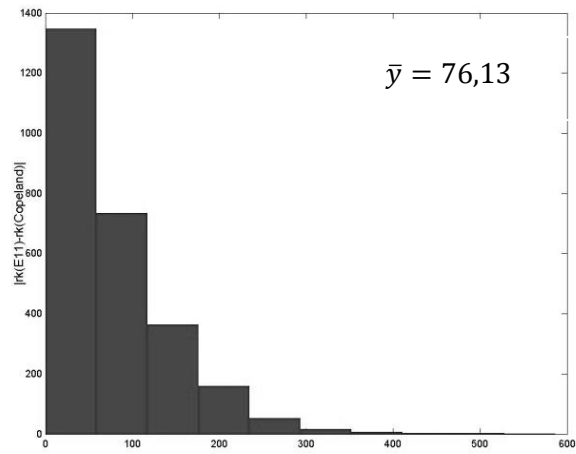


Figura 10.9. Histograma de las diferencias en posiciones del escenario referencial E11 y las posiciones aplicando Copeland

Anexo D. Secuencias de baja discrepancia

El concepto **discrepancia** proporciona una métrica cuantitativa de la desviación de una secuencia de números respecto a la distribución uniforme. La discrepancia es una propiedad fundamental en los métodos Monte Carlo.

Sea un conjunto de puntos $P = x_1, \dots, x_n \in C$ con $I^s = [0,1]^s$. Para un subconjunto arbitrario B en I^s se define la función de conteo que indica cuántos elementos de P pertenecen a B mediante la expresión,

$$A(B; P) = \sum_{n=1}^N c_B(x_n)$$

donde c_B es la función característica del conjunto B ($c_B(x) = 1 \leftrightarrow x \in B$).

Para una familia no vacía de subconjuntos $\mathcal{B} \subseteq I^s$ medibles en el sentido de Lebesgue, la métrica discrepancia para un conjunto de puntos P se define mediante la expresión,

$$D_N(\mathcal{B}; P) = \sup_{B \in \mathcal{B}} \left| \frac{A(B; P)}{N} - \lambda_s(B) \right|$$

donde λ_s es la medida de Lebesgue s -dimensional. De forma conceptual, puede interpretarse que una secuencia de baja discrepancia tiende a no agrupar puntos y, por tanto, a lograr un mejor recubrimiento del hipercubo de dimensión s que una secuencia de puntos aleatorios. En un conjunto de puntos de baja discrepancia (o de números cuasialeatorios), cada punto es generado por un mecanismo que considera la restricción de alta correlación con números precedentes, evitando los fenómenos de agrupamiento y baja cobertura de las secuencias de puntos generadas por mecanismos aleatorios. El objetivo de cada nuevo punto es, por tanto, tratar de cubrir los mayores espacios vacíos existentes entre los números previamente generados. Por tanto, un conjunto de punto P se considera de baja discrepancia si el valor $D_N(\mathcal{B}; P)$ es pequeño.

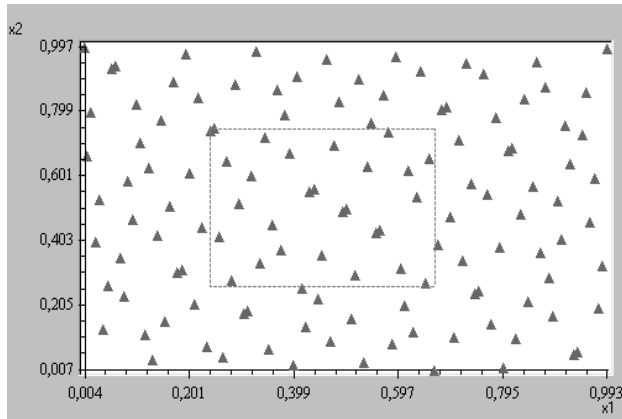
En particular, las **secuencias de Sobol'**, LP_τ , son un ejemplo de secuencias de baja discrepancia que tienen la propiedad de cubrir de forma uniforme el dominio de trabajo. Este tipo de secuencias intenta evitar los agrupamientos de puntos.

Considérese la representación en base 2 de un entero positivo $n > 1$ utilizando dígitos d_k ; $0 \leq d_k(n) < 2$:

$$n = \sum_{k=0}^{L-1} d_k(n) \cdot 2^k$$

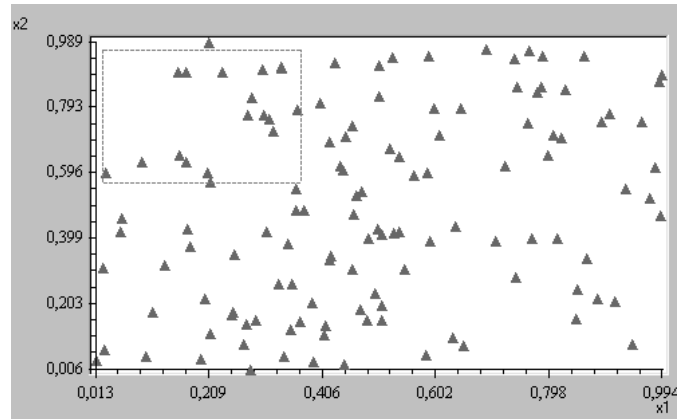
La secuencia de Sobol' queda definida por $g_2(1), \dots, g_2(N)$, donde $g_2(n)$ es la función que revierte la representación en base 2 del entero n :

$$g_2(n) = \sum_{k=0}^{L-1} d_k(n) \cdot 2^{-(k-1)}$$



(a)

(a) secuencia de Sobol' con $N=128$;



(b)

(b) secuencia aleatoria con $N=128$. Software: Simlab 2.2.1

Anexo E. Escenarios de trabajo propuestos para la construcción del IAA

ESCENARIO	MODELO	PESOS Y RESTRICCIONES	NORMALIZACIÓN	SOLUCIÓN A LA PRESENCIA DE CEROS EN LAS DIMENSIONES DE INVESTIGACIÓN Y GESTIÓN
UPV	Agregación de dimensiones con pesos fijos	$IAA_{UPV_c} = 67,60 * \frac{IAD_c^*}{43,48} + 26,40 * \frac{VAI_c^*}{38,11} + 6 * \frac{AG_c^*}{1,56}$	Dividir las puntuaciones de las dimensiones por las puntuaciones medias de las dimensiones correspondientes para todo el profesorado	Ninguna
E1	BOD	$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \forall j = 1, \dots, c, \dots, 2.020$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	Ninguna
E2	BOD	$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \forall j = 1, \dots, c, \dots, 2.020$ $0,25 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,75$ $0,25 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,75$ $0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	<p>Calcular el IAA correspondiente a aquellos profesores k con una investigación y gestión nula con la siguiente fórmula:</p> $IAA_{E2_k}^* = \max_{w_{k,1}} \frac{w_{k,1} I_{k,1}}{\max_{(I_{1,1}, \dots, I_{j,1}, \dots, I_{n,1})} w_{k,1} I_{j,1}} * 0,75$ $= \frac{I_{k,1}}{\max_{(I_{1,1}, \dots, I_{j,1}, \dots, I_{n,1})} I_{j,1}} * 0,75$
E3	BOD	$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \forall j = 1, \dots, c, \dots, 2.020$ $0,25 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,75$ $0,25 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,75$ $0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} y_{c,3}} \leq 0,5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación

ESCENARIO	MODELO	PESOS Y RESTRICCIONES	NORMALIZACIÓN	SOLUCIÓN A LA PRESENCIA DE CEROS EN LAS DIMENSIONES DE INVESTIGACIÓN Y GESTIÓN
E4	BOD	$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020$ $0,01 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,95$ $0,01 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,95$ $0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,8$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación
E5	BOD	$\sum_{i=1}^3 w_{c,i} I_{j,i} \leq 1 \quad \forall j = 1, \dots, c, \dots, 2.020$ $0,01 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,95$ $0,01 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,95$ $0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,8$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	Traslación en todas las dimensiones
E6	BOD modificado	$w_{c,1} + w_{c,2} + w_{c,3} = 100$ $0,25 \leq \frac{w_{c,1} I_{c,1}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,75$ $0,25 \leq \frac{w_{c,2} I_{c,2}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,75$ $0 \leq \frac{w_{c,3} I_{c,3}}{w_{c,1} I_{c,1} + w_{c,2} I_{c,2} + w_{c,3} Y_{c,3}} \leq 0,5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	Ninguna	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación

ESCENARIO	MODELO	PESOS Y RESTRICCIONES	NORMALIZACIÓN	SOLUCIÓN A LA PRESENCIA DE CEROS EN LAS DIMENSIONES DE INVESTIGACIÓN Y GESTIÓN
E7	BOD modificado	$w_{c,1} + w_{c,2} + w_{c,3} = 100$ $0,25 \leq \frac{w_{c,1}I_{c,1}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,75$ $0,25 \leq \frac{w_{c,2}I_{c,2}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,75$ $0 \leq \frac{w_{c,3}I_{c,3}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	z-score con traslación para evitar números negativos	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación
E8	BOD modificado	$w_{c,1} + w_{c,2} + w_{c,3} = 100$ $0,25 \leq \frac{w_{c,1}I_{c,1}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,75$ $0,25 \leq \frac{w_{c,2}I_{c,2}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,75$ $0 \leq \frac{w_{c,3}I_{c,3}}{w_{c,1}I_{c,1} + w_{c,2}I_{c,2} + w_{c,3}Y_{c,3}} \leq 0,5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	min-max con traslación para evitar ceros en la base de datos	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación
E9	BOD modificado	$w_{c,1} + w_{c,2} + w_{c,3} = 100$ $25 \leq w_{c,1} \leq 75$ $25 \leq w_{c,2} \leq 75$ $0 \leq w_{c,3} \leq 5$ $w_{c,i} \geq 0 \quad i = 1, \dots, 3$	min-max con traslación para evitar ceros en la base de datos	Asignar una puntuación mínima de 5 puntos a todo el profesorado evaluado en la dimensión de investigación

Tabla 10.12. Escenarios de trabajo propuestos para la construcción del IAA

