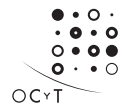


Encuestas, datos y descubrimiento de conocimiento sobre la
innovación
en Colombia

Encuestas, datos y descubrimiento de conocimiento sobre la
innovación
en Colombia



Encuestas, datos y descubrimiento de conocimiento sobre la innovación en Colombia

ISBN: 978-958-716-307-0

Primera edición: Bogotá, 2009

Esta publicación es resultado del proyecto de investigación *Descubrimiento de conocimiento sobre la innovación en Colombia, a partir de las encuestas de innovación y desarrollo tecnológico, la Encuesta Anual Manufacturera y la base de datos scienTI*, realizado con el apoyo financiero de Colciencias y con recursos de la Universidad Nacional de Colombia (sedes Medellín y Bogotá), la Pontificia Universidad Javeriana, la Universidad Pontificia Bolivariana y el Observatorio Colombiano de Ciencia y Tecnología. El proyecto fue dirigido por Jorge Robledo Velásquez.

Jorge Robledo Velásquez
Florentino Malaver Rodríguez
Marisela Vargas Pérez
Editores

Ella Suárez
Corrección de estilo

Ángel David Reyes Durán
Diseño y diagramación

Fundación Cultural Javeriana de Artes Gráficas (Javegraf)
Impresión

Prohibida la reproducción total o parcial de esta obra o su difusión telemática sin citar la fuente.

Impreso y hecho en Colombia
Printed and made in Colombia

Contenido

Agradecimientos.....	9
Introducción	11
Capítulo 1. Sobre las nociones de dato, información y conocimiento.....	15
<i>Jorge Charum Díaz</i>	
Capítulo 2. Metadatos de una bodega de datos para descubrir conocimiento	33
<i>Claudia Jiménez Ramírez, Fernán Villa Garzón y Martín Rico Herrera</i>	
Capítulo 3. Gestión de datos e información para el análisis de la innovación a partir de encuestas: aprendizaje y retos	53
<i>Claudia Nelcy Jiménez Hernández, Martín Rico Herrera y Claudia Jiménez Ramírez</i>	
Capítulo 4. Relación entre capacidades de innovación tecnológica y el desempeño empresarial y sectorial	71
<i>Jorge Robledo Velásquez, Fredy Alexander Gómez Jiménez y Juan Felipe Restrepo Arias</i>	
Capítulo 5. Caracterización del relacionamiento empresas-instituciones de educación superior según la Segunda Encuesta Colombiana de Innovación..	101
<i>Manuel Santiago Arbeláez Cuartas, Fredy Alexander Gómez Jiménez y Lida Isabel Tamayo Giraldo</i>	
Capítulo 6. Las capacidades de absorción tecnológica: una mirada centrada en la adquisición de la tecnología	123
<i>Florentino Malaver Rodríguez, Marisela Vargas Pérez y Felipe Ardila Camelo</i>	

Capítulo 7. Internacionalización del financiamiento de las actividades de desarrollo tecnológico de las empresas establecidas en Colombia	157
<i>Fredy Alexander Gómez Jiménez y Jorge Robledo Velásquez</i>	
Capítulo 8. La financiación de la innovación: un análisis a partir de la Encuesta de Innovación de Bogotá y Cundinamarca.....	175
<i>Jaime Humberto Sierra González, Florentino Malaver Rodríguez y Marisela Vargas Pérez</i>	
Capítulo 9. Interés de las empresas por la financiación de sus actividades de I+D+i: un análisis en el marco de las redes complejas para el sector industrial manufacturero colombiano	211
<i>Diego Leonardo Otálora Montenegro, Rafael Germán Hurtado Heredia y Carlos Quimbay Herrera</i>	
Capítulo 10. Análisis Comparativo entre Árboles de Regresión y Clasificación (CART) y Regresión Logística: Aplicación a la Caracterización de Empresas Innovadoras Colombianas	235
<i>Sandra Carolina Serna Pineda, Juan Carlos Correa Morales y Jorge Robledo Velásquez</i>	
Capítulo 11. Aplicación exploratoria de técnicas basadas en sistemas neurodifusos al estudio de la innovación a partir de encuestas realizadas en Colombia	265
<i>Luis Felipe Duque Álvarez, Jorge Manrique Henao y Jorge Robledo Velásquez</i>	
Capítulo 12. Tratamiento de datos y técnicas para analizar la relación entre el uso energético y la innovación a partir de encuestas oficiales colombianas	287
<i>Jorge Manrique Henao y Jorge Robledo Velásquez</i>	
Capítulo 13. Los datos, la información y el conocimiento sobre la innovación en Colombia y sus perspectivas de desarrollo.....	329
<i>Jorge Robledo Velásquez y Florentino Malaver Rodríguez</i>	
Autores.....	363

Agradecimientos

Los autores agradecen al Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias), por la financiación otorgada al proyecto *Descubrimiento de conocimiento sobre la innovación en Colombia a partir de las encuestas de innovación y desarrollo tecnológico, la Encuesta Anual Manufacturera y la base de datos Scien TI*, en el marco del cual se llevó a cabo el trabajo aquí presentado. Agradecen, igualmente, a los demás investigadores y estudiantes integrantes del equipo del Proyecto que no figuran como autores, pero cuyas críticas, comentarios y sugerencias constituyeron un insumo de gran importancia para el trabajo; su aporte intelectual y su apoyo humano contribuyeron, sin duda, de forma decisiva a la creación de un ambiente de trabajo cordial y estimulante de creatividad y colaboración. Un reconocimiento explícito también merecen las instituciones que hicieron posible el acceso a los datos, particularmente Colciencias, el Departamento Administrativo Nacional de Estadística (DANE), el Departamento Nacional de Planeación (DNP) y la Cámara de Comercio de Bogotá, responsables de la aplicación de las encuestas, cuyos datos son el insumo fundamental del trabajo aquí reportado. Finalmente, los autores agradecen a las instituciones a las que pertenecen, sin cuyo respaldo hubiera sido imposible llevar a cabo el trabajo: la Universidad Nacional de Colombia (sedes Medellín y Bogotá), la Pontificia Universidad Javeriana, la Universidad Pontificia Bolivariana y el Observatorio Colombiano de Ciencia y Tecnología; particularmente, agradecen a los directivos y personal de este último, que realizó con gran profesionalismo la gestión administrativa y financiera del Proyecto.

Introducción

Variadas y cambiantes son las formas en que se innova, debido a los múltiples actores que intervienen en su generación, a las relaciones que establecen entre ellos y a las particularidades de los contextos (históricos, geográficos e institucionales) en que actúan y que los condicionan. Por ello, la innovación es un fenómeno complejo, contextualizado y dinámico. Para captar esos rasgos, así como las dinámicas y transformaciones de la innovación en Colombia, se requieren herramientas metodológicas y marcos de análisis e interpretación más amplios y flexibles.

A efectos de enfrentar ese desafío, se llevó a cabo el proyecto *Descubrimiento de conocimiento sobre la innovación en Colombia* (Proyecto DESCUBRIMIENTO), en el cual se exploraron técnicas para descubrir conocimiento en bases de datos (KDD, por su sigla en inglés), con el propósito general de contribuir a desarrollar los análisis de los procesos de innovación en Colombia. Con este fin se acudió a las amplias posibilidades que ofrecen dichas herramientas para escudriñar en los datos (de las encuestas de innovación), por ejemplo, por patrones no evidentes y no lineales de comportamiento y de asociación entre variables, los cuales, por lo demás, son congruentes con las características de los procesos de innovación señaladas; sin embargo, son instrumentos de análisis muy poco utilizados en el país.

Como fuentes principales de información se acudió a las encuestas de innovación, que constituyen los esfuerzos más significativos realizados en el país, por la cantidad de establecimientos encuestados, por la periodicidad

de su aplicación, por su vocación de permanencia y por las potencialidades de desarrollo que ofrecen. También se acudió a las encuestas con el propósito de aprovechar las enormes posibilidades que ofrecen de convertirse en un instrumento útil para la formulación y evaluación de políticas sobre la innovación en el país. En el estudio se acudió, en particular, a la Primera y Segunda Encuestas de Innovación, aplicadas a la industria manufacturera nacional, en 1996 y 2005 (EDT I y EDIT II) y a la Encuesta de Innovación en la Industria de Bogotá y Cundinamarca (EIByC), en 2005.¹

Adicionalmente, se acudió a otras fuentes de información: a CvLAC y GrupLAC, de la plataforma ScienTI, con el fin de identificar los vínculos entre la innovación y la investigación, y a la Encuesta Anual Manufacturera, para establecer la relación entre la innovación y el desempeño económico industrial. Debe advertirse, sin embargo, que el uso efectivo de los datos provenientes de tan diferentes fuentes se logró sólo parcialmente, y más en lo informático y de desarrollo metodológico que analíticamente, por las dificultades de acceso a los metadatos y a la estructura de las bases de datos ScienTI, al igual que por la imposibilidad de cruzar los datos de las encuestas de innovación con los de la Encuesta Anual Manufacturera.

Un objetivo central del proyecto fue construir una bodega de datos debidamente fundamentada en las nociones de dato, información y conocimiento, a fin de coadyuvar al avance en los procesos de normalización y estandarización requeridos para captar el significado, sentido y alcance de la información incorporada en las bases de datos que alimentan la bodega;

¹ La EDT I fue adelantada por Colciencias y el Departamento Nacional de Planeación (DNP); se aplicó en una muestra de 885 establecimientos industriales del país, y fue representativa al nivel de tres dígitos del Código Industrial Internacional Uniforme (CIIU), revisión 2, de un universo de 4.501, que reportó la Encuesta Anual Manufacturera (EAM), de 1995. La EDIT II fue adelantada por Colciencias, el DNP y el Departamento Administrativo Nacional de Estadística (DANE); se encuestaron 6.172 empresas industriales, correspondientes a la totalidad de las empresas reportadas por la EAM, en 2003. Finalmente, la EIByC fue realizada por la Cámara de Comercio de Bogotá (CCB) y el Observatorio Colombiano de Ciencia y Tecnología (OCyT); se aplicó en 404 establecimientos industriales, que fueron representativos, al nivel de tres dígitos de la CIIU revisión 2, de un universo conformado por 2.404 establecimientos de Bogotá y Cundinamarca, con más de 10 empleados, según la EAM de 2002 y el Registro Mercantil de la CCB.

así como para que la información suministrada (metadatos) sobre cada uno de los datos almacenados permita su estandarización y, de este modo, facilitar a los investigadores “trabajar con los datos”, esto es, leerlos desde sus propias nociones y perspectivas teóricas. La bodega de datos también fue construida pensando en que a partir de una “interfaz amigable” se facilitara en el futuro a los investigadores, incluso a quienes no dominan los lenguajes especializados de consulta, acceder a distancia a la bodega, y de esta manera racionalizar sus esfuerzos e incrementar la productividad de su trabajo.

Así, desde las perspectivas teórica y técnica, la bodega de datos se construyó con el propósito de mejorar tanto las posibilidades de comprensión de los datos disponibles y del trabajo con ellos como de facilitar el uso de la rica variedad de técnicas de descubrimiento de conocimiento existentes. Este propósito se sumó al doble desafío de aprovechar la información obtenida en las encuestas de innovación y de hacerlo abordando tanto el estudio de la innovación como de los procesos y relaciones que la originan, en toda su complejidad. Todo ello con el objetivo de generar información útil para la toma de decisiones de política pública y de estrategia empresarial.

Para asumir esos desafíos se conformó un equipo de investigación interdisciplinario, conformado por investigadores vinculados a grupos de investigación de las áreas de sistemas, estadística, ingeniería, física, economía y administración, que se complementan en los ámbitos disciplinar, técnico y en las temáticas estudiadas. Este equipo desarrolló sus actividades en una dinámica de trabajo típica de una red de investigación; contó con la financiación de Colciencias y con el apoyo de un convenio interinstitucional firmado por el OCyT y las universidades Nacional (sedes Medellín y Bogotá), Javeriana y Pontificia Bolivariana. Además, contó con la colaboración informal, pero rigurosa y fecunda, de investigadores de las universidades EAFIT y de Antioquia.

Los principales hallazgos generados por la investigación colaborativa y en red adelantada en el marco del Proyecto DESCUBRIMIENTO se consignan en los trece capítulos que conforman este libro. En los tres primeros se presentan los resultados de los esfuerzos de construcción de la bodega de datos y adecuación de la información requerida para suministrar los

datos utilizados en los siguientes nueve capítulos, en los cuales se explora en las técnicas de descubrimiento para profundizar en la indagación sobre las características y dinámica de la innovación. La principal característica de estos capítulos es la variedad en los temas tratados,² en las técnicas de descubrimiento usadas y en los énfasis en los unos o en las otras, lo que refleja el carácter abierto del Proyecto y de su objeto de estudio.

Finalmente, en el último capítulo se realiza un esfuerzo de síntesis de los resultados, se efectúa una reflexión sobre las posibilidades de investigación que estos abren y se formulan recomendaciones orientadas a fortalecer las encuestas de innovación, para que se adecúen en mayor grado a las exigencias de marcos y herramientas analíticas más desarrollados. En concordancia con el carácter exploratorio del estudio, en este capítulo, antes que concluir, se señalan hallazgos que revelan desafíos.

² Estos cubren la financiación de la innovación, las capacidades de innovación, las capacidades de absorción, los perfiles de las empresas innovadoras, la relación entre las empresas y universidades, la cooperación internacional, la energía, la innovación y el desempeño tecnológico y del negocio, etc.

Capítulo 1

Sobre las nociones de dato, información y conocimiento

Jorge Charum Díaz

La opción de *descubrir el conocimiento presente en las bases de datos* suscita múltiples preguntas que ya están insinuadas en el mismo título del proyecto de investigación. En efecto, las acciones de descubrir, conocer y representar, o la manipulación de objetos aparentemente sin complejidad, como “base de datos” e incluso “dato”, que están en el centro de las actividades desplegadas en las ciencias de la información, deben establecer el alcance que tendrán en el desarrollo del trabajo. Antes de señalar lo descubierto, es conveniente reflexionar sobre algunos de los elementos que serán puestos en movimiento durante el decurso del Proyecto DESCUBRIMIENTO.

El uso del término *norma*, que con tanta frecuencia se encuentra en los textos sobre la información, debe ser aquí más cuidadoso de lo acostumbrado para lograr los acuerdos sobre el significado, el sentido y el alcance de cada uno de los conceptos que califican o representan los objetos puestos en operación. Así se hará una presentación de la noción de norma que se utilizará y de su puesta en escena para representar como dato cada uno de los objetos en juego, a fin de señalar, por una parte, el carácter limitado, basado en el conocimiento alcanzado sobre el objeto,

de esa representación y, por la otra, el carácter vinculante de toda norma, por cuanto su aceptación permite conocer tanto el uso propio y el que los otros le dan a cada una de las normas, y que está entonces en la base de la construcción de su legitimidad.

La diferenciación entre forma y contenido permite pasar del conocimiento puesto en la norma a su representación concreta en los objetos tecnológicos diseñados para recoger la información de los múltiples representantes de la misma especie, inicialmente como formularios o fichas (la *forma*), y luego como metadatos que recogen las características y cualidades particulares de los objetos individuales, en cuanto *contenido*.

También se consideró pertinente reflexionar sobre el estatus del dato, sobre todo con el propósito de hacer hincapié en su carácter de *constructo* que debe lograr su validez social. El dato alcanza su forma para poder recoger los objetos de la misma especie y alcanza su contenido para individualizar a cada uno, conservando las cualidades propias y las características comunes que pueden ponerse en relación según las teorías de la investigación, de la innovación, por ejemplo, que permitan interpretar.

En efecto, si se tienen los *datos* (esto es, el primer nivel), es posible analizarlos, en cuanto agregaciones, síntesis, clasificaciones o como resultados de la aplicación de procedimientos estadísticos univariados, bivariados o multivariados. Esto a fin de alcanzar la *información elaborada* (o segundo nivel). Luego, para llegar al *conocimiento* (tercer nivel) es preciso contar con teorías que, por una parte, respondan por las modalidades como entran en relación los diferentes elementos que representan los datos, sus representaciones, sus combinaciones y, por la otra, permitan su interpretación.

En este último nivel se lograría poner de presente relaciones inéditas entre los datos y, quizás, es el momento en que se comienzan a construir las nuevas preguntas que, probablemente, están en la frontera del descubrimiento. Por supuesto, estas interpretaciones vuelven a poner en tensión los datos y al sujeto que interpreta.

Luego de concluir el recorrido, de aplicar las nociones de dato, información, conocimiento a los datos presentes en las bases de datos, se puede intentar responder a las cualidades de los datos con que se cuenta, a su capacidad de representación, a la satisfacción de las normas para formular

las conclusiones, los éxitos y las limitaciones que permitirán proponer la segunda fase del estudio.

NORMAS Y ESTÁNDARES

En el *Diccionario de la lengua española*, de la Real Academia, cuando se quiere precisar el sentido del término *norma*, se encuentra: “... // 2. Regla que se debe seguir o a que se deben ajustar las conductas, tareas, actividades, etc. [...] // 3. *Der.* Precepto jurídico”. Entre tanto, *normalizar* es “Regularizar, o poner en orden lo que no lo estaba [...] // 3. *Tipificar*, ajustar a un tipo, modelo o norma”. Para el término *Estándar*, luego de recordar que proviene del inglés, se encuentra: “Dícese de lo que sirve como modelo, *norma*, patrón, o referencia”. El verbo derivado, *estandarizar*, corresponde a la acción de “*tipificar*, ajustar a un tipo, modelo o norma”. Así, a juzgar por los resultados de esta consulta, no hay marcadas diferencias entre los significados de norma y estándar; por el contrario, al menos en la tercera acepción, prácticamente son iguales. ¿Significa esto que se pueden utilizar indiferentemente?

Para su empleo en el dominio de la información es necesario, sin embargo, hacer y precisar distinciones entre estos dos términos y sus usos. En efecto, en el *Dictionnaire encyclopédique de l'information et de la documentation* (1997) se encuentra que *norma* es “un documento establecido por consenso y aprobado por un organismo reconocido, que provee, para usos comunes y repetidos, reglas, líneas directrices o características, para actividades o sus resultados, que garantizan un nivel óptimo en un contexto dado”. En la práctica esta definición de *jure* no se encuentra por lo general en su estado puro. Aparecen, y es el caso más frecuente, normas de *facto*, es decir, no aprobadas por un organismo reconocido, sino resultado de acuerdos en el interior de grupos de interés, que regulan actividades, definen elementos, dan cuenta de sus características y garantizan en su uso niveles óptimos de precisión.

Ejemplos de la primera situación son las normas de organismos nacionales, como la Association Française de Normalisation (Afnor), o

internacionales, como la Fédération Internationale des Associations de Normalisation (ISA) o la International Organization for Standardization (ISO). Ejemplos de la segunda situación son las normas que deben ser satisfechas en el interior de sistemas de información nacionales, para responder a sus necesidades, en su campo de interés.

Las normas de que se tratarán en este documento tienen, en su mayoría, este carácter de norma técnica. El paso de ser una norma de facto a una norma de jure es un problema que en muchas ocasiones debe enfrentarse mediante la constitución de instituciones nacionales, encargadas de responder por las actividades de normalización, con el fin de asegurar la aceptación social de la norma.¹

Por otra parte, también en el dominio de la información, pero ahora asociado a los bienes y servicios, apareció la necesidad de establecer normas para las tecnologías, orientada a darles forma, consistencia y confianza a los objetos tecnológicos sobre los que se basa la llamada *sociedad de la información*. Esto a fin de asegurar la compatibilidad de los servicios y de los productos desarrollados, en los que tienen intereses los diferentes actores en competencia. En este caso, conviene hablar ya no de normas, sino de estándares; no obstante, dada la lentitud de los organismos nacionales e internacionales para definirlos, en el caso de la tecnología² se busca llegar

¹ Si el uso generalizado puede significar la aceptación, no se debería dejar de establecer explícitamente el significado, el campo de acción y los alcances de cada una de las normas. Muchas de las discusiones en el momento de su aplicación se deben a la carencia de instituciones de referencia que, además de ponerlas en operación, tengan la función y la capacidad de explicarlas y de justificarlas.

² El proceso de elaboración de una norma internacional se mide en años. Las normas documentales tienen largas historias. En el dominio del Intercambio de Documentos Digitalizados (EDI), la oficina de las Naciones Unidas encargada de definir las normas documentales se tomó 15 años para llegar a la definición de una norma operativa; en el caso de las normas sobre libros, Fayet-Scribe y Canet (1999) muestran que el proceso de normalización se inició luego de la Revolución Francesa y que sólo alcanzó la posibilidad de llegar a una normalización concertada en los años setenta del siglo pasado, con la participación de la ISO, que abrió la posibilidad de elaborar métodos concertados para definir, aprobar una norma y cambiarla cuando se precisara, por ejemplo, para seguir las evoluciones tecnológicas. Respecto a la ISO, el proceso de definición y aprobación de una norma internacional lo desarrolla un comité o

a acuerdos consensuales entre los actores económicos, que son puestos rápidamente en operación, antes de una validación por parte de organismos nacionales o internacionales.

En síntesis, se tienen normas sobre los objetos y sobre los procedimientos que entran en la recolección, la integración, el tratamiento y la circulación de la información, que en muchas ocasiones son de facto, y se tienen estándares sobre los objetos tecnológicos en los que se apoyan concretamente esos procedimientos y que sirven para asegurar el funcionamiento técnico de los sistemas de información.

LOS OBJETOS INFORMACIONALES

Las formas como se usan los términos *base de datos*, *servicio de información* y *sistema de información* no corresponden a nociones precisas y diferenciadas. En realidad estos términos son utilizados por quienes los proponen, de manera indiscriminada, y esto ha dado lugar a una polisemia que hace difícil su caracterización. Por ello, a fin de superar esta situación, se estableció la noción abarcadora de *objeto informacional*, cuya función es proveer, para diferentes tipos de usuarios, datos o información organizada según distintos niveles de elaboración o estructuración.

Se considera que para lograr sus objetivos debe responder por la calificación de los usos de la información y contar con las herramientas técnicas, por lo regular apoyadas en medios informáticos, para integrar los datos primarios, tratarlos y generar los informes dirigidos a responder por las

subcomité técnico (CT o ST), en seis etapas: (i) propuesta, por al menos cinco miembros del CT; (ii) elaboración de un documento de trabajo que recoge la aceptación de miembros del grupo de trabajo; (iii) aprobación del documento por los miembros del CT/ST implicado: el texto votado tiene el status de *Draft International Standard* (DIS); (iv) etapa de encuesta sobre la aprobación: distribución del DIS a todos los miembros del ISO para voto y comentarios (cinco meses de duración). Así, siguiendo un procedimiento preciso se acepta o se rechaza, en tanto se llega al *Final Draft International Standard* (FDIS); (v) aprobación: voto final que toma dos meses; (vi) aprobación final, hecha por la secretaria central de la ISO. Las normas deben ser objeto de una evaluación cada cinco años (ISO, 2003).

necesidades de información de diversos tipos de usuarios. Así, una hoja de cálculo, una base de datos o un servicio o sistema de información se pueden tratar teniendo una misma referencia conceptual. En este capítulo sólo se utiliza la expresión *objeto informacional*, pero se debe tener en cuenta que cubre todo el espectro de posibilidades que va desde una simple base de datos hasta la complejidad de un sistema de información. Se debe, entonces, hacer las especificaciones necesarias para cada caso particular, cuando se trata de mostrar sus propiedades o cualidades.

En un objeto informacional se pueden identificar, desde el punto de vista de su estructuración interior, la dimensión del uso y la de la concepción. La primera está en relación con los tipos de usuarios que pretende atender, las modalidades de atención y de formación de los usuarios derivadas ya sea de las opciones definidas interiormente o del vínculo que ellos establecen con el objeto. La segunda dimensión, la concepción, tiene en cuenta diferentes variables que califican al objeto tecnológico específico: su estructuración interior en cuanto tal; las herramientas para incorporar, tratar los datos y generar los informes generales o específicos.

Desde el punto de vista de su operación, el objeto informacional integra los datos, los pone internamente en circulación, los asocia y genera información elaborada, es decir, información que después de varias fases de tratamiento (evaluación, agrupación o síntesis) ha adquirido un nuevo valor, la que a su vez es puesta en circulación para su uso por parte de individuos, de colectivos o de grupos sociales con prácticas comunes.

La posibilidad de interpretar la información generada está, en primera instancia, en el conocimiento que tienen los interesados sobre las orientaciones políticas que presiden la concepción del instrumento informatizado, sobre la selección o la construcción de los procedimientos técnicos regulados por nociones, conceptos y teorías y, en segunda instancia, en las concepciones y teorías que presiden sus propias investigaciones. Así, se tiene una jerarquía cognitiva entre los datos, la generación de información elaborada con base en los datos y su interpretación, que es el asunto del usuario.

NORMALIZACIÓN

La acción de normalizar la información está en el centro de los intereses de los sistemas de información. En efecto, toda su función cognitiva está sostenida por la información, la que, a su vez, se apoya en los datos. El término *normalización* referido a los sistemas de información tiene múltiples interpretaciones. Una primera acepción, que responde a los procesos a los que se someten los datos tan pronto como son integrados y antes de cualquier tipo de tratamiento, cubre las tareas de depuración y de su presentación unificada: se trata de corregir los datos que se han integrado cuando hay errores en su presentación (debidos, por ejemplo, a una mala digitación) o cuando se unifican las múltiples modalidades posibles de presentación. Esto permitirá, posteriormente, hacer su tratamiento y realizar agregaciones entre los datos.

Esta acepción, de uso común por los ingenieros, que responde a una laboriosa y cuidadosa actividad de control que consume muchos esfuerzos y que debe ser siempre enfrentada, no será objeto de una especial consideración aquí. La generación de métodos informatizados para hacer la depuración, la unificación de términos y la identificación de posibles errores es un problema que debe ser enfrentado en un proyecto particular.

Una segunda acepción corresponde al proceso de concepción, diseño y aplicación de las normas para cada uno de los elementos que entran en los procesos relativos a la información, diríamos *informacionales*, del sistema de información. Se entiende entonces por normalización el proceso por el cual se adoptan las normas sobre los elementos relativos a la información. La normalización adoptada depende de las formas de la representación consideradas para los elementos, que deben estar integradas y disponibles para almacenar, combinar y poner en circulación la información elaborada. Así, toda la función cognitiva de los sistemas de información está sostenida por la información, la que, a su vez, se apoya en los datos. Ahora bien, los datos deben ser fiables, consistentes, pertinentes y, sobre todo, deben estar normalizados.

Las normas, entendidas como formas organizadas y estructuradas que sirven de orientación para dar cuenta de los elementos que están presen-

tes o son puestos en movimiento, a través de procedimientos a menudo complejos, apelan a nociones, movilizan teorías y ponen en relación a distintos grupos sociales que se identifican por sus prácticas comunes (comunidades de prácticas) y que están interesados en la producción de los datos, en su organización y circulación, en la generación de resultados y en su utilización.

Por otra parte, los dispositivos que generan, integran, tratan, ponen en movimiento y hacen disponible la información también son sometidos a procesos de normalización —se dirá entonces normalización técnica— con el fin de asegurar el flujo de la información, desde su estado primario en cuanto dato hasta su elaboración. Hay, así, un proceso de estandarización de los objetos tecnológicos. Se trata de las interfaces entre el productor y el usuario, que aumentan el valor de los productos relativos a la información y que son parte de las redes en que las interrelaciones técnicas hacen compatibles los diferentes componentes, para asegurar la interconectividad y la interoperabilidad técnica.

Es posible distinguir los datos brutos en cuanto entrada o acumulación, materia prima que se va a someter a procesos orientados con ayuda de instrumentos para llegar a atender las necesidades de los usuarios. En un lado se tienen las informaciones-recursos. En el otro, los métodos de tratamiento de la información: las informaciones-métodos.

Desde un punto de vista general, hay una oposición entre la normalización de los objetos y la diversidad presente en cada uno de ellos. En efecto, la atribución de características comunes a un conjunto de objetos, que es propia de las normalizaciones propuestas para los objetos de la información, siempre entraña una cierta reducción de la diversidad. Por otra parte, tanto los procesos de normalización de los datos como la estandarización de las infraestructuras materiales están sujetos a diversas tensiones que ponen en juego los intereses de quienes conciben desde las teorías las orientaciones que se deben dar a la estructuración de los datos y a los informes orientados a los usuarios, de quienes conciben y diseñan los objetos informacionales en cuanto objetos tecnológicos.

NOCIÓN DE DATO PARA EL OBJETO INFORMACIONAL

En el inicio de la red de producción de la información elaborada se encuentra el dato. Se empieza con una primera reflexión sobre el estatus del dato y su constitución como elemento básico y materia prima de todo objeto informacional.

La noción de dato científico puede usarse con provecho para delimitar la noción de dato dentro del objeto informacional. El dato científico no es una simple y directa expresión de la naturaleza; no resulta tampoco de la aplicación inmediata de un método científico riguroso. Es, más bien, un *constructo* social. Contra el racionalismo que pretende que los objetos de la naturaleza están allí y que la tarea del científico es desentrañar sus cualidades y propiedades, se encuentra la corriente relativista, que considera que los datos, los hechos y los conocimientos cambian en función de las relaciones en las que entran y que no tienen un valor unívoco o universal, pues no existen a priori criterios absolutos y universales de racionalidad. Conviene interrogarse sobre los actores que han participado en su elaboración, en sus relaciones.³

En el caso de los datos en el objeto informacional se considera que ya integran un conjunto de conocimientos que están disponibles, a los que se puede apelar y que, en algunas ocasiones, permiten darle una forma al dato, que ya no está sujeta a discusiones. Por el contrario, en otros casos el conocimiento desarrollado es escaso y es todavía relativamente informe su

³ Un ejemplo interesante es el de los datos de energía. La idea misma de energía no existía antes de mediados del siglo XIX. Se tenía entonces una confusa noción de fuerza, de esfuerzo y ella, con el desarrollo de la mecánica, fue dando lugar a la de trabajo, clara sólo a partir del siglo XVIII, cuando fuerza, esfuerzo, etc. se veían como sinónimos. Por otro lado, se tenían las ideas confusas de temperatura y calor, que evolucionaron y se diferenciaron desde la invención del termómetro, en el siglo XVI, hasta la noción de calórico como una sustancia particular a finales del XVIII. Sólo cuando Rumford y Joule se dieron cuenta, ya en el siglo XIX, con máquinas térmicas de por medio, que había una equivalencia entre el calor y el trabajo, que podían ser vistos como dos formas de “algo más profundo” que vino a llamarse *energía*, surgió este concepto que hoy nos parece tan familiar, pero que fue toda una construcción social que necesitó teorías, instrumentos y procesos para llegar a ser concebida. (Esta observación es deudora de formulaciones hechas por José Luis Villaveces)

posible caracterización. El paso del estado relativamente informe del objeto al que se refiere el dato a su forma cada vez más precisa es el resultado de las inversiones previamente realizadas precisamente para darle forma.⁴

Estas inversiones, que toman una referencia concreta en los estudios realizados, son por lo regular costosas —en tiempo y en esfuerzos conceptuales— y están acompañadas de la puesta en común de opciones teóricas, de explicitación de procesos y procedimientos que han permitido buscar y, en ocasiones, alcanzar una relativa aceptación por parte de los grupos sociales que en ellas han participado, para darles forma, y que permiten a quienes tienen interés en el dato integrarlo en sus actividades prácticas.

FORMA DE LOS DATOS EN LOS OBJETOS INFORMACIONALES

La operación de normalización está precedida de una objetivación del objeto. En las ciencias no se opera inmediatamente sobre los objetos naturales; tampoco sobre los objetos sociales o culturales. Estos son resultado de una construcción que no corresponde a una lógica lineal del tipo:

Naturaleza —————> objeto —————> representación

Esta asume que los objetos son naturales, es decir, generados directamente por la naturaleza, la sociedad o la cultura, en la que el científico aparece como si sólo tuviera la tarea de extraer sus atributos, cualidades,

⁴ Las inversiones para dar forma (*investissements de forme*) a los objetos de los que se habla, además de poner en evidencia características y atributos del objeto, tienen otras consecuencias. Primera, para hablar del objeto hay que conocer el estado de desarrollo del conocimiento sobre él, lo que eleva la exigencia para participar en sus adecuaciones o en las posibles discusiones o cuestionamientos. Segunda, los consensos pueden ser alcanzados apoyándose en el conocimiento explícito de sus características y atributos. Los desacuerdos, en muchas ocasiones, tienen que ver con la dificultad de no tener un principio superior al cual referirse para superar las posiciones interesadas no coincidentes. Los objetos que están representados en las bases de datos pueden no haber sido objeto de estas inversiones para conformarlos, lo que da lugar a posibles múltiples interpretaciones (Boltanski y Thévenot, 1991).

rasgos distintivos que dan cuenta de su existencia y de sus formas propias, naturales, se diría. Por el contrario, a diferencia de este dualismo entre el objeto y su representación, el proceso de la constitución de los objetos es precisamente el resultado de una inversión de esa lógica, por la que se realiza su objetivación (Vinck, 1994).

Así, hay múltiples operaciones por medio de las cuales se construye la representación. La primera, por la alimentación de las diversas trazas e inscripciones de que se dispone —como descripciones de sus manifestaciones que han sido integradas a documentos que circulan públicamente, derivadas precisamente por esas inversiones— y que son el resultado de combinaciones, acercamientos y asociaciones que hacen aparecer un objeto, lo traen a la existencia y le dan forma:

Inscripciones —————> objeto

Por la segunda operación, luego de la constitución del objeto, se lo representa como alguna cosa que ya está allí, independiente de las inscripciones a partir de las cuales ha sido constituido, disociando el objeto de sus inscripciones. La tercera operación consiste en la inversión de la relación entre el objeto y las inscripciones: si bien el objeto es una emergencia de las inscripciones, se supone ahora que las inscripciones son producidas por el objeto.

Finalmente, para mantener la inversión, las tres primeras etapas son minimizadas, incluso olvidadas, y queda sólo la certidumbre de que la naturaleza ha producido trazas e inscripciones que imponen al observador su aceptación como provenientes del objeto (Latour y Woolgar, 1988):⁵

Objeto —————> inscripciones

⁵ La representación del objeto es así deudora de las inscripciones, las que a su vez se derivan de manipulaciones cuyos resultados son consignados como trazas escritas en los textos. Los instrumentos producen así las inscripciones (que aparecen como listas de cifras, marcas en tubos, gráficas, etiquetas, descripciones, ecuaciones, modelos) que luego van a ser puestas públicas en los documentos.

REPRESENTACIÓN DE LOS OBJETOS DENTRO DE LOS OBJETOS INFORMACIONALES: SU RELACIÓN CON LA INFORMACIÓN Y EL CONOCIMIENTO

Dentro de los objetos informacionales, los objetos se representan como datos. Diremos que son datos informacionales. La relación entre datos, información y conocimiento ha sido tratada desde múltiples opciones teóricas (Zins, 2007).⁶ Como elementos fundamentales de la ciencia de la información es crucial, tanto para el desarrollo de una concepción sistemática de ella como para la construcción de un conocimiento sistemático del campo, formular concepciones sistemáticas de dato, información y conocimiento. Aquí, sin embargo, sólo se diferenciará entre estos tres elementos, sin entrar en todas las discusiones, necesarias, porque superan los límites que se ha impuesto en su fase inicial el Proyecto DESCUBRIMIENTO. Sólo el concepto de dato se trata con una cierta profundidad y se hacen rápidas afirmaciones sobre los conceptos de información y conocimiento.

Se considera el dato la materia prima para la información, y a esta, la del conocimiento. Aquí se asume que la información es el resultado de las relaciones significativas entre los datos presentes en una o en un conjunto de bases de datos que tienen un mismo objeto de referencia: la investigación o la innovación, que los tratamientos de los datos se hacen mediante procedimientos orientados conceptualmente por teorías y que, entonces, los resultados de estos tratamientos pueden, en primera instancia, ser controlados lógicamente, apoyados en las teorías que han dado lugar a los procedimientos. Se dirá entonces que los datos son analizados. Por ejemplo, los datos pueden ser sometidos a tratamientos orientados

⁶ En el estudio de Zins (2007), “Conceptual approaches for defining data, information, and knowledge”, el autor contó con la participación de 45 estudiosos en el campo que hicieron 130 definiciones de dato, información y conocimiento desde diversas bases teóricas: “muchas de las definiciones reflejan un pensamiento sistemático y exhaustivo y están basadas en sólidas bases teóricas y filosóficas. Algunas, sin embargo, son incompletas, inconsistentes, lógicamente falsas y filosóficamente problemáticas. Parece como si las definiciones mostraran que la comunidad académica hablara diferentes lenguas” (p. 487). Una razón de más para definir y volver comunes las nociones que están en la base del Proyecto DESCUBRIMIENTO.

por la estadística, que los hace aparecer bajo nuevas presentaciones, organizaciones, agregaciones o estructuraciones. No obstante, si bien se trata de dar una organización a los datos que ahora aparecen dotados de una nueva estructuración, según relaciones significativas conducidas por los procedimientos, no hay producción de nuevos conocimientos, sino de una información.

Para que aparezca el conocimiento es preciso que el usuario interprete la información desde teorías de tipo económico, sociológico o histórico. Así, se tiene una diferencia entre el análisis de los datos y la interpretación de los resultados. El primero es un proceso interno que pone en juego a los datos y su tratamiento y los procedimientos para obtenerlos, lo que da lugar a la información; entre tanto, el segundo pone en juego conceptos o teorías exteriores a los datos y los resultados y produce conocimiento.

FORMA DE LOS DATOS EN UN OBJETO INFORMACIONAL

Dentro de un objeto informacional, los datos pueden presentarse como una lista estructurada de metadatos, información asociada a los datos, y, en consecuencia, a los elementos que representan, lo cual proporciona un conocimiento tan completo como sea posible de sus características, atributos o propiedades. La identificación y la selección de los atributos inscritos en un dato permiten avanzar en la caracterización del elemento del que es su representación. Estos atributos son los metadatos del dato, información estructurada de sus características que mejoran la comprensión y permiten operar procedimientos para relacionar los metadatos de un conjunto de datos, es decir, de los atributos de los elementos que ellos representan.

La identificación y la estructuración de los metadatos de un dato están precedidas de estudios sobre el elemento por ellos representado. En efecto, el dato no es el elemento mismo, sino una representación, tanto más fiel cuanto más se ha avanzado en su conocimiento, que hace posible preci-

samente producir su representación.⁷ Así, por ejemplo, la representación del elemento *artículo*, como un dato en una base de datos, cuenta con una serie de metadatos que permiten su descripción más o menos precisa, más o menos consistente, debido a todas las *inversiones* realizadas para avanzar en su caracterización, en los esfuerzos teóricos para precisarlo (para darle forma), en los esfuerzos técnicos para poder integrar todos sus atributos explícitamente, según los procedimientos técnicos utilizados, a través de los *softwares* que están en la interfaz de quienes producen el dato primario y hacen su integración en la base de datos. Dicho de otro modo, el avance en la calidad interna del dato está acompañado de las inversiones para darle forma.⁸

En el contexto de un objeto informacional se debe desarrollar un modelo de datos de acuerdo con la temática específica que ponga en relación los objetos y los actores implicados, así como las actividades desarrolladas. En este caso se trata de un sistema de información. En el caso de la investigación, por ejemplo, se consideran las actividades propiamente de investigación, las políticas nacionales en el campo, las temáticas específicas de trabajo, las actividades de investigación de las instituciones, de las

⁷ La representación del objeto no es el objeto mismo, pero puede considerarse como una definición parcial. La noción de definición parcial está presentada en Pawlowski (1980), quien la ha tomado de Rudolf Carnap (1937), y la utiliza para mostrar la estructura lógica de los *concepts con aire de familia* (*concepts with family meaning*), en el sentido de Wittgenstein, es decir, conceptos que no pueden ser definidos en el sentido lógico de una relación de equivalencia del tipo *A es B*, pero sí de una relación de implicación del tipo *Si a, y, b, y c, y, ... entonces B*, donde *a, b, c...* son metadatos o características esenciales del objeto B. Su uso aquí abre la posibilidad de manipular diferentes subconjuntos de características, de inscripciones o de metadatos de cada uno de los datos o de un conjunto de datos. Se podrá hablar de adecuación del dato a la representación del elemento dado por el conjunto de metadatos que lo representan y de definición parcial del elemento.

⁸ Es posible considerar el formulario utilizado para recoger los metadatos como la *forma* del elemento, que corresponde a la definición parcial alcanzada y que se sujeta a cambios según se va avanzando en la determinación de un elemento genérico, *forma* que permite integrar los metadatos específicos de un elemento concreto. Así, siguiendo el mismo ejemplo, se tiene la forma del elemento artículo, y los metadatos concretos de un artículo particular, que constituyen entonces el *contenido*, representan sus características concretas. Otro ejemplo, ahora en el campo de la innovación, es el de empresa innovadora.

unidades de investigación (centros, institutos, grupos de investigación o investigadores individuales) y las relaciones internas y externas en estas unidades.

El diseño del modelo de datos y la construcción práctica de los instrumentos de lo que es el sistema de información de una temática específica deben ser objeto de un programa desarrollado progresivamente,⁹ que cuente con un lugar central que sirva de referencia para la generación de la información básica normalizada, coordine los procesos de integración de los datos primarios y tenga capacidades técnicas para asegurar su operación y para conformar y conservar la memoria informacional, ponerla disponible para los interesados y para generar permanentemente informes orientados a responder a las necesidades de los usuarios privilegiados.

TIPOS DE DATOS EN EL SISTEMA DE INFORMACIÓN

Según el ejemplo de la construcción del sistema de información sobre la investigación nacional, las orientaciones políticas, por una parte, dan sentido a su existencia y a sus finalidades, y, por la otra, permiten identificar los elementos necesarios de la infoestructura para responder a la construcción de la información elaborada, con el fin de (a) conocer la estructuración de la investigación nacional y establecer las características de sus elementos activos y sus relaciones, (b) determinar las dinámicas de la investigación en los diferentes dominios de la ciencia y la tecnología y (c) construir informaciones elaboradas y ponerlas a disposición pública para los diferentes tipos de usuarios, en particular bajo la forma de indicadores.

⁹ Estos programas, por lo regular, son desarrollados *top-down* como proyectos racionalmente orientados por sujetos particulares, como cuando se propone la construcción de un sistema de información para responder a las necesidades de información de una empresa para sostener sus actividades; no obstante, también pueden ser el resultado de un progresivo crecimiento orgánico a lo largo del tiempo a través de la integración de “herramientas producidas por muchos agentes de manera no jerárquica y no coordinada como el generado alrededor de la comunicación científica o, en el internet” (Björk, 2007: 6).

Un resultado importante del objeto informacional es la posibilidad de disponer en el espacio público de las *formas* de representación de cada uno de los elementos que entran en la producción de la información elaborada. Así, la normalización tiene otros efectos que van más allá de construir las formas explícitas que permiten integrar los datos. En particular, las normas tienen la propiedad de ser vinculantes en cuanto:

- a. Sirven como orientación para presentar información básica, a fin de integrarla en las bases de datos. En este sentido, se constituyen en filtros ordenados y estructurados para incorporarla, lo que es aprovechado cuando se necesita hacer agregaciones de elementos del mismo tipo (por ejemplo, la agregación de todos los artículos presentes).
- b. Permiten construir conjuntos de datos que participan de los mismos atributos.
- c. Orientan la acción, cuando, por ejemplo, se escriben documentos teniendo el cuidado de satisfacer todos los metadatos propios de un artículo. Dicho de otra forma, se constituyen en referencias frente a las que es posible hacer ajustes en las actividades que están en la base de la producción del elemento considerado.
- d. Sirven como referencia para evaluar los resultados de las acciones de los otros miembros de la comunidad.
- e. Permiten hacer relaciones y construir indicadores relacionales, pues se requiere que los datos estén normalizados para que relaciones entre elementos disímiles como “inversión en ciencia y tecnología como porcentaje del PIB”, “número de investigadores por cada mil habitantes” y otras más elaboradas tengan sentido.

Si no se sabe a qué se llama un “investigador” o qué se considera una “inversión en ciencia y tecnología”, estos indicadores no tienen sentido, como de hecho es el caso actualmente en la mayoría de los países, sobre todo en los que no hacen parte de la Organización para la Cooperación y el Desarrollo (OCDE). En este sentido, se ha considerado que las normas son bienes colectivos particulares y junto con los datos y su acumulación

en el tiempo hacen parte del patrimonio *informativo* de una sociedad, “en tanto que factor de identidad, elemento estructurante de una cultura y creador de sentido” (Ribault, 1993, citado en Horn, 1999: s. p.).

Ahora bien, puesto que las normas pueden cambiar en el tiempo, los datos pueden dejar de usarse cuando se ha escrito un informe con base en ellos y poco a poco estar dirigidos a alimentar archivos muertos o permanecer en herramientas técnicas que llegan a ser obsoletas; por ello, una de las funciones del objeto informativo es su conservación, a fin de asegurar su permanente disponibilidad.

La memoria de las normas, de sus transformaciones, de los estándares, del espacio de sus usos, de los datos, de la información elaborada, de los procesos para producirla, así como de todos los métodos y procedimientos intermedios para tratarla y ponerla en circulación, que hacen parte del patrimonio informativo, debe estar disponible en todo momento en el objeto informativo.

REFERENCIAS

- Björk, B. (2007). A model of scientific communication as a global distributed information system, *Irinformation Research*, 12 (2). Recuperado el 1 de mayo de 2009, de <http://informationr.net/ir/12-2/paper307.html#ele06>.
- Boltanski, L. y Thévenot, L. (1991). *De la justification: les économies de la grandeur*. París: Gallimard.
- Carnap, R. (1937). Testability and meaning. En B. Cornelius, *Philosophy of science*. New York: McMillan.
- Charum, J. et al. (1999). *Sobre la gestión estratégica de la investigación: un modelo de análisis de las actividades científicas y tecnológicas*. Bogotá: Departamento de Matemáticas de la Universidad Nacional de Colombia.
- Dictionnaire encyclopédique de l'information et de la documentation* (1997). París: Nathan.
- Fayet Scribe, S. y Canet, C. (1999). Histoire de la normalisation autour du livre et du document: l'exemple de la notice bibliographique et catalographique. De la bibliographie générale et raisonnée de la France à la description bibliographique internationale normalisée. *Solaris* (6). Recuperado el 1 de septiembre de 2009, de <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d06/6fayet.html>.

- Horn, F. (1999). Diversité des informations traités par des moyens informatiques, standardisation optimale et acteurs du processus de standardisation. *Communications & Stratégies* (33). Recuperado el 1 de septiembre de 2009, de <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d06/6horn.html>.
- International Organization for Standardization (ISO), (2003). *Stades de l'élaboration des normes internationales*. Recuperado el 1 de septiembre de 2009, de http://www.iso.org/iso/fr/standards_development/processes_and_procedures/stages_description.htm.
- Latour, B. y Woolgar, S. (1988). *La vie de laboratoire: la production des faits scientifiques*. París: La Découverte.
- Pawloski, T. (1980). *Concept formation in the humanities and the social sciences*. Dordrecht: Holanda y Reidel Publishing Company.
- Vinck, D. (1994). La donnée scientifique: outil scientifique et outil de gestion. *Environment et Société*, 75-87.
- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and technology*, 58 (4), 479-493.

Capítulo 2

Metadatos de una bodega de datos para descubrir conocimiento

Claudia Jiménez Ramírez
Fernán Villa Garzón
Martín Rico Herrera

En relación con los sistemas y procesos de innovación, en Colombia se han venido acumulando múltiples datos procedentes de fuentes diversas y en formatos variados, que sugieren la posibilidad de empezar a explorar alternativas de integración, sistematización y métodos de análisis. Entre dichas fuentes se pueden mencionar las encuestas colombianas de innovación (1996, 2005 y 2008), la Encuesta de Innovación y Desarrollo Tecnológico en Bogotá y Cundinamarca (2005), la Encuesta Anual Manufacturera (particularmente desde 1992), la Primera Encuesta de Innovación en el Sector Servicios (2006), la Encuesta Anual del Comercio y la base de datos ScienTI Colombia, que almacena información sobre los grupos de investigación y los investigadores del país.

Sin embargo, a pesar de los esfuerzos e inversiones que se han realizado para recoger tales datos y sistematizarlos, en el país aún existe un alto grado de desconocimiento de los sistemas y procesos de innovación y desarrollo tecnológico. Es preciso hacer un esfuerzo mucho mayor para convertir

los datos “crudos” en conocimiento útil para la toma de decisiones y el establecimiento de nuevas políticas y estrategias que permitan fortalecer o aumentar la competitividad de nuestro país en el ámbito mundial.

La falta de un análisis más riguroso y profundo de la información recopilada sobre la innovación y temas relacionados como la investigación y el desarrollo tecnológico se debe, entre otros factores, a la carencia de un conjunto de programas y herramientas informáticas lo suficientemente potentes y amigables para que los interesados (investigadores y entes del gobierno) puedan, por sí mismos y en forma interactiva, acceder a los datos disponibles de varias fuentes, relacionarlos, describir los objetos de su interés, resumir datos, descubrir relaciones entre variables y comprobar hipótesis y supuestos, sin necesidad de tener que recurrir a expertos en paquetes o lenguajes de programación y en las técnicas multivariantes empleadas para el análisis de los datos.

El dominio de los lenguajes de programación y la aplicación de técnicas sofisticadas requieren tiempo y recursos que hacen que el proceso de análisis de los datos sea ineficiente no sólo por los costos, sino por los tiempos de respuesta a cada una de las consultas o preguntas que se podrían plantear en un momento determinado.

Una posibilidad ampliamente usada ha sido recurrir a los lenguajes de consulta y a bases de datos, como el lenguaje de consulta estructurado (SQL, por su sigla en inglés), considerados más sencillos y de mayor potencia expresiva que los lenguajes de programación procedimental como el C++ o el Java, por ser declarativos. Estos lenguajes de consulta permiten especificar qué se requiere, sin decir cómo.

Dicha sencillez ha permitido crear interfaces humano-máquina (HCI, por su sigla en inglés) a bases de datos, de alta amigabilidad con los usuarios finales (como la que nos ofrece Google o cualquier otro buscador de páginas web). Sin embargo, estas todavía son muy limitadas para analizar información y descubrir nuevo conocimiento, puesto que el lenguaje de la interfaz no tiene la riqueza semántica suficiente para plantear todos los interrogantes que se le pudieran ocurrir a los consultores de los datos. Sólo unos pocos sistemas gestores de bases de datos, como el SQL Server de Microsoft u Oracle, permitirían, gracias al módulo de *minería de datos*,

ofrecer maneras de agrupar los datos y de establecer relaciones de asociación entre los objetos de interés.

En el Proyecto **DESCUBRIMIENTO** hemos buscado avanzar en el desarrollo de una bodega de datos que integre datos sobre innovación, provenientes de las dos primeras encuestas colombianas de innovación (1996 y 2005), la EAM y la base de datos ScienTI Colombia, accesibles mediante un sistema gestor y una interfaz de consulta. Esto facilitará procesos futuros de integración de nuevos datos a la bodega y su análisis, a fin de consolidar metodologías de *descubrimiento de conocimiento sobre la innovación* en Colombia.

En el presente capítulo se presenta la parte más significativa del trabajo realizado en esa dirección. En particular, el centro de atención es una propuesta de metamodelo para la bodega de datos. Tal metamodelo puede considerarse una declaración explícita y completa de cómo está constituida la bodega de datos —incluidas algunas estadísticas de resumen sobre cada atributo—, con el objeto de no sólo facilitar la búsqueda de los datos requeridos en un análisis determinado, sino de depurar, seleccionar las técnicas de análisis aplicables en cada caso particular e interpretar los resultados.

El capítulo se estructura de la siguiente manera: al principio se introduce el enfoque de inteligencia de negocios, que se adopta para el diseño de la bodega de datos; luego se hacen precisiones conceptuales y metodológicas sobre los sistemas de bodegas de datos y el proceso de descubrimiento de conocimiento en bases de datos (KDD); a continuación se comentan algunas limitaciones de los metadatos actuales, para luego presentar el metamodelo propuesto y enunciar las conclusiones más relevantes.

ENFOQUE DE INTELIGENCIA DEL NEGOCIO

La propuesta realizada en el contexto del Proyecto **DESCUBRIMIENTO** exige un cambio de enfoque respecto al que tradicionalmente ha adoptado el trabajo sobre los datos relacionados con la innovación en Colombia, para aproximarnos más al enfoque gerencial, que busca un mejor aprovechamiento de las potencialidades de la informática moderna. A este enfoque

gerencial se le denomina *inteligencia de negocios* (*business intelligence*, abreviadamente BI) y se basa en el uso de la información almacenada en las bases o bodegas de datos y en otras fuentes de información, para tomar decisiones con diagnósticos más precisos y soluciones más inteligentes (Loshin, 2003).

Con este nuevo enfoque se persigue que las personas con poder de análisis y toma de decisiones puedan reaccionar con mayor efectividad y más rápidamente, para llegar a soluciones más objetivas u obtener mayores ventajas competitivas. El enfoque BI, como se pudiera erróneamente pensar, no se restringe al ámbito empresarial, pues se puede aplicar en cualquier tipo de organización de carácter público o privado y en cualquier área de conocimiento, y de ahí su relevancia.

El enfoque BI se basa en la utilización de un sistema informático que se forma con datos relacionados con la organización y su entorno, extraídos de diferentes fuentes. Dicha información histórica se integra en una bodega de datos para facilitar su manipulación y análisis. Esta colección de datos se conoce bajo el nombre de *bodega*, con el objeto de diferenciarla de las bases de datos tradicionales, que sirven de apoyo a la operatividad cotidiana de una organización, pero poco usadas con el propósito de realizar procesos de descubrimiento de nuevo conocimiento.

El modelo lógico subyacente de una bodega de datos para el acceso y manejo de los datos es el mismo que sirve de soporte a los sistemas de bases de datos actuales: el modelo relacional. En este modelo, los datos se guardan en relaciones o tablas y los datos que describen los datos almacenados, llamado *diccionario de datos* o *catálogo del sistema*, también se deben guardar de la misma manera (Date, 2001).

En las bases de datos relacionales, el propio sistema gestor de bases de datos se encarga de mantener actualizado el diccionario de datos de acuerdo con las órdenes de definición de datos emitidas. En ellas, el usuario final puede consultar los metadatos de la misma manera como se consultan los datos almacenados en la bodega usando el mismo lenguaje de consulta. Esto constituye una gran ventaja, aunque lo que contiene un conjunto de metadatos típico es una descripción tan limitada de los datos que el proceso de descubrimiento de nuevo conocimiento no se facilita para nada.

Los metadatos que usualmente permiten especificar las herramientas informáticas actuales para la construcción de la bodega de datos, como Oracle o SQL Server, se refieren a lo que debe saber el propio sistema para minimizar el tiempo de acceso a los datos. Más recientemente han surgido algunas propuestas de especificación de los metadatos que amplían los metadatos actuales, aunque estas se han concebido para facilitar más que todo el intercambio y la interoperabilidad entre bodegas de datos, sin considerar su utilidad en las actividades que implica el proceso completo de descubrimiento de nuevo conocimiento.

De acuerdo con lo anterior, los metadatos de las bodegas de datos han estado más orientados a las máquinas que a los seres humanos, que es lo que aquí se pretende cambiar, para apoyar el trabajo de los analistas de los datos o investigadores. Se concibió un conjunto de metadatos que pudiera ser considerado una declaración lo suficientemente explícita de los datos de una organización que facilite el proceso KDD, fundamental para poder aplicar el enfoque BI en cualquier ámbito o dominio de aplicación. Adicionalmente, el nuevo metamodelo de una bodega de datos que se propone incluye la información necesaria para que los procesos puedan ser más automatizados, con el objeto de buscar mayor oportunidad del nuevo conocimiento en la toma de decisiones.

SISTEMAS DE BODEGAS DE DATOS

Un sistema informático basado en una bodega de datos (*data warehousing*) constituye la base tecnológica del enfoque BI, pues facilita la consulta, la generación de informes tabulares o gráficos y la creación de cubos e hiper-cubos para el análisis multidimensional de los datos. Los cubos de datos o cubos OLAP (como sigla de *On-Line Analytical Processing*) funcionan de manera similar a los cubos del juego de Rubick, donde se trata de armar el cubo por colores, mientras en la bodega de datos se trata de obtener resúmenes de los datos, considerando múltiples dimensiones o criterios de agrupamiento de los datos. Este tipo de análisis tiene como eje central las mediciones o valores observados de un objeto de investigación (los

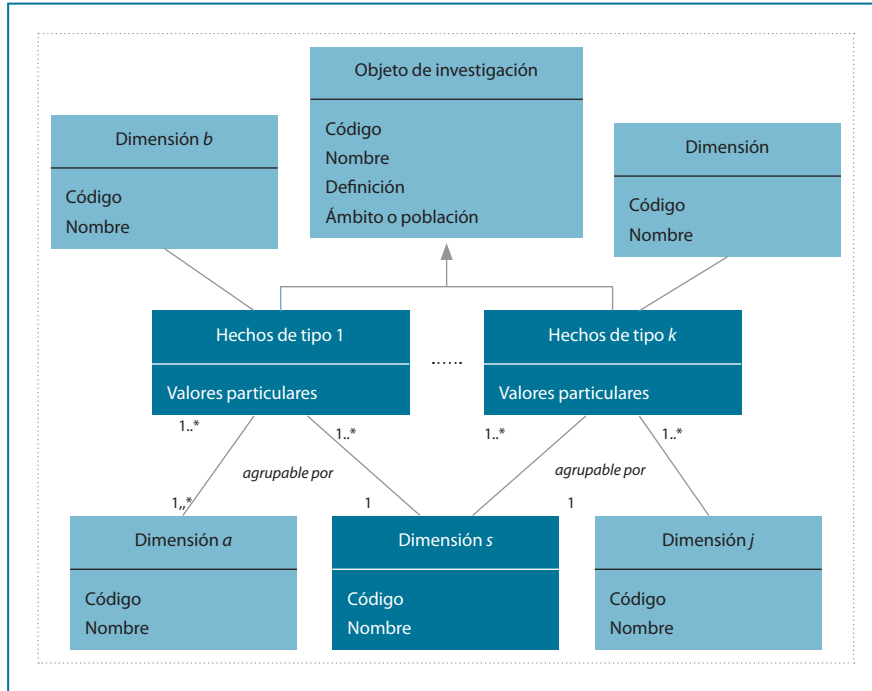
hechos), sobre el cual se identifican y representan las dimensiones que se pueden considerar en su análisis y que permiten resumir o agrupar los datos, especificando diferentes criterios o condiciones de filtrado.

La estructura más simple de una bodega de datos para el análisis de un solo objeto de interés (como las ventas) lo constituye el modelo en estrella, en el cual cada una de ellas corresponde a una dimensión o característica del objeto, que nos permite agrupar los datos o resumirlos en un análisis multidimensional. La estructura *copo de nieve*, un poco más compleja que la estrella, se tiene cuando alguna dimensión se descompone en otras y forma una jerarquía. Esta estructura permite que los agrupamientos o resúmenes de los datos se puedan realizar considerando diferentes granularidades; sin embargo, cuando se tienen varios objetos de interés en los análisis de los datos, como los establecimientos industriales, los grupos de investigación y los proyectos de investigación, la estructura de la bodega de datos irremediablemente se vuelve más compleja y genera una *constelación* o un modelo híbrido como el que se muestra en el Gráfico 1. Como se puede observar, este modelo híbrido tiene una estructura similar a la de una base de datos típica.

En el Gráfico 1 se muestra que los hechos de tipo 1 y los de tipo k comparten la dimensión s , como ocurre con los establecimientos industriales colombianos de la Encuesta Anual Manufacturera y los grupos de investigación de la base de datos ScienTI, pues estos comparten la dimensión “departamento/región”. Aunque las regiones del país a las cuales pueden pertenecer los objetos no tienen el mismo dominio o la misma convención para codificarlos, dicha dimensión permite establecer relaciones o asociaciones entre estos objetos que no se declararon explícitamente, a fin de hacer una transformación a los datos. Por esto el esquema de constelación ofrece mucha flexibilidad, aunque puede aumentar la dificultad para la consulta o manejo de los datos, incluso para el personal informático, si no se cuenta con una buena colección de metadatos. De ahí, la importancia de su buen diseño.

Gráfico 1

Constelación de hechos en una bodega de datos



Fuente: elaboración propia.

Los sistemas de bodegas de datos más recientes ofrecen otras herramientas estadísticas y de la inteligencia artificial que permiten análisis más profundos de los datos. Sirven para descubrir tendencias, patrones o asociaciones en los datos, teniendo el efecto de convertirlos en nuevo conocimiento. Por esto, este nuevo conjunto de herramientas y metodologías es conocido bajo el nombre de KDD.

DESCUBRIR NUEVO CONOCIMIENTO

Descubrir nuevo conocimiento es el proceso de análisis de los datos disponibles en una base o bodega de datos con el objetivo de extraer conocimiento útil para la toma de decisiones (Mitra y Acharya, 2003). Permite

caracterizar los objetos pocos conocidos; descubrir patrones inusuales en los datos, tendencias y asociaciones entre variables; realizar inferencias, o comprobar hipótesis. Así mismo, el proceso KDD es usado para analizar los datos en cualquier área del conocimiento, obtenidos de múltiples fuentes por métodos experimentales, por la captura de señales, de encuestas o del juicio de expertos.

La *minería de datos* es el proceso central del KDD para extraer el nuevo conocimiento. La totalidad de los pasos en el proceso de descubrir nuevo conocimiento son, en su orden (Han y Kamber, 2001):

- *Comprensión del dominio de la aplicación.* Incluye recolectar la información relevante sobre los objetos de interés en el trabajo que se aborda, del dominio del problema y de los supuestos que se cumplen o las restricciones que se tienen.
- *Selección de datos.* Se eligen los datos que se consideran relevantes para un análisis y se integran, si están almacenados en distintas fuentes, como es lo más corriente.
- *Preprocesamiento de datos.* Se necesita para depurar la información, chequear inconsistencias o preparar los datos para la minería. La “limpieza” de los datos consiste en la remoción de ruido (eliminación de errores), en la detección y en el manejo de los datos inconsistentes y faltantes. Además de esto, en algunas ocasiones, los datos deben ser transformados o consolidados en una forma apropiada para la minería. También puede ser necesario resumir la información recolectada, hacer cambios de escala o reducir la dimensión del problema antes de aplicar una técnica de minería de datos en particular. Este proceso es de vital importancia para garantizar la calidad de los datos y la validez del nuevo conocimiento derivado de los datos guardados en una bodega de datos
- *Minería de datos.* Es el proceso donde se aplican los métodos estadísticos o de la inteligencia artificial para encontrar o cuantificar relaciones implícitas o patrones previamente desconocidos. Permite llevar a cabo análisis exploratorios o corroborativos de los datos e incluye una amplia gama de técnicas. Algunas exploran los datos

de forma más autónoma o no supervisada, como en la búsqueda de relaciones de implicación con el algoritmo a priori, y otras son técnicas supervisadas. En este tipo de análisis, a veces, se invierte la dinámica del método científico, puesto que antes de formular la hipótesis se parte de los datos recolectados a través de los años por una organización, y en el método científico se comienza con las hipótesis y luego se diseña el experimento, para recolectar los datos que las confirmen o refuten.

- *Evaluación de los hallazgos.* Este proceso es necesario para validar los modelos construidos o las hipótesis planteadas. En la validación de modelos es frecuente determinar el grado de bondad del ajuste de datos reales, empleando muestras de aprendizaje y datos de prueba.
- *Interpretación y presentación del conocimiento.* En este paso se aplican técnicas para visualizar y representar el nuevo conocimiento de forma tal que permitan su fácil comprensión.

En los procesos de KDD confluyen las disciplinas de la inteligencia artificial y de la estadística, aliadas muy eficaces para robustecer el lenguaje convencional de consulta a bases de datos. Estas disciplinas ofrecen múltiples métodos para resumir los datos o visualizar los modelos descriptivos. También proporcionan una gran variedad de métodos para hallar modelos predictivos, no sólo para predecir valores futuros, sino también para estimar los valores en las variables para otros casos no observados. Por esto el KDD es el proceso fundamental para aplicar el enfoque de BI en cualquier dominio.

LIMITACIONES DE LOS METADATOS ACTUALES EN EL PROCESO DE KDD

El hallazgo, la veracidad y la utilidad del nuevo conocimiento en un proceso de KDD dependen de la calidad y completitud de los datos y sus metadatos, por lo cual es fundamental enfocar el trabajo no sólo en la

consecución e integración de los datos, sino en el uso y disponibilidad de la información correcta y necesaria para producir nuevo conocimiento. Por esto una de las componentes más importantes de la arquitectura de una bodega de datos son los metadatos, que se definen como “datos acerca de los datos”. También han sido definidos como “cualquier cosa acerca de los datos necesaria para facilitar su manejo y uso” (Hartenauer, 2007: 3). Los metadatos describen el significado y las características de los datos en una base o bodega de datos, al igual que sus orígenes y su confiabilidad.

Los metadatos, cuando sirven de apoyo a los mineros de datos en el uso apropiado de los datos contenidos en una bodega de datos, se denominan *metadatos estadísticos* (Westlake, 2006). Los metadatos estadísticos tienen la propiedad de ser útiles no sólo en la selección de las variables o atributos de los objetos para un análisis particular, sino en todas las fases del descubrimiento de nuevo conocimiento: en la captura e integración de los datos, en la fase de limpieza y posibles transformaciones de los datos, en la selección de las técnicas de minería de datos más apropiadas en cada caso particular, en la interpretación y presentación de los resultados y en el intercambio de datos con otros sistemas. Adicionalmente, permiten que los procesos de depuración o transformación de datos puedan ser más automatizados, con lo cual se busca mayor eficiencia y oportunidad del nuevo conocimiento en la toma de decisiones.

En la literatura sobre el tema se encuentran varias propuestas de especificación de los metadatos para las bodegas de datos como el *Common Warehouse Meta-model* (CWM), propuesto por el Comité de Estándares del Object Management Group (OMG, 2003). También existen otros dos modelos reconocidos en el ámbito mundial, uno de ellos desarrollado por MetaData Coalition, llamado *MetaData Interchange Specification* (MDSI), que describe unos metadatos para el intercambio de información (Brock, 2002).

El otro es el dado por la propuesta de Microsoft Corporation, llamado *Open Information Model* (OIM), basado en el metamodelo del lenguaje estándar *Unified Modeling Language* (UML), que se usa en la ingeniería de *software* para describir sistemas informáticos (Rumbaugh, Jacobson y Booch, 1998). Sin embargo, la principal limitación de las propuestas

para la representación de los metadatos es que están orientadas a facilitar el intercambio y la interoperabilidad entre sistemas de bodegas de datos, sin considerar la utilidad de los metadatos en las otras tareas que demanda el descubrimiento de nuevo conocimiento, mencionadas en la sección anterior.

Los sistemas gestores de bases de datos (SGBD) actuales cuentan con un conjunto de vistas sobre la estructura de los datos, comúnmente denominada *diccionario de datos*. La principal función de este catálogo es almacenar toda la información de la estructura lógica y física de la base de datos, desde los objetos existentes, el estado de los archivos físicos (*data-files*) y los privilegios de los usuarios para poder ejecutar las peticiones de los usuarios y buscar los planes de ejecución de las consultas más eficientes. Por esta razón, los metadatos están más orientados al propio gestor de bases de datos que a los analistas. Para documentar cada atributo o tabla de la bodega o base de datos es común que sólo se ofrezca la columna “comentario” o “descripción”, donde se puede especificar, en texto libre y sin seguir ninguna convención, un conjunto limitado de propiedades de los datos. Pero como no es obligatorio añadir comentarios, infortunadamente, la mayoría de las bases de datos no contienen esa documentación. Las bases de datos del ScienTI —el Currículum Vitae Latinoamericano y del Caribe (CvLAC) y Grupo Latinoamerica y el Caribe (GrupLAC)—, proporcionadas por Colciencias, por ejemplo, y conformadas por un número apreciable de tablas con nombres poco mnemotécnicos, no poseen ninguna documentación.

Para ilustrar lo anterior, en el Gráfico 2 se presentan los metadatos que SQL Server 2005 de Microsoft permite guardar sobre una columna de una tabla de una bodega de datos. Allí se pueden observar los datos de la columna denominada IG_SST de la Encuesta Anual Manufacturera. Sin embargo, por ejemplo, no se sabe la unidad de medida utilizada en su representación, si se trata de miles de pesos o de dólares. De los detalles presentados para una columna o atributo se aprecia que la descripción es uno de los pocos metadatos que le pueden ser útiles al usuario.

Gráfico 2
Metadatos de una columna en SQL Server, 2005

Propiedades de columna	
(Nombre)	IG_SST
Descripción	INDICADORES GENERALES - SUELDOS Y SALARIOS TOTAL
Determinístico	Sí
<input checked="" type="checkbox"/> Especificación de columna calculada	
<input checked="" type="checkbox"/> Especificación de identidad	No
<input checked="" type="checkbox"/> Especificación de texto completo	No
Indizable	Sí
Intercalación	<base de datos predeterminada>
Longitud	150
No disponible para replicación	No
Permitir valores nulos	Sí
Publicado por combinación	No
Publicado por DTS	No
Replicado	No
RowGuid	No
Suscriptor que no es de SQL Server	No
Tamaño	300
Tipo de datos	nvarchar
Tipo de datos comprimido	nvarchar(150)
Valor o enlace predeterminado	

Fuente: elaboración propia.

La falta de metadatos dentro de la bodega de datos obliga a los investigadores a recurrir a documentos anexos, digitales o en papel, para poder elegir apropiadamente los datos, lo que dificulta y demora considerablemente el análisis que se quiera llevar a cabo. Si se deseara saber, por ejemplo, el porcentaje medio de las ventas destinado a la inversión en investigación y desarrollo (I+D) de los establecimientos industriales antioqueños, primero, se tendría que identificar el nombre del archivo o tabla donde se almacena la Primera Encuesta de Innovación para luego buscar en los anexos cuál columna representa esta variable. Después de una tediosa búsqueda en los anexos se encontraría el nombre (PVI_I+D), pero además se tendría que volver a ellos para buscar el código correspondiente al departamento de Antioquia, en esa encuesta.

Además de la utilidad de los metadatos en el proceso de integración o intercambio de los datos con fuentes externas, estos también son esenciales en su proceso de depuración, porque si los datos no son válidos, tampoco serán las descripciones o las inferencias que se hagan basándose en ellos,

lo que conduce a decisiones erróneas que ocasionan pérdidas de tiempo, dinero o credibilidad. De poco o nada sirve la minería de datos si se aplica a datos de baja calidad.

Para determinar la calidad de los datos se deben considerar diferentes factores: su validez o grado de corrección, su consistencia (si la misma pieza de datos está almacenada en múltiples localizaciones con el mismo valor) y su integridad de dominio o referencial (Cohen, 1998). Los errores de digitación suelen ser frecuentes a pesar del cuidado que se le preste a esta tarea, y son más factibles si no se tienen diseñadas interfaces para la captura de los datos que hagan por lo menos algunas validaciones, en línea.

En la Encuesta Anual Manufacturera se encontraron todos los tipos de error recién mencionados y por eso fue necesario enriquecer la bodega de datos para incluir los valores máximos y mínimos que puede tomar cada uno de los atributos de tipo continuo, con el propósito de saber si existían valores por fuera del dominio. Igualmente, para las variables de tipo discreto se especificó el conjunto de valores posibles, así como las convenciones usadas para representar los valores faltantes, entre otras características.

En el proceso de depuración de datos también se deben analizar los valores extremos, a fin de determinar si son errores o no, pues la ocurrencia de valores extremos puede tener drásticos efectos en la minería de los datos, si no son considerados (Weisberg, 2005). Como ejemplo, en la Encuesta Anual Manufacturera de 1995 se encontró que existía un valor 13 para el código del tipo de establecimiento, cuando el máximo valor posible era 12. En el porcentaje de las ventas destinado a la inversión en I+D no se encontraron valores por fuera del dominio que indicaran valores errados, pero se detectaron dos valores extremos (92% y 87%), que pueden catalogarse como errores de digitación, pues ningún establecimiento industrial destinaría unos porcentajes tan altos de sus ventas a la investigación.

Otro metadato faltante en el diccionario de datos —relevante en el proceso de minería de datos— consiste en saber si los atributos se refieren a variables cuantitativas o cualitativas (ordinales y nominales). Esto porque muchas técnicas no son aplicables a cualquier tipo de datos. Por ejemplo, para las variables cualitativas nominales no tiene sentido calcular una estadística descriptiva de la tendencia central, como la media aritmética.

En cambio, sí es plausible calcular la moda o las modas. Tampoco se puede aplicar el algoritmo a priori, diseñado para buscar asociaciones entre variables cualitativas, usarlo con variables numéricas sin antes realizar un proceso de discretización.

En la Encuesta Anual Manufacturera el tamaño de los establecimientos industriales se define de acuerdo con el número de empleados, como una variable cualitativa que usa una codificación que permite catalogarlas en cuatro categorías, siendo: A=entre 20 y 49; B=entre 50 y 99; C=entre 100 y 199, y D=200 o más. Es fácil intuir que estas definiciones del tamaño de un establecimiento pueden cambiar con el tiempo e incluso pueden involucrar otras variables. Debido a esto, los metadatos deben incluir una definición de las convenciones usadas y el registro de los cambios que hayan ocurrido en el tiempo. En cualquier contexto es posible que una dimensión o atributo cambie de dominio o de significado, y eso debe especificarse en los metadatos, para que se tomen como referencia los datos apropiados, en un análisis cualquiera.

Otra característica fundamental para saber cuál técnica elegir en la minería de los datos es saber si la distribución de los datos de una variable tiene un ajuste aceptable a la distribución normal. Es necesario, entonces, implementar algunas de las técnicas existentes, a fin de probar la normalidad de los datos y guardar ese valor derivado como otra característica relevante de los metadatos, para evitar la ejecución del mismo procedimiento en múltiples ocasiones. También es recomendable generar y guardar, como otro metadato, el histograma de frecuencias o el diagrama de cajas y bigotes, gráficas descriptivas muy útiles para determinar la forma de la distribución de los datos.

Además, en la visualización y la representación del conocimiento encontrado igualmente es vital tener una declaración explícita de los metadatos (entre estos las unidades de medida de las variables) si se desea evitar la edición manual de cada uno de los resultados tabulares o gráficos que arroje un análisis. Como ejemplo, en el Cuadro 1 se muestran los resultados de un análisis descriptivo como lo produciría el sistema, sin la intervención humana. Ahí vemos la gran dificultad para entender qué puede significar V1612V y su valor máximo de 0,92.

Cuadro 1

Estadísticas de resumen

	PEC	VI612V	VI613A01	VI613A02
Q1	0	0	0	0
Mínimo	0	0	0	0
Máximo	0,76	0,92	12	70
Q3	0,06	0	0	1
Mediana	0,03	0	0	0

Fuente: elaboración propia.

En los metadatos también es recomendable documentar cada una de las técnicas que se ofrecieron para el proceso de KDD y cuáles se usaron para obtener los modelos con los datos que se dispongan en una bodega de datos. Esto puede servir para otros análisis en un mismo trabajo de investigación o en trabajos futuros. Así mismo, se deberían conservar las fuentes bibliográficas que sirvieron de referencia y las publicaciones o informes derivados del proceso de descubrimiento de conocimiento realizado.

METAMODELO PROPUESTO

De acuerdo con lo recién expuesto sobre las limitaciones de los metadatos de los sistemas de bodegas actuales y los requerimientos para respaldar los procesos de KDD, en el Gráfico 3 se presenta el metamodelo propuesto (un subconjunto de los metadatos), expresado mediante un diagrama de clases UML, que puede ser considerado una declaración explícita de cómo está constituida la bodega de datos para el Proyecto DESCUBRIMIENTO.

En el diseño de metamodelo también se tuvieron en cuenta las consideraciones hechas por Charum, en el primer capítulo de este libro. En el diagrama, un rombo o diamante significa una agregación de elementos y la relación de especialización o generalización se dibuja como una flecha. Una encuesta, por ejemplo, se considera un agregado de capítulos o de preguntas cuando no existan capítulos y una publicación se considera una

clase de producción académica. Así, como se puede observar en el Gráfico 3, en la solución propuesta para el Proyecto DESCUBRIMIENTO, para cada pregunta de la encuesta o medición almacena información útil, por ejemplo, su tipo, las opciones de respuesta, las unidades de medida, entre otras características.

Como en la Encuesta Anual Manufacturera se presentaron cambios en los códigos de la Clasificación Internacional Industrial Uniforme (CIIU), que identifican la actividad económica de las empresas (de la revisión 2 a la revisión 3), en el Gráfico 3 aparece una clase denominada “Vigencia”, que sirve para especificar el significado y los períodos de validez de cada código CIIU, de acuerdo con el número de la revisión o versión.

Por último, se puede apreciar que el modelo estructural de la bodega de datos contempla que los metadatos estén almacenados junto con los datos de los objetos de interés para una organización o grupo de investigación. Esto es ventajoso para consultar o filtrar los datos requeridos en algún proceso de descubrimiento de nuevo conocimiento, puesto que son datos más visibles y, por lo tanto, de más fácil acceso para cualquier usuario.

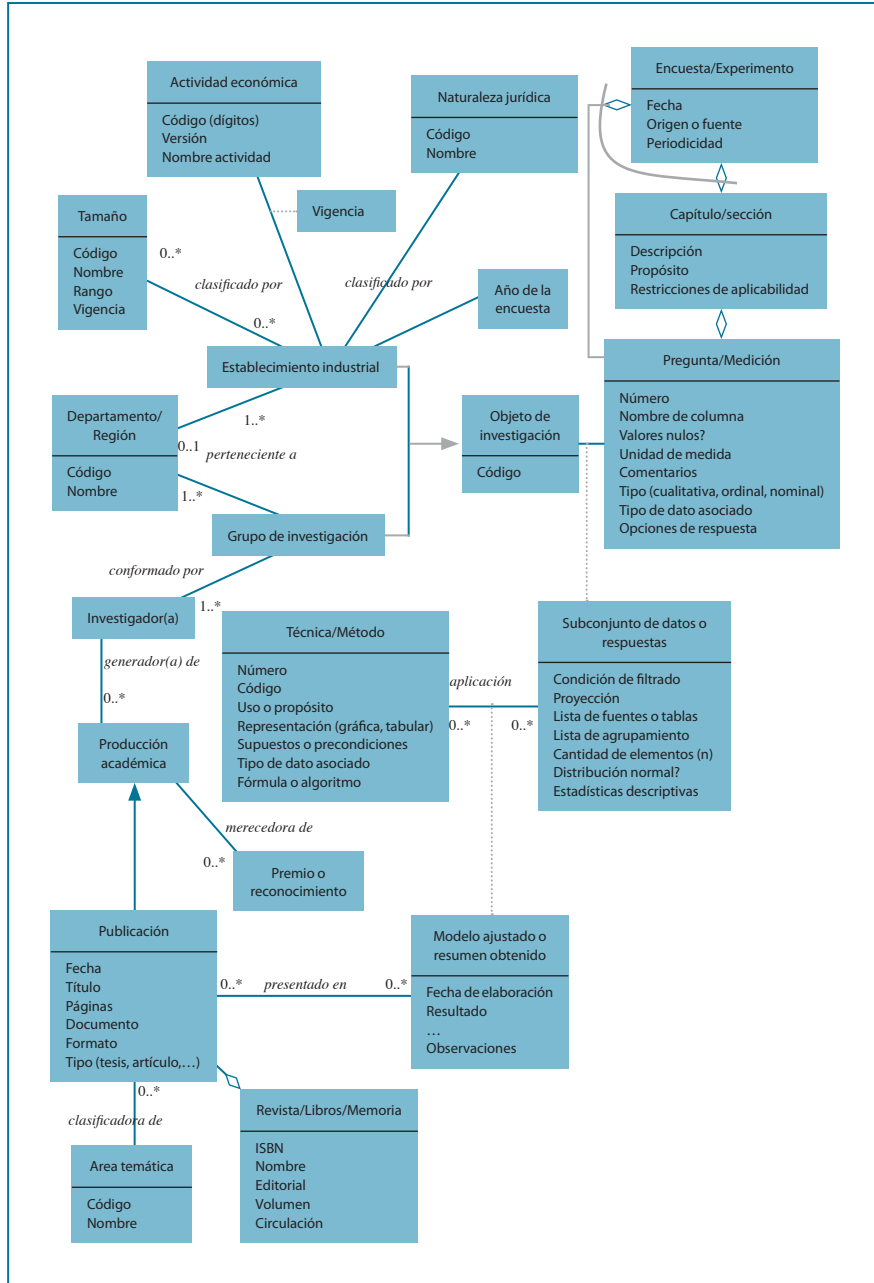
CONCLUSIONES

El análisis de la información recopilada a lo largo de los años permite conocer tendencias en variables de interés, sus relaciones o asociaciones, así como evaluar el impacto de decisiones tomadas en el pasado. Por ello, es necesario facilitar todas las fases que implica el enfoque BI para la toma de decisiones oportunas y basadas en conocimiento, no en juicios subjetivos.

El metamodelo de la bodega de datos propuesto para el Proyecto DESCUBRIMIENTO incluye información básica sobre cada variable de interés, lo cual facilita integrar los datos y limpiarlos, así como visualizar los resultados para su mejor comprensión. Así mismo, se posibilita almacenar en la bodega de datos una representación gráfica de la distribución de los datos y otras estadísticas de resumen (como metadatos), con el fin de facilitarles a los analistas la selección de las técnicas que se van a usar en la minería de los datos. Esto evita la repetición de las mismas tareas previas a

Gráfico 3

Esquema de la bodega de datos sobre la innovación en Colombia



Fuente: elaboración propia.

un análisis por parte de distintos investigadores o en proyectos de investigación futuros. Es decir, el metamodelo propuesto apoya completamente el proceso de descubrimiento de nuevo conocimiento, en todas sus fases.

El enfoque BI propende por la accesibilidad a la información. Debido a esto, lo primero que se debe garantizar en este tipo de herramientas y técnicas será el acceso de los usuarios a los datos; pero se busca ir más allá para que los usuarios puedan seleccionar y manipular los datos que les interesen en un análisis determinado, sin necesidad de la mediación de personal informático o estadístico. Todo esto se logra con unos metadatos completos y bien estructurados.

Por último, dada la complejidad del diseño y la alimentación de la bodega de datos del Proyecto DESCUBRIMIENTO, el modelo presentado en este trabajo está en constante enriquecimiento y mejoramiento, lo cual es necesario para posibilitar la integración de nuevos datos procedentes de aplicaciones recientes de las encuestas o de nuevas fuentes de información.

REFERENCIAS

- Brock, D. (2002). *Survey of languages, specifications and standards for database and network communication*. Recuperado el 31 de julio de 2009, de <http://www.autoidlabs.org/uploads/media/MIT-AUTOID-WH-015.pdf>.
- Cohen, W. (1998). *Integration of heterogeneous databases without common domains using queries based on textual similarity*. Documento procedente de ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '98). Recuperado el 1 de septiembre de 2009, de <http://www.seas.upenn.edu/~zives/03s/cis650/P201.PDF>.
- Date, C. (2001). *Introducción a los sistemas de bases de datos* (7a ed.). México: Alhambra Mexicana.
- Han, J. y Kamber, M. (2001). *Data mining: concepts and techniques*. New York: Morgan Kaufmann.
- Hartenauer, J. (2007). *Introduction to business intelligence: concepts and tools*. Berlín: Verlag Muller.
- Loshin, D. (2003). *Business intelligence and information exploitation, business intelligence*. San Francisco: Morgan Kaufmann Publisher.
- Mitra, S. y Acharya, T. (2003). *Data mining*. Multimedia, Soft Computing and Bioinformatics. New York: John Wiley & Sons.

- Object Management Group (OMG), (2003). *Common Warehouse Metamodel: CWM Specification*. Version 1.1, Volume 1 formal/03-03-02.
- Rumbaugh, J.; Jacobson, I. y Booch, G. (1998). *The unified modeling language reference manual*. New York: Addison-Wesley.
- Trosset, M. (2001). *An introduction to statistical inference and data analysis*. Williamsburg: Department of Mathematics of College of William & Mary.
- Weisberg, S. (2005). *Applied linear regression* (3a ed.). Minneapolis: John Wiley & Sons.
- Westlake, A. (2006). *Proposals for metadata for generic support of statistical modelling in statistical databases*. Recuperado el 14 de octubre de 2009, de <http://www.sasc.co.uk/Guides/index.htm#Managing%20Model%20Metadata>.

Capítulo 3

Gestión de datos e información para el análisis de la innovación a partir de encuestas: aprendizaje y retos

Claudia Nelcy Jiménez Hernández

Martín Rico Herrera

Claudia Jiménez Ramírez

Autores como Burton-Jones (1999) y Gelauff (2003) han señalado que la economía está pasando de la manufactura y la producción de bienes tangibles al manejo de información, la generación de conocimiento y la producción de bienes simbólicos o intangibles, con el empleo creciente de tecnologías de información y comunicación (TIC), que se despliegan como tecnologías de propósito general, junto a nuevas formas sociales basadas en el conocimiento.

En el actual contexto de la llamada *sociedad del conocimiento*, resulta imprescindible un análisis de la información que lleve a la creación, acumulación y asimilación de nuevo conocimiento. En el proceso de gestión de datos e información, la identificación de relaciones entre diferentes situaciones, descritas por los datos recopilados, es una actividad clave para la solución de problemas en las organizaciones, lo que permitirá la optimización de recursos y el respaldo a la toma de decisiones.

Davenport y Prusack (1998) establecen que los datos se convierten en información cuando son ordenados, estandarizados bajo reglas predeterminadas y mostrados de una forma que permitan la comunicación a través de representaciones del lenguaje, como gráficas, tablas, etc. (una discusión más profunda sobre los conceptos de datos, información y conocimiento la presenta Jorge Charum en el Capítulo 1).

Se argumenta que la información es un conjunto de datos al cual se le ha añadido valor a partir del desarrollo de tareas de contextualización, categorización, cálculo, corrección y condensación (Tiwana, 2000, citado en Baker y Badamshina, 2002). Para el procesamiento o estructuración de datos e información, las TIC son una herramienta facilitadora, debido a su capacidad para realizar las distintas operaciones de generación de valor en períodos mucho menores, porque disminuyen considerablemente la probabilidad de error en los cálculos.

Este capítulo tiene el propósito de describir el proceso seguido en el marco del Proyecto DESCUBRIMIENTO, respecto al manejo de los datos e información, provenientes de fuentes como las encuestas colombianas de innovación, la Encuesta Anual Manufacturera y la base de datos ScienTI. Dentro del mencionado proyecto se avanza en la creación de una bodega de datos relativos a la innovación, lo cual permitirá mejorar la disponibilidad de dichos datos e información, así como aplicar diferentes tipos de técnicas para su análisis, como las conocidas bajo el nombre de *descubrimiento de conocimiento en bases de datos* (*Knowledge Discovery in Databases* o abreviadamente KDD).

El proceso que aquí se describe se refiere a la gestión de datos e información y se ha entendido como aquel que incluye operaciones de extracción, manipulación, tratamiento, depuración y conservación de la información adquirida por una organización a través de diferentes fuentes, y que gestiona el acceso y los derechos de los usuarios sobre la misma (Information Management, 2006). Con base en la caracterización de este proceso, se evidencia el aprendizaje generado a lo largo de su desarrollo, así como las falencias detectadas, incluso desde la misma captura de los datos por parte de los organismos responsables, las cuales influyen en la calidad de la información y, por tanto, en la legitimidad de los análisis. A partir de allí, se formulan recomendaciones que se constituyen en retos

tanto para las entidades gubernamentales proveedoras de los datos como para los investigadores en los temas relativos al proyecto.

DESCRIPCIÓN DEL PROCESO DE GESTIÓN DE DATOS E INFORMACIÓN

En el marco del Proyecto **DESCUBRIMIENTO** se ha desarrollado un proceso de manejo de datos e información, en el que se ha estructurado una bodega de datos para integrarlos y facilitar su análisis mediante técnicas estadísticas y de inteligencia artificial, empleando el enfoque de la inteligencia de negocios. Este enfoque, además de utilizarse en entornos gerenciales, se considera apropiado para analizar información de tipo institucional, a fin de diseñar y evaluar políticas y estrategias públicas o para el trabajo investigativo en cualquier área de conocimiento (véase Jiménez, Villa y Rico, 2008; Capítulo 2 de este libro).

En el proceso de gestión de datos e información del proyecto en mención se identificaron cinco etapas: (1) revisión de la estructura y semántica inicial de cada fuente de datos, (2) diseño de la estructura de la bodega de datos y adecuación de datos, (3) enriquecimiento inicial de datos, (4) depuración de errores y (5) documentación del proceso. En los siguientes acápite se describe en qué consiste cada etapa, presentando como ejemplo las características más relevantes de su aplicación a una de las fuentes de datos del proyecto, la Primera Encuesta Colombiana de Innovación. Así mismo, se plantean algunas observaciones respecto a otra de las bases empleadas en el proyecto, partiendo de sus particularidades dentro del proceso de gestión de datos e información.

REVISIÓN DE LA ESTRUCTURA Y SEMÁNTICA INICIAL DE CADA FUENTE DE DATOS

Esta primera etapa del proceso de gestión de datos e información contempló la exploración de las bases de datos y las hojas electrónicas suministradas por el Departamento Administrativo de Ciencia, Tecnología

e Innovación (Colciencias), el Departamento Nacional de Estadística (DANE) y el Departamento Nacional de Planeación (DNP). Se estudiaron los tipos de archivos recibidos con el objeto de describir su contenido y presentación, al tiempo que se consideraron los atributos existentes, el volumen de información, el formato de cada archivo, entre otros aspectos.

En esta etapa se evidenciaron algunas deficiencias en la gestión inicial de los datos, realizada por parte de los proveedores de la información, como su falta de validación o su ausencia, así como algunas inconsistencias. Para mostrar algunos de los hallazgos, en el Cuadro 1 se resumen las principales características de la base de datos correspondiente a la Primera Encuesta Colombiana de Innovación —la Encuesta sobre Desarrollo Tecnológico en el Establecimiento Industrial Colombiano (EDT)—, efectuada por el DNP y Colciencias en 1996, donde se observa el gran volumen de datos, dada la cantidad de establecimientos industriales encuestados y de preguntas que componen la encuesta. La información fue suministrada en libros electrónicos de Excel, clasificada en nueve agrupaciones, las cuales se identificaron como capítulos de la encuesta.

Cuadro 1

Principales características de la base de datos de la EDT

Característica	Descripción
Formato de archivo	Excel
Campos existentes	472 preguntas 885 establecimientos
Volumen de información	417.720 datos (respuestas)

Fuente: elaboración propia con base en la información de la EDT.

Al revisar la EDT se encontraron problemas relativos a la integridad referencial, que consistieron en la identificación de algunas preguntas que no tenían respuesta asociada, así como la situación inversa. De igual manera, se constató la ausencia de datos para relacionar cada establecimiento encuestado con el área geográfica donde estaba ubicado, así como falta de información sobre el tamaño de algunos establecimientos.

DISEÑO DE LA ESTRUCTURA DE LA BODEGA DE DATOS Y ADECUACIÓN DE LOS DATOS

Una bodega de datos se define como una colección de datos históricos, orientados a un dominio, integrados y organizados para dar un respaldo objetivo a la toma de decisiones (Inmon, 2002). El diseño de la bodega de datos para el Proyecto DESCUBRIMIENTO consideró las características de las fuentes de datos disponibles, así como otras variables requeridas para su mejor manejo. En el Capítulo 2 se describe el metamodelo de la bodega de datos diseñada para el presente proyecto.

La forma como se optó por almacenar los datos en la bodega de datos resultó distinta a aquella como estaban dispuestos en las fuentes originales, lo que implicó un importante trabajo manual y técnico para organizarlos de manera que se acomodaran al diseño planteado. Durante este proceso se observó que las bases de datos estaban incompletas o simplemente algunas variables no habían sido consideradas, tal es el caso de las *temáticas* para la primera y la segunda encuestas de innovación, las cuales no existían y fueron identificadas para este proyecto en el formulario de cada encuesta. De igual manera, se encontraron problemas asociados con la presentación de los datos y la información —por ejemplo, la redacción de algunas de las preguntas de la EDT—, para lo cual se efectuaron correcciones teniendo en cuenta el formulario de dicha encuesta.

ENRIQUECIMIENTO INICIAL DE DATOS

La bodega de datos del Proyecto DESCUBRIMIENTO se diseñó para atender las expectativas de los investigadores participantes, puesto que se hizo pensando en el manejo de los datos por parte de cualquier usuario, no sólo de conocedores del lenguaje y las herramientas informáticas. Además, se buscó que esta bodega incluyera los diferentes trabajos técnicos e investigativos realizados alrededor de la temática de la innovación y que facilitara una posterior actualización.

Por lo anterior, y con el propósito de ampliar la descripción de los datos originales, se identificaron cuáles deberían ser los metadatos básicos (características o datos acerca de los datos existentes) relacionados con cada pregunta, el tipo de pregunta, las opciones de respuesta, las unidades de medida, las preguntas filtro y su descripción, así como los capítulos y subcapítulos o temáticas en las cuales están agrupadas o pueden agruparse las preguntas y respuestas. Como ejemplo, en el Cuadro 2 se presenta una de las 74 temáticas identificadas en la EDT, con la descripción asignada, es decir, el metadato correspondiente.

Cuadro 2

Ejemplo de metadatos de la base de datos de la EDT

Temática EDT	Descripción
Inversión en proyectos de I+D entre 1993 y 1995	Identifica los recursos invertidos en actividades de I+D (humanos, económicos, tiempo), así como las fuentes de financiación empleadas. También indaga sobre la existencia de laboratorios de I+D y el impacto de las actividades de I+D en la innovación

I+D: investigación y desarrollo.

Fuente: elaboración propia con base en la información de la EDT.

El enriquecimiento de la bodega de datos con los metadatos permite mejorar la gestión de la información y, por ende, ser más efectivos en el proceso de descubrir nuevo conocimiento. Igualmente, le facilita al investigador seleccionar las variables o atributos que se deben considerar, así como elegir las técnicas de minería de datos que deben aplicarse en un determinado análisis. Además de esto, el enriquecimiento de datos ayuda en la visualización de los resultados y, por tanto, en la comprensión del conocimiento minado. Por ello esta fase se convirtió en parte fundamental del proceso de gestión de datos e información, al considerar su utilidad tanto a corto como a largo plazos.

DEPURACIÓN DE ERRORES

En la etapa de depuración de errores, del proceso de gestión de datos e información, se emplean diversas técnicas y procedimientos para identificar

fallas o datos faltantes e implementar acciones correctivas, de manera que se incremente su calidad, como condición sine qua non para la validez de las inferencias obtenidas a partir de dichos datos.

Algunas de las fallas en los datos pueden deberse a la estructura de las propias encuestas. En el caso de la EDT, por ejemplo, se encontró que se formularon dos preguntas para determinar si las fuentes de innovación son de origen nacional o extranjero, en lugar de consultarlo usando una sola pregunta, como se muestra en el Cuadro 3.

Cuadro 3

Ejemplo de la detección y depuración de un error en la EDT

Código	Variable	Cambio o acción	Solución dada
IV416B01	Pregunta del capítulo 4, para conocer si el origen de la fuente de innovación es nacional (=1) o no (=2)	Unificación de las preguntas	¿El origen de la fuente de innovación es nacional (=1) o extranjero (=2)?
IV416C01	Pregunta del capítulo 4, para conocer si el origen de la fuente de innovación es extranjero (=1) o no (=2)		

Fuente: elaboración propia con base en la información de la EDT.

A continuación se describen algunos de los procedimientos seguidos en la depuración de los datos, considerando distintos aspectos: incumplimiento de reglas de integridad, análisis de inconsistencias, verificación de la completitud de los datos y la información y medidas correctivas.

Incumplimiento de reglas de integridad

Las reglas de integridad garantizan una adecuada correspondencia y relación entre los datos, de manera que ello es parte vital para certificar su veracidad. La integridad se maneja en gran medida gracias a las ventajas de los sistemas de almacenamiento, en asocio con unas definiciones establecidas.

Durante el proceso de depuración de los datos, se observó la falta de manejo de reglas de integridad, lo que refleja dificultades en la gestión

inicial de la información por parte de las entidades encargadas de la recolección y almacenamiento, como ya se había mencionado. Estas fallas repercutieron en el trabajo realizado respecto a la construcción de la bodega de datos, puesto que originaron demoras y, en algunos casos, imposibilitaron trabajar con la integridad referencial planteada en el diseño de esta.

El caso más particular ocurrió en la EDT, donde se detectó, empleando las bondades de los sistemas de almacenamiento, que existían preguntas sin respuesta por parte de los establecimientos encuestados o que no estaban contempladas como una pregunta dentro de los archivos donde se encuentran las correspondientes respuestas; sin embargo, lo más destacado fue encontrar respuestas sin una pregunta asociada, es decir, en los archivos respectivos no existían preguntas definidas para dichas respuestas.

Análisis de inconsistencias

El análisis de inconsistencias busca identificar discrepancias en códigos, nombres o relaciones entre los datos. Para lograrlo en el Proyecto DESCUBRIMIENTO se cumplieron dos pasos: el primero, basado en la revisión manual de los datos, y el segundo, posterior a la carga de la información en la bodega de datos, realizando consultas mediante Lenguaje de Consulta Estructurado (SQL, por su sigla en inglés), que constituyen tareas de minería de datos. Para el caso de la EDT, en el primero de estos pasos, mediante la revisión manual de los datos se detectaron inconsistencias como:

- Existencia de un código para identificar el tamaño de los establecimientos que no estaba definido (se encontró que en realidad dicho código se asociaba con los establecimientos que no respondieron a la pregunta sobre su tamaño).
- Códigos originalmente creados para referirse a respuesta negativa, pero que se usaron para expresar otras circunstancias, como la ausencia del dato en el establecimiento.

Luego de la carga de los datos de la EDT a la bodega de datos, se hicieron consultas con sentencias SQL, y a través de ello se evidenció la existencia de preguntas sin respuesta asociada y viceversa, lo cual fue señalado previamente en este documento.

Verificación de la completitud de los datos y la información: valores atípicos

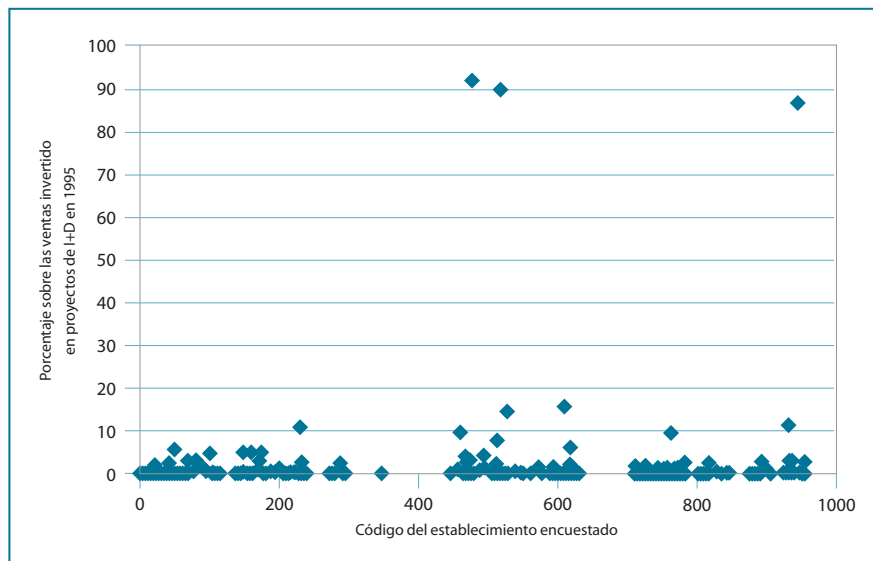
En etapas previas del proceso de gestión de datos e información se estableció que existían problemas tanto en la recolección de datos como en su transferencia al grupo de investigación, que ocasionaron la pérdida de cierta información relevante. Se comprobó que en la base de datos de la EDT, por ejemplo, no existían datos que permitieran relacionar cada establecimiento encuestado con el área geográfica donde se ubica o conocer el tamaño de algunos de los establecimientos industriales. Esta información es muy relevante para los análisis comparativos que demanda la caracterización de la innovación en el país y que quedó, infortunadamente, incompleta.

Por otro lado, una manera de detectar errores en los datos de un atributo cuantitativo consiste en determinar los valores extremos para saber si existen valores por fuera del dominio, puesto que el tipo de dato puede ser insuficiente para restringirlos. Los valores extremos pueden ser correctos aunque atípicos y, por eso, se deben analizar para determinar si corresponden a la realidad y para tomar medidas correctivas cuando no sea así.

La ocurrencia de valores extremos puede tener drásticos efectos en la minería de los datos, si no son considerados (Weisberg, 2005). A manera de ejemplo, en las respuestas a la pregunta de la EDT sobre el porcentaje de las ventas destinado a la inversión en proyectos de I+D se detectaron algunos valores extremos (92%, 90% y 87%) que pueden catalogarse claramente como errores de digitación, tal como se aprecia en el Gráfico 1.

Gráfico 1

Diagrama de dispersión con valores atípicos



Fuente: elaboración propia con base en la información de la EDT.

Medidas correctivas

En el proceso de KDD se proponen algunas acciones para manejar los inconvenientes o errores en los datos. Estas acciones se realizan en la etapa de preprocesamiento de los datos, conocida como limpieza de datos (*data cleaning*) y se presentan en el Cuadro 4.

Cuadro 4

Acciones incluidas en la limpieza de dato

Denominación	Acciones incluidas
Detección de ruido	Análisis de distribución de datos con histogramas Análisis de clúster Funciones de regresión
Manejo de inconsistencia de datos	Uso de dependencias funcionales conocidas Revisión del proceso de integración de diferentes fuentes, donde frecuentemente se manejan distintos nombres para el mismo atributo

Continúa

Denominación	Acciones incluidas
Manejo de datos ausentes	Uso de la media del atributo Uso del dato más probable basado en los datos existentes (árboles de decisión, regresión bayesiana)

Fuente: Goharian y Grossman (2003).

DOCUMENTACIÓN DEL PROCESO

Como etapa final del proceso de gestión de datos e información se incluye aquella que permite generar metadatos semánticos para la bodega de datos del Proyecto DESCUBRIMIENTO, los cuales asignan un significado a la información. Si bien es la última etapa descrita, no significa que sólo se desarrolle al final del proceso, ya que la documentación se debe ir generando en etapas previas e, incluso, en este caso, algunos de tales documentos fueron provistos por los organismos gubernamentales encargados de las encuestas analizadas en este proyecto.

Como ejemplos de la documentación generada pueden mencionarse los informes y boletines del DANE y Colciencias sobre las encuestas de innovación y la plataforma ScienTI, las ponencias y artículos preparados por los investigadores en el marco del Proyecto DESCUBRIMIENTO, así como este mismo documento. Adicionalmente, se incluyen dentro de la documentación las guías y manuales sobre la creación y manejo de la bodega de datos de la innovación en Colombia, formulados con el fin de que todos los integrantes del proyecto y demás interesados puedan acceder de forma sencilla a dicha bodega.

PARTICULARIDADES DE LA PLATAFORMA SCIENTI

Las bases de datos de investigadores (CvLAC) y grupos de investigación (GrupLAC) se encuentran en la plataforma ScienTI, sistema de información coordinado por Colciencias sobre investigadores y grupos de investigación colombianos en ciencia, tecnología e innovación. Dichas bases

hacen parte de la información manejada en el Proyecto DESCUBRIMIENTO, y en el proceso de gestión de estas bases se identificaron algunas particularidades que ameritan ser reseñadas en este capítulo.

Las bases de datos de ScienTI, con información actualizada hasta el 15 de septiembre de 2008, fueron suministradas directamente por Colciencias al equipo investigador de este proyecto en el sistema Oracle. Por esta razón fue necesario migrarlas al sistema SQL Server, donde se manejaron las demás bases de datos. Infortunadamente, estas bases no se recibieron completas, por lo cual, luego de varios intentos por darles claridad a los datos allí contenidos, el equipo del proyecto desistió de obtener algún análisis sobre la innovación en Colombia a partir de estas fuentes.

La información faltante hace referencia al modelo entidad-relación, fundamental para conocer la estructura y el esquema conceptual de cada base de datos. Adicionalmente, no se contó con diccionarios de datos que suministraran el significado de cada uno de los campos de las bases. Por tanto, fue imposible intentar reconstruir los modelos entidad-relación, más aún si se tiene en cuenta la cantidad de cuadros contenidos en cada base (por ejemplo, la base de datos CvLAC tiene 98).

Así mismo, el hecho de que en sus inicios la plataforma ScienTI haya tomado como modelo un sistema similar desarrollado en Brasil implica que aún se incluyan bastantes palabras en idioma portugués, que pueden incrementar la dificultad para el manejo de las bases. Esto, junto con que en los cuadros y campos en ocasiones no se utilizan términos de fácil comprensión o poco nemotécnicos, dificultó todavía más cualquier posible análisis. En el Gráfico 2 se muestra un ejemplo de los cuadros contenidos en estas bases de datos y los nombres dados a los campos.

Por otra parte, debido a la manera como la plataforma ScienTI captura la información, las bases de datos de investigadores y grupos de investigación contienen datos cuyos análisis pueden presentar problemas de validez. Por ejemplo, cabe mencionar el caso de los nombres de instituciones en las cuales han estudiado o están afiliados los investigadores, que en un primer momento pueden ser seleccionados de un listado que muestra el

sistema, pero que proviene de la información digitada libremente por otros investigadores. Entonces, si la persona que está llenando su hoja de vida no encuentra una coincidencia en la lista desplegada, se le permite registrar libremente el nombre de la institución donde estudió o donde trabaja, lo cual ocasiona que haya diferentes formas de identificar una misma entidad.

Gráfico 2

Ejemplo de los cuadros contenidos en la base de datos de GrupLAC

Table Name	Fields
DW_RE_GRUPO_GR_RH	PK, NRO_ID_GRUPO, NRO_ID_CN PQ, NME_GRUPO, STA_LIDER_GRUPO, TPO_RH, HRS_DEDICACION_GRUPO, MES_INICIO_ATUACION, ANO_INICIO_ATUACION, MES_FIM_ATUACION, ANO_FIM_ATUACION
AUX_TEMP_CRITICAS	NRO_ID_GRUPO, TXT_COMENTARIO
DW_RE_PROD_REVISTA	COD_ISSN, COD_SIR, TPO_SIR, NME_SIR
DW_RE_PRODUCTO_GR	PK_PRODUCTO, NRO_ID_GRUPO, SEQ_PRODUCION, COD_SECTOR_PADRE, NME_SECTOR_PADRE, COD_SECTOR, NME_SECTOR
AUX_PRODUCTO_RH_GR	SEQ_PRODUCION, NRO_ID_CN PQ, SEQ_ORDEM_AUTORIA
DW_EN_RECURSO_HUMANO_GR	NRO_ID_CN PQ, COD_RH, NME_RH, NME_PRI_SEG_RH, NME_PRIMEIRO_APELLIDO, NME_SEGUNDO_APELLIDO, TPO_NACIONALIDADE, CPF_RH, DTA_NASCIMENTO, COD_PAIS_NASC, NME_PAIS_NASC, COD_DEPTO_NASC, NME_DEPTO_NASC, NME_CUIDAD_NASC, COD_PAIS_RESID, NME_PAIS_RESID, COD_DEPTO_RESID, NME_DEPTO_RESID, COD_CUIDAD_RESID, NME_CUIDAD_RESID, TXT_CUIDAD_RESID, COD_PAIS_INST, NME_PAIS_INST, COD_DEPTO_INST, DW_EN_RECURSO_HUMANO_GR, NME_DEPTO_INST, COD_CUIDAD_INST, NME_CUIDAD_INST, TXT_CUIDAD_INST
DW_EN_GRUPO_GR	NRO_ID_GRUPO, NME_GRUPO, ANO_FORMACION, MES_FORMACION, TPO_CONVO_GRUPO_02, TPO_ESTADO_GR, NRO_ICOL, SGL_LIF, DSC_LIF, COD_MUNICIPIO, NME_MUNICIPIO, COD_INST_LINO, NME_INST_LINO, COD_PAIS_LINO, NME_PAIS_LINO, COD_INST_DOS, NME_INST_DOS, COD_PAIS_DOS, NME_PAIS_DOS, COD_INST_TRES, NME_INST_TRES, COD_PAIS_TRES, NME_PAIS_TRES, COD_GRAN_AREA, NME_GRAN_AREA, COD_AREA, NME_AREA, COD_PROGRAMA_PRINC, NME_PROGRAMA_PRINC

Fuente: bases de datos de la plataforma ScienTI, suministradas por Colciencias.

Ocurre una situación similar cuando cada persona registra los proyectos en los que ha trabajado. Si varios investigadores han participado en el mismo proyecto, cada uno de ellos lo deberá digitar al momento de diligenciar su propio CvLAC (hoja de vida), con la posibilidad de que haya cambios por adición o eliminación de alguna palabra, e incluso por cuestiones de ortografía. Este tipo de fallas podrían evitarse si la interfaz de captura de datos presentara listados de opciones normalizados.

APRENDIZAJE A PARTIR DEL PROCESO DE GESTIÓN DE DATOS E INFORMACIÓN DESARROLLADO

En primer lugar, se evidenció que el proceso de gestión de datos e información es complejo y que para su ejecución se requiere tanto de habilidades y conocimientos técnicos en el campo informático como de vínculos con el contexto de la información que se está gestionando; en este caso, con el tema de la innovación y la gestión tecnológica, a fin de garantizar la correspondencia entre los datos y la realidad que se intenta abstraer a través de ellos.

En segundo lugar, es fundamental que los datos que se constituyen en los insumos del proceso de KDD sean de calidad, lo cual justifica la inversión requerida en recursos humanos, económicos y de tiempo para su gestión. En este orden de ideas, si se minimizan las fallas tanto en la captura de los datos como en su presentación inicial, el proceso para gestionarlos se facilitará y agilizará en gran medida.

En tercer lugar, debido a los interrogantes que pueden surgir en el proceso de gestión de los datos, los cuales pueden no ser tan fácilmente solucionables, se hace indispensable el contacto permanente con los proveedores de dichos datos, que favorezca el conocimiento de aspectos clave para su comprensión y manejo por parte de terceros; en este caso, los investigadores del Proyecto DESCUBRIMIENTO.

Teniendo en cuenta que, como se mencionó, para el Proyecto ha sido necesaria la participación tanto de expertos en el campo informático como en el tema de la innovación y la gestión tecnológica, se evidenció que si desde las primeras etapas del proceso de gestión de datos e información se hubiera contado con una interfaz de usuario para acceder a la bodega de datos, se hubiera facilitado la participación de otras personas diferentes a los técnicos y así se habría agilizado el proceso.

Por último, cabe anotar que en la gestión de datos e información, particularmente en la generación de una bodega de datos, se observa que se trata de un desarrollo continuo, es decir, dicha bodega es dinámica, lo cual implica que haya ajustes permanentes en la estructura de los datos que se van cargando.

RETOS EN LA GESTIÓN DE DATOS E INFORMACIÓN SOBRE INNOVACIÓN EN COLOMBIA

Con base en la descripción del proceso de gestión de datos e información desarrollado en el marco del Proyecto DESCUBRIMIENTO y en la identificación del aprendizaje logrado a través de dicho proceso, es posible sugerir cambios que se relacionan con el mejoramiento o definición de nuevos sistemas de recolección de los datos, mayor coordinación entre las entidades involucradas y capacitación y asesoría para la interpretación de los datos. Por tanto, se formulan a continuación varios retos orientados a elevar la calidad de los datos manejados y a garantizar la validez de los análisis resultantes.

El primero de estos retos se formula para las entidades responsables de la captura de los datos sobre la innovación en el país, y se relaciona con el *mejoramiento de procesos de gestión de dichos datos desde su misma recolección*, más aún si se considera que quienes están interesados en usarlos deben asumir un costo económico para obtenerlos. En consecuencia, se plantea la *generación de interfaces de captura de los datos*, donde se puedan definir reglas de integridad de dominio, referencial, de valores no nulos y demás reglas particulares que requiera una encuesta o conjunto de datos particular, por parte del personal de informática. Este trabajo implica una mayor inversión en las fases tempranas de la gestión de los datos e información, pero se justifica plenamente por la calidad de las descripciones e inferencias que se deriven de ellos.

A manera de ejemplo, en el Gráfico 3 se presenta un formulario de captura de datos, donde los errores de digitación son mínimos, porque no demanda que el digitador recuerde números o convenciones para codificar los valores de las variables. También se puede verificar la validez de datos como la fecha de nacimiento, así como exigir que sea un dato obligatorio para impedir que algún encuestado carezca de él.

Gráfico 3
Ejemplo de una interfaz de captura de datos

The screenshot shows a web-based data capture form titled "Cuestionario". At the top, there is a menu bar with options: Archivo, Edición, Ver, Insertar, Formato, Registros, Herramientas, Ventana, and Adobe PDF. The form contains several input fields and a dropdown menu:

- Estudiante Nro:** 200725569
- Fecha de nacimiento:** 02/08/1991
- Ideología Política:** Radio buttons for Izquierda (checked), Centro, Derecha, and Ninguna.
- Religioso:** Radio buttons for Sí and No (checked).
- Prefieres:** Radio buttons for Día (checked) and Noche.
- Máximo nivel educativo alcanzado por uno de los padres:** A dropdown menu currently showing "secundaria".
- Lo que más me gusta de la universidad:** Input field containing "todo".
- Carrera:** Input field containing "sistemas".
- Idioma para aprender:** A dropdown menu currently showing "alemán".
- País para conocer:** Input field containing "Italia".

The dropdown menu for "Máximo nivel educativo" is open, showing the following options: Primaria, Secundaria (highlighted), Técnico o Tecnológico, Pregrado, and Postgrado.

Fuente: elaboración propia.

Por otra parte, se ha evidenciado la necesidad de *mejorar la disponibilidad y mecanismos de acceso a los datos por parte de las entidades del Estado responsables de su captura y manejo*, teniendo en cuenta que actualmente debe recurrirse a distintas entidades y satisfacer los requerimientos de sus gestiones internas, en muchas ocasiones sin mecanismos bien definidos, para obtener los datos de interés en la cantidad y con la calidad y presentación que estas determinen.

Esto dificulta a investigadores y entidades como las participantes en el Proyecto al que se hace referencia en este documento desarrollar de manera más eficiente sus trabajos de investigación. Por tanto, el mejoramiento que se plantea como reto se refiere principalmente a los sistemas de almacenamiento que se manejan en la actualidad, considerando la posibilidad de centralizar la información y los mecanismos y procedimientos de acceso a estos en las condiciones requeridas por los procesos investigativos.

De igual manera, como ya se anotó, es fundamental contar con una vía de contacto permanente y con la asesoría de los entes proveedores de los

datos. Por ende, el reto es que estas entidades establezcan *un canal directo y ágil de comunicación que permita que los investigadores puedan realizar consultas y recibir orientación en el manejo e interpretación de los datos*, sean estos relacionados con el tema de la innovación y el desarrollo tecnológico en el país o con otras temáticas sobre la realidad colombiana.

Finalmente, en el proceso de gestión de datos e información en el marco del actual proyecto de investigación y proyectos posteriores, se presenta el desafío de *desarrollar herramientas o interfaces que faciliten la interacción de usuarios no técnicos con la bodega de datos*, para participar tanto en el análisis de dichos datos como en su misma gestión, lo cual acerca a los investigadores de diferentes campos a este tipo de repositorios y sus enormes potencialidades, a la vez que mejora y agiliza el proceso de gestión.

REFERENCIAS

- Baker, K. y Badamshina, G. (2002). Knowledge management. En *Management benchmark study* (pp. 1-24). Alabama: Air University, Office of Planning and Analysis. Recuperado el 15 de mayo de 2009, de <http://www.au.af.mil/au/awc/awcgate/doe/benchmark/>.
- Burton-Jones, A. (1999). *Knowledge capitalism: Business, work, and learning in the new economy*. Midsomer Norton: Oxford University Press.
- Davenport, T. y Prusack, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School.
- Gelauff, G. M. (2003). The changing position of firms in the knowledge-based economy. En *Innovation management in the knowledge economy* (pp. 5-28). London: Imperial College Press.
- Goharian, N. y Grossman, D. (2003). *Data preprocessing*. Chicago: Illinois Institute of Technology.
- Information Management (2006). *¿Qué es la gestión de la información?* Recuperado el 31 de mayo de 2009, de <http://informationmanagement.wordpress.com/category/gestion/gestion-de-la-informacion/>
- Inmon, W. (2002). *Building the data warehouse*. New York: Wiley & Sons.
- Jiménez, C. S.; Villa, F. y Rico, M. (2008). *Metamodelo de una bodega de datos para el descubrimiento de conocimiento*. Documento procedente del I Congreso Internacional de Gestión Tecnológica e Innovación, Bogotá: Universidad Nacional de Colombia.
- Weisberg, S. (2005). *Applied linear regression* (3a ed.). New York: Wiley.

Capítulo 4

Relación entre capacidades de innovación tecnológica y el desempeño empresarial y sectorial

Jorge Robledo Velásquez
Fredy Alexander Gómez Jiménez
Juan Felipe Restrepo Arias

Desde las propuestas teóricas de la economía evolucionista y la visión de la empresa basada en recursos y capacidades (*resource-based view of the firm*), se trabaja un grupo de factores fundamentales que recoge las especificidades explicativas de las dinámicas de una organización, bajo el concepto de capacidad. En este contexto teórico, las *capacidades* desempeñan un papel clave para que la organización adapte, integre y reconfigure habilidades organizacionales, recursos y competencias funcionales en un ambiente cambiante, para el logro de sus objetivos misionales (Tece, Pisano y Shuen, 1997). Así, la relación entre capacidades y desempeño organizacional queda planteada como una explicación estructural al comportamiento de las empresas.

En particular, la literatura especializada llama la atención sobre la importancia de aquellas capacidades organizacionales relacionadas con las dinámicas de innovación tecnológica que, por extensión, se denominan

capacidades de innovación tecnológica (CIT). La relación entre este tipo de capacidades y varias formas de desempeño empresarial se está investigando extensamente, dado el papel crítico que hipotéticamente cumplen las capacidades y la importancia de generar procesos de acumulación de capacidades mediante políticas y estrategias adecuadas en el país, el sector y la empresa.

Es en este contexto teórico donde surge la pregunta central que orienta el presente capítulo: ¿están relacionadas las CIT con el desempeño de las empresas y sectores industriales colombianos? ¿Cuáles de dichas capacidades y con qué tipo de desempeño? ¿Qué sectores se destacan, tanto positiva como negativamente, en cuanto a sus capacidades y desempeño? ¿Qué evolución se puede identificar en el tiempo? Hasta ahora, las respuestas en esta dirección han sido buscadas principalmente en el contexto de estudios de caso (véase, por ejemplo, Vargas, Malaver y Zerda, 2003; Robledo y Aguilar, 2007). En este capítulo se propone abordar este tema utilizando los datos de las encuestas colombianas de innovación (EDT I y EDIT II) y buscando evidencias empíricas generales con significancia estadística. Para ello, se recurre al estudio de dos unidades de análisis: las empresas y los sectores. Respuestas en la dirección señalada permitirían no solamente mejorar la comprensión de la innovación en Colombia, sino generar recomendaciones de política industrial y de gestión empresarial, a fin de incrementar la competitividad de las empresas y sectores y su contribución a la generación de riqueza, bienestar y desarrollo.

En particular, aquí interesa explorar la relación entre las CIT y el desempeño empresarial medido desde dos ópticas: una enfocada en el desempeño innovador, el cual está más directamente vinculado con las CIT, y otra enfocada en el desempeño del negocio, relacionado este con el impacto de la innovación en las ventas, las utilidades, la participación en el mercado y las exportaciones.

A continuación se presenta el contexto conceptual y teórico que sustenta la propuesta investigativa y el análisis e interpretación de los resultados. Luego se describe la metodología empleada y se presentan y analizan los resultados obtenidos, para finalizar con las conclusiones y algunas recomendaciones que emergen del trabajo.

INNOVACIÓN TECNOLÓGICA Y CAPACIDADES DE DESEMPEÑO EMPRESARIAL

Según la definición del *Manual de Oslo*, la innovación se refiere a la implementación de un producto, proceso, método organizacional o método de mercadeo nuevo o significativamente mejorado (Organización para la Cooperación y el Desarrollo Económico [OCDE], 2005). En particular, la innovación de productos y procesos cuya novedad o mejora significativa está asociada con cambios tecnológicos se conoce como *innovación tecnológica*. Hasta su segunda edición, el *Manual de Oslo* hacía referencia principalmente a la innovación tecnológica de productos y procesos, y sólo en la tercera edición fueron introducidas al *Manual* las innovaciones organizacionales y mercadotécnicas como objeto de medición. Por ser una categoría más estudiada, el presente trabajo considera exclusivamente la innovación tecnológica de productos y procesos.

Desde la perspectiva de los recursos y capacidades, se entiende que detrás de la innovación como proceso empresarial están las capacidades organizacionales que la hacen posible. El concepto de CIT se refiere a aquellas capacidades genéricas y específicas que posee la empresa para producir innovaciones tecnológicas como resultado de la gestión estratégica u operativa de la organización. Sin embargo, es necesario precisar que este es un concepto complejo, elusivo y con mucha incertidumbre, que es difícil de determinar y cuya medición requiere considerar simultáneamente múltiples criterios de orden cuantitativo y cualitativo aplicados a la organización (Wang, Lu y Chen, 2008).

Tradicionalmente, según Christesen (1995), la capacidad de producir innovaciones tecnológicas había sido asociada a la capacidad de investigación y desarrollo (I+D) de la empresa. Desde este punto de vista, gran parte de la literatura había concentrado su atención en la I+D desde una perspectiva de política y de gestión, como la única actividad capaz de producir innovaciones tecnológicas. Sin embargo, este enfoque tradicional se queda corto frente al amplio espectro de capacidades que se consideran hoy necesarias para la innovación tecnológica. De acuerdo con Christesen, las CIT de una empresa deben comprender no sólo la capacidad de I+D,

sino aquellas capacidades ligadas a áreas funcionales de la cadena de valor y a las fortalezas interfuncionales y generales de gestión de la organización.

En la literatura reciente se identifican múltiples fuentes de capacidades organizacionales directamente asociadas con la innovación tecnológica, dentro de las cuales se destacan las que se relacionan a continuación:

Capacidad de I+D: Yam *et al.* (2004) definen esta capacidad como la habilidad de la firma para integrar la estrategia de I+D, la implementación de proyectos, la gestión de portafolios de proyectos y los gastos de I+D. Por su parte, Sher y Yang (2005) definen la capacidad de I+D como la suma de dos componentes asociados a los insumos que tradicionalmente ha definido el *Manual de Frascati* (OCDE, 2002): la intensidad de I+D y el personal de I+D. El primer componente (intensidad de I+D) está definido como la relación entre los gastos en I+D y el número total de empleados de la empresa; este es un componente frecuentemente correlacionado de manera positiva con las medidas de desempeño innovador de la empresa (*i. e.*, Hitt, Hoskisson y Kim, 1997; Keizer, Dijkstra y Halman, 2002, citados por Sher y Yang, 2005). El segundo componente (personal de I+D) es medido como la relación entre el número de empleados dedicados de tiempo completo a actividades de I+D y el número total de empleados de la empresa.

Capacidad de gestión de recursos: Yam *et al.* (2004) y Wang, Lu y Chen (2008) se refieren a esta capacidad como a la habilidad de la firma para adquirir y asignar apropiadamente capital, experiencia y tecnología a los procesos de innovación.

Capacidad de aprendizaje organizacional: Yam *et al.* (2004) definen esta capacidad como la habilidad del establecimiento para identificar, asimilar y explotar el conocimiento proveniente del ambiente circundante. Desde otra óptica, DiBella, Nevis y Gould (1996) la definen como la habilidad de mantener o mejorar el desempeño basado en la experiencia, y hacen referencia exclusivamente al *learning-by-doing*. Esta capacidad posibilita adquirir conocimiento implícito y explícito, el traspaso de conocimiento y el uso del conocimiento.

Capacidad de planeación estratégica: Yam *et al.* (2004) se refieren a esta capacidad como la habilidad de la firma para identificar las fortalezas y debilidades internas, así como las amenazas y oportunidades externas; formular planes de acuerdo con la misión y visión corporativa, y ajustar los planes para su implementación. Guan y Ma (2003) la definen como la capacidad de formular e implantar diferentes tipos de estrategias que se adapten a los cambios externos para sobresalir en los ambientes competitivos actuales.

Capacidad de producción: tanto Guan y Ma (2003) como Yam *et al.* (2004) definen esta capacidad como la habilidad de la empresa para transformar los resultados de I+D en productos que satisfagan los requerimientos del mercado, integrando los requisitos de diseño y las limitaciones y posibilidades del sistema de manufactura disponible para la empresa.

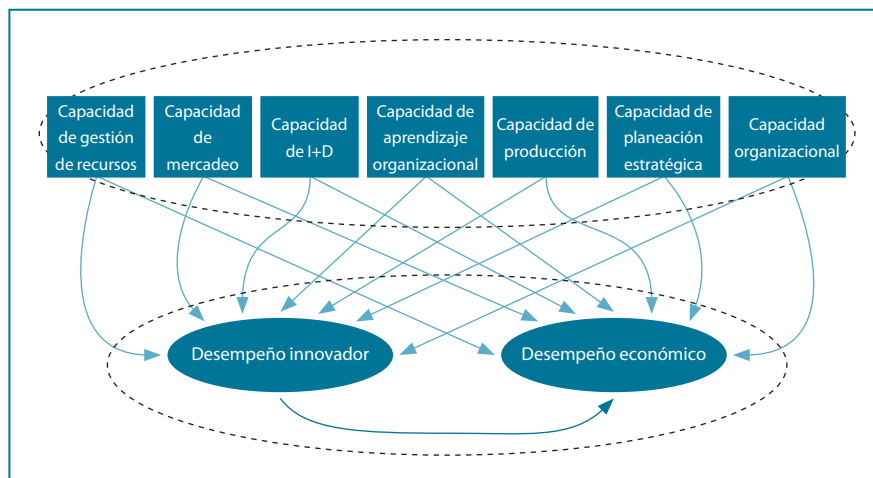
Capacidad de mercadeo: según Yam *et al.* (2004) y Guan y Ma (2003), esta capacidad representa la habilidad de la firma para publicitar y vender productos según la comprensión de las necesidades del mercado (tanto presentes como futuras), el ambiente competitivo, los costos y beneficios y la aceptación de la innovación.

Capacidad organizacional: Yam *et al.* (2004) definen esta capacidad como la habilidad de la empresa para asegurar el mecanismo y la armonía organizacional, cultivando la cultura organizacional y adoptando buenas prácticas de gestión.

Finalmente, es claro que la acumulación de las anteriores capacidades no es un fin en sí mismo, sino más bien un medio para lograr ciertos objetivos empresariales. El logro de tales objetivos corresponde a lo que se denomina aquí desempeño empresarial, definido desde dos ópticas: (1) como desempeño innovador, entendido como el logro de innovaciones tecnológicas de producto y proceso, y (2) como desempeño del negocio, relacionado con el impacto de la innovación en las ventas, las utilidades, la participación en el mercado y las exportaciones. En el Gráfico 1 se presenta el modelo que hace explícitos y relaciona los conceptos introducidos en esta sección.

Gráfico 1

Capacidades de innovación tecnológica (CIT) y desempeño empresarial



Fuente: elaboración propia con base en Guam y Ma (2003), Yam *et al.* (2004), Shery Yang (2005) y Wang, Lu y Chen (2008).

ASPECTOS METODOLÓGICOS

Las relaciones entre las CIT y el desempeño empresarial a partir de los datos de las encuestas colombianas de innovación se exploran aquí aplicando técnicas estadísticas de análisis de asociación entre variables. Para ello existen varias medidas para los casos en que los atributos de la tabla de contingencia presentan sus modalidades ordenadas siguiendo un orden natural, como el coeficiente gamma (γ) de Goodman-Kruskal, la tau-b de Kendall, la tau-c de Kendall y la tau-d de Somers. En este trabajo se utiliza el coeficiente gamma (γ) propuesto por Goodman y Kruskal (1954), para asociar las variables objeto de análisis, dado que su campo óptimo de aplicación es la relación entre dos variables ordinales de dos o más niveles.

Este coeficiente toma valores entre -1 y 1, donde 1 es la perfecta asociación positiva, -1 es la perfecta asociación negativa y 0 es la independencia. Adicionalmente, la significancia estadística de la asociación prueba la hipótesis nula de que la fuerza de la asociación no es diferente de la que sería esperada debido al azar de las muestras aleatorias. Depende del tamaño de

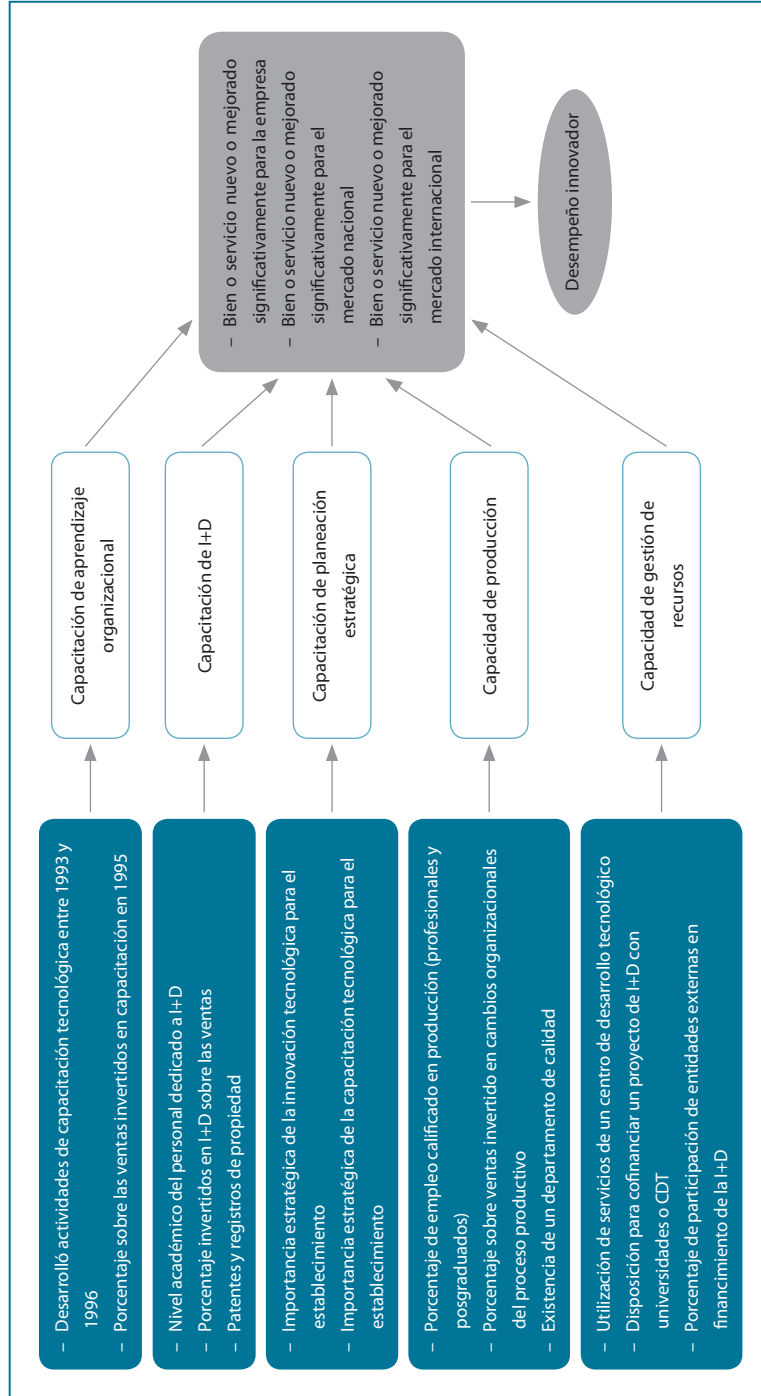
la muestra y otros parámetros. Puesto que gamma tiene una distribución conocida aproximadamente normal para muestras grandes, es posible calcular su error estándar y el grado de significancia (Liebetrau, 1983).

Las variables que se analizaron se obtuvieron de aplicar el marco conceptual y teórico —presentado en la sección anterior y esquematizado en el Gráfico 1— a los datos de las encuestas colombianas de innovación de 1996 (EDT I) y 2005 (EDIT II). Para ello fue necesario identificar las variables que conceptualmente se aproximan más a la medición de las distintas categorías de CIT y desempeño empresarial. En este proceso se distinguieron dos tipos de variables: las variables latentes, que corresponden conceptualmente a las distintas categorías de CIT, y las variables observables, que se obtienen directamente de los resultados de las encuestas y permiten construir las variables latentes. El resultado se presenta en el Gráfico 2 para la EDT I y en el Gráfico 3 para la EDIT II.

Sobre el uso de los datos de las encuestas de innovación, es necesario hacer las siguientes precisiones:

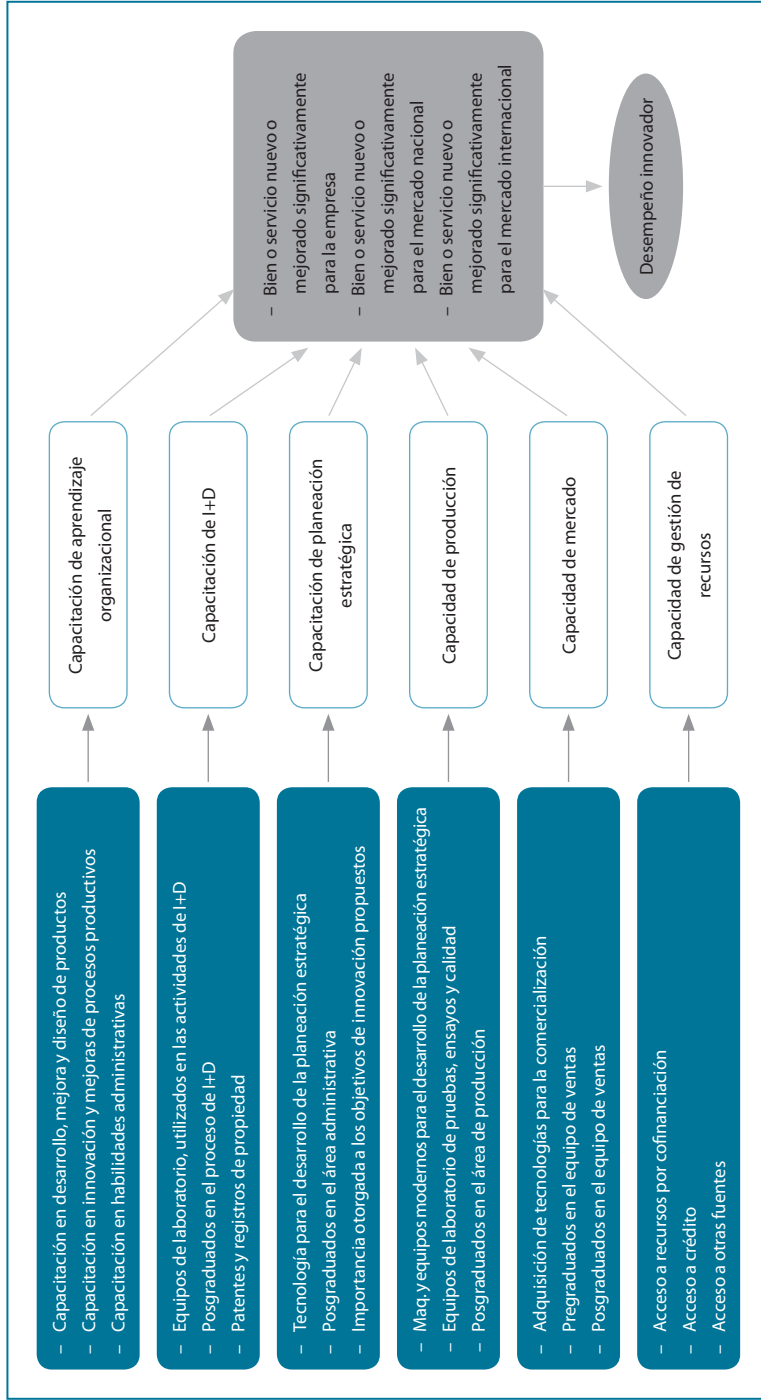
1. Respecto al desempeño empresarial, de la Primera Encuesta sólo se obtuvo una variable observable utilizando el valor de las utilidades (o pérdidas) sobre las ventas del establecimiento en 1995. La Segunda Encuesta no permite el acceso a datos sobre ventas, utilidades (pérdidas) y exportaciones de la empresa, por lo que no es posible construir variables latentes de desempeño del negocio. Una posible solución para esta limitación sería poder cruzar los datos de la encuesta con los de la Encuesta Anual Manufacturera, pero esto no es viable de forma práctica en Colombia.
2. La Primera Encuesta no presenta datos que puedan relacionarse con la capacidad de mercadeo, por lo que esta variable latente queda sin representación en el modelo.
3. Ninguna de las dos encuestas presenta datos que puedan relacionarse claramente con la capacidad organizacional; los datos existentes se recogen mejor en la variable latente capacidad de planeación estratégica.

Gráfico 2
Variables y modelo conceptual para la Primera Encuesta de Innovación (EDT I)



Fuente: elaboración propia.

Gráfico 3
Variables y modelo conceptual para la Segunda Encuesta de Innovación (EDIT II)



Fuente: elaboración propia.

4. A la capacidad de aprendizaje organizacional se asocian variables observables relativas a la capacitación, dimensión que es central en esta capacidad, aunque no suficiente para recoger plenamente el concepto.
5. Se puede observar que no hay una total correspondencia entre las variables observables de la Primera y Segunda encuestas. Sin embargo, dada la suficiente consistencia conceptual de las variables observables respecto a las respectivas variables latentes, es de esperar que esta falta de total continuidad en las variables observables entre una y otra encuesta no genere resultados definitivamente inconsistentes.

Debido a que las encuestas presentan una mezcla de tipos de variables observables, con predominio de variables nominales y escalas dicotómicas, fue preciso redefinir las variables y sus escalas, de forma que fuera posible aplicar las técnicas estadísticas de asociación para variables ordinales. La solución adoptada fue llevar todas las variables observables a una escala ordinal dicotómica, con valores de 2 (para respuestas positivas que reflejan presencia o fortaleza de la capacidad) o 1 (para respuestas negativas que reflejan ausencia o debilidad de la capacidad).

A partir de estos valores se calcularon las contribuciones de todas las variables observables agrupadas para cada capacidad y se obtuvieron resultados para las variables latentes que varían entre 1 y 2, lo cual permitió concluir que aquellas capacidades que poseen un valor cercano a 2 es debido a la presencia o fortaleza de sus componentes, según la medida de sus variables observables. Por el contrario, si los valores de las variables latentes se acercan a 1, son el resultado de ausencia o debilidad de sus componentes.

Dado que la medida de cada capacidad es el resultado de computar entre dos y tres variables observables, los posibles resultados que se pueden obtener se ilustran en el Cuadro 1, junto con la conversión cualitativa ordinal asignada. Este tratamiento a las variables observables permite asignar un valor cualitativo ordinal a cada variable latente y hacer viable la aplicación de las técnicas estadísticas de asociación para variables ordinales.

Cuadro 1

Construcción de variables latentes para medir las CIT

Capacidad medida con dos variables		Capacidad medida con tres variables	
Valores posibles	Conversión cualitativa ordinal	Valores posibles	Conversión cualitativa ordinal
1,0	Bajo	1,0	Bajo
1,5	Medio	1,3	Medio-bajo
2,0	Alto	1,7	Medio-alto
		2,0	Alto

Fuente: elaboración propia.

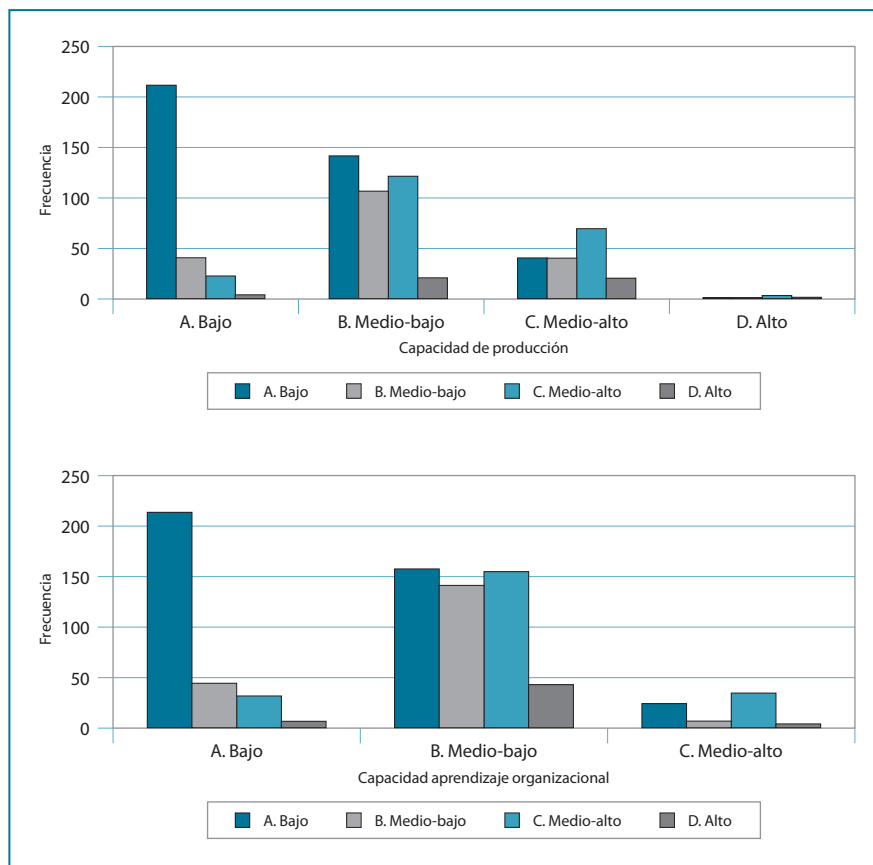
Una vez obtenidas las variables objeto de estudio, estas se sometieron a un análisis de tablas de contingencia y medidas de asociación, que reveló qué tan correlacionadas estaban las CIT con las respectivas medidas de desempeño empresarial. Este análisis fue llevado a cabo con el *software* estadístico SPSS.

ANÁLISIS PARA EL ÁMBITO EMPRESARIAL: RELACIÓN ENTRE LAS CIT Y EL DESEMPEÑO

En cuanto a los datos empresariales, los resultados obtenidos en el análisis estadístico de la relación entre CIT y desempeño empresarial se dividen en tres casos: (1) relación CIT-desempeño innovador, según la Primera Encuesta; (2) relación CIT-desempeño innovador, según la Segunda Encuesta, y (3) relación desempeño innovador-desempeño del negocio, según la Primera Encuesta. El Gráfico 4 muestra dos de las tablas de contingencia para el primer caso, y el Cuadro 2, las medidas de asociación y significancia para los dos primeros casos.

Gráfico 4

Tablas de contingencia entre las capacidades de producción y aprendizaje organizacional y el desempeño innovador para la Primera Encuesta de Innovación



Fuente: elaboración propia.

Los resultados revelan dos elementos de importancia: el primero es que todas las pruebas realizadas fueron estadísticamente significantes (pruebas cuyo $p < 0,05$), lo cual permitió confiar en el resultado dado sobre la relación existente; el segundo es que la relación entre cada una de las CIT es positiva respecto al desempeño innovador de las empresas. En otras palabras, a medida que las empresas muestran mayores (menores) CIT, aumenta (disminuye) el desempeño innovador.

Cuadro 2

Medidas de asociación y significancia para la relación entre las CIT y el desempeño innovador para la Primera y Segunda encuestas de innovación

CIT	1ª Encuesta		2ª Encuesta	
	γ	p	γ	p
Capacidad de planeación estratégica	0,456	0,000	0,844	0,000
Capacidad de I+D	0,461	0,000	0,299	0,000
Capacidad de gestión de recursos	0,332	0,000	0,212	0,000
Capacidad de producción	0,515	0,000	0,310	0,000
Capacidad de aprendizaje organizacional	0,526	0,000	0,228	0,000
Capacidad de mercadeo	0,223	0,000

Fuente: elaboración propia.

Para la Primera Encuesta, la relación más fuerte se presentó con las capacidades de producción y de aprendizaje organizacional, lo cual es un indicio de que el desempeño innovador de las empresas manufactureras colombianas en la época de la Primera Encuesta se basaba fuertemente en el empleo calificado en labores de producción, la inversión en cambios organizacionales en la producción y las fortalezas de sus sistema de gestión de la calidad, por una parte, y la inversión en actividades de capacitación tecnológica, por la otra.

No obstante, ya en esa época era significativo el aporte de las capacidades de I+D y de planeación estratégica, cuyos coeficientes gamma son respectivamente 0,461 y 0,456. Estos resultados sustentan la hipótesis de que la I+D, en el marco de una planeación estratégica adecuada, es una de las principales fuentes de éxito innovador empresarial. Aún así, los resultados del análisis de la Primera Encuesta muestran cómo, por encima de la I+D, la innovación en las empresas colombianas se asociaba más con fortalezas en la producción y la capacitación, lo que revela el carácter incremental de la innovación, su fuerte dependencia de la incorporación de conocimiento y tecnología generado externamente a la empresa y su proximidad a las dinámicas de los sistemas de gestión de la calidad.

Para la Segunda Encuesta también se observó una relación positiva y estadísticamente significativa entre las CIT y el desempeño innovador, aunque las magnitudes del coeficiente gamma tendieron a ser menores que para la Primera Encuesta. Sin embargo, en este caso es bastante notable la asociación positiva muy alta existente entre el desempeño innovador y la capacidad de planeación estratégica, lo cual amerita un análisis futuro más minucioso. Es de observar que también sobresalió la asociación entre el desempeño innovador y la capacidad de producción, seguida muy de cerca por la capacidad de I+D. Esto tiende a confirmar el análisis realizado sobre los resultados de la Primera Encuesta. Sin embargo, a diferencia de sus resultados, la capacidad de aprendizaje mostró una asociación relativamente más débil con el desempeño innovador respecto a las demás capacidades. A continuación aparecen, en su orden pero a un nivel mucho menor, asociaciones positivas con las capacidades de mercadeo y gestión de recursos.

De manera interesante, en los resultados del análisis de ambas encuestas, la capacidad de gestión de recursos presentó la asociación relativa más débil con el desempeño innovador, resultado que amerita, igualmente, un análisis posterior más detallado, pero que, en primera instancia, constató el hecho ya conocido de que los esfuerzos de I+D e innovación de las empresas colombianas están fundamentalmente sustentados financieramente en recursos propios, sin que haya una participación significativa del sistema financiero comercial ni los esquemas gubernamentales de apoyo financiero a estas actividades.

Finalmente, es importante hacer unos comentarios específicos sobre la capacidad de I+D, atendiendo a la importancia que la literatura le otorga a este tema. Al respecto, es notable el alto porcentaje de empresas con una pobre capacidad de I+D, según la Segunda Encuesta (3.224 con baja capacidad y 422 con capacidad media-baja, de las 4.279 que componen la muestra). Por otro lado, se observa también un número significativo de empresas con alto desempeño innovador que presentan bajas capacidades de I+D. Este resultado es consistente con los resultados que asocian de manera notable el desempeño innovador con la capacidad de producción y que señalan un posible carácter incremental y relacionado con los sistemas de gestión de la calidad que presenta la innovación tecnológica en

la industria colombiana. En cuanto al tercer caso (relación desempeño innovador-desempeño del negocio según la EDT I), el Cuadro 3 muestra las medidas de asociación y significancia. Se muestran también en este Cuadro, con fines exploratorios, estas medidas para la relación CIT-desempeño del negocio.

Cuadro 3

Medidas de asociación y significancia para la relación CIT/ desempeño innovador-desempeño del negocio para la Primera Encuesta de Innovación

CIT/desempeño innovador	γ	p
Capacidad de planeación estratégica	-0,49	0,531
Capacidad de aprendizaje organizacional	0,010	0,868
Capacidad de gestión de recursos	-0,34	0,514
Capacidad de I+D	0,190	0,018
Capacidad de producción	0,123	0,016
Desempeño innovador	0,01	0,977

Fuente: elaboración propia.

Se observa que no existe una asociación estadísticamente significativa entre los desempeños innovador y del negocio; sin embargo, hay dos capacidades con una asociación positiva y significativa ($p < 0,05$) con el desempeño del negocio del establecimiento: las capacidades de I+D y de producción. No obstante, queda por analizar si esta es una asociación meramente espuria o si tiene una explicación racional.

ANÁLISIS SECTORIAL: RESULTADOS A PARTIR DE LA PRIMERA ENCUESTA DE INNOVACIÓN

Los sectores industriales estudiados para la Primera Encuesta corresponden a los grupos del Código Industrial Internacional Uniforme revisión 2 (CIU rev. 2) a tres dígitos. Se excluye el sector de refinerías de petróleo

(CIU 353), dada la pequeña muestra que lo representa (dos establecimientos).

ANÁLISIS HORIZONTAL: CIT VS. DESEMPEÑO INNOVADOR POR SECTOR

Este análisis lleva su nombre porque evalúa cada sector de forma independiente respecto a las diferentes capacidades acumuladas. Específicamente, el análisis horizontal muestra las capacidades que mejor se asocian al desempeño innovador de las empresas del sector respectivo.

En el Cuadro 4 se presentan los resultados de evaluar, para cada sector industrial, el coeficiente gamma de Goodman-Kruskal y su respectiva significancia estadística entre las diferentes CIT y el desempeño innovador del sector (líneas horizontales). Como ya se explicó, el coeficiente gamma de Goodman-Kruskal muestra el grado de asociación de dos variables ordinales, variando entre -1 y 1.

Cuadro 4

EDT I: coeficiente gamma del análisis de la CIT vs. el desempeño innovador

CIU	Planeación estratégica		I+D		Gestión recursos		Producción		Aprendizaje organizacional	
	γ	Valor p	γ	Valor p	γ	Valor p	γ	Valor p	γ	Valor p
311	0,46	0,14	0,65	0,06	0,29	0,16	0,832	0,00	0,56	0,01
312	0,00	1,00	0,00	1,00	0,00	1,00	0,54	0,47	0,82	0,25
313	1,00	0,04	0,82	0,15	0,36	0,24	0,964	0,00	0,58	0,14
321	0,60	0,17	0,44	0,30	0,10	0,71	0,754	0,00	0,827	0,00
322	0,70	0,35	0,26	0,55	0,607	0,01	0,19	0,48	0,711	0,00
323	1,00	0,14	-1,00	0,14	0,54	0,17	1,00	0,00	0,44	0,33
324	1,00	0,00	0,37	0,54	0,833	0,00	0,839	0,00	0,58	0,03
331	0,64	0,13	1,00	0,10	0,63	0,06	0,964	0,00	0,68	0,10
332	0,07	0,85	0,09	0,89	-0,10	0,78	0,29	0,39	-0,74	0,03
341	0,53	0,24	0,24	0,57	-0,10	0,69	0,24	0,34	0,542	0,04

Continúa

CIU	Planeación estratégica		I+D		Gestión recursos		Producción		Aprendizaje organizacional	
	γ	Valor p	γ	Valor p	γ	Valor p	γ	Valor p	γ	Valor p
342	0,05	0,90	-0,13	0,82	0,797	0,00	0,740	0,00	0,653	0,00
351	0,826	0,01	0,752	0,01	-0,25	0,34	0,609	0,00	0,715	0,00
352	-0,33	0,27	0,25	0,42	0,413	0,04	0,553	0,01	0,798	0,00
354	1,00	0,23	0,50	0,30	0,20	0,69	0,636	0,03	0,00	1,00
355	0,22	0,69	-1,00	0,11	0,793	0,00	0,897	0,00	0,581	0,03
356	0,24	0,53	0,85	0,00	0,457	0,01	0,655	0,00	0,611	0,00
361	1,00	0,23	-1,00	0,02	1,00	0,22	0,75	0,13	1,00	0,00
362	0,22	0,64	0,36	0,34	-0,08	0,83	0,39	0,20	0,46	0,28
369	0,61	0,10	-0,12	0,73	0,32	0,17	0,595	0,01	0,540	0,01
371	0,72	0,46	0,25	0,55	0,36	0,21	0,43	0,15	0,525	0,06
372	0,47	0,30	1,00	0,27	0,57	0,10	0,647	0,04	0,740	0,02
381	-0,01	0,98	0,25	0,35	0,27	0,22	0,36	0,09	0,07	0,76
382	0,777	0,00	0,696	0,00	0,647	0,00	0,28	0,13	0,619	0,00
383	0,32	0,45	-0,32	0,66	0,33	0,18	0,39	0,13	0,698	0,00
384	0,763	0,04	0,52	0,28	0,508	0,02	0,461	0,03	0,512	0,08
385	1,00	0,12	0,36	0,63	0,39	0,20	0,39	0,19	0,589	0,02
390	0,30	0,39	0,96	0,14	0,615	0,01	0,35	0,18	0,563	0,06

Fuente: elaboración propia.

En el Cuadro 4 se resaltan las pruebas con significatividad del 5% (valor de $p < 0,05$). Además, este es claro en mostrar que las dos capacidades que mejor se relacionan con el alto desempeño innovador de los diferentes sectores industriales son la capacidad de producción y la capacidad de aprendizaje. Es decir, para la gran mayoría de los sectores industriales, una buena capacidad de producción (personal calificado en producción, inversión en el proceso productivo y gestión de la calidad) o de la capacidad de aprendizaje (actividades de capacitación tecnológica a los empleados e inversión en capacitación) está relacionada con un desempeño innovador elevado.

Llama la atención el coeficiente negativo (-0,774) para el sector 332 (fabricación de muebles y accesorios), que indica que una menor capacidad de aprendizaje de las empresas del sector está asociada con un mayor desempeño innovador. Sin embargo, como se verá más adelante, el sector 332 es de los que menores capacidades acumula, por lo que sus resultados deben ser sujetos a un análisis sui géneris.

Se observa también que existen tres sectores cuyo desempeño innovador no está asociado con las CIT. Estos sectores son el 381 (fabricación de productos metálicos), el 362 (fabricación de productos de vidrio) y el 312 (fabricación de productos alimenticios). De nuevo, es un resultado de amerita un examen detallado posterior.

Además, es relevante mencionar la baja relación que posee la capacidad de I+D con el desempeño innovador de los diferentes sectores estudiados. Sólo los sectores 351 (fabricación de sustancias químicas industriales básicas), 356 (fabricación de productos plásticos) y 382 (construcción de maquinaria) presentan coeficientes respectivamente de 0,752, 0,850 y 0,696, positivos y de alta significancia. Como se verá más adelante, el sector 351 es el más sobresaliente en cuanto a capacidades se refiere y, al mismo tiempo, es el sector más innovador.

ANÁLISIS VERTICAL: CIT Y DESEMPEÑO INNOVADOR ENTRE SECTORES

El análisis vertical del Cuadro 4 llevó a examinar los diferentes sectores industriales respecto a la acumulación de capacidades y desempeño, a fin de identificar los sectores líderes frente a los más rezagados. Como cada capacidad se mide con variables ordinales, se aplica un test de hipótesis de comparación de medias de todos los sectores industriales bajo estudio, buscando encontrar evidencia significativa de alguna diferencia entre la media de los sectores.

El Cuadro 5 muestra los resultados de un test no paramétrico (prueba de Kruskal-Wallis) que permite identificar diferencias estadísticamente significantes entre las medias. Como se puede observar, la capacidad de producción, de I+D, de aprendizaje y el desempeño innovador poseen diferencias significantes, por lo menos, en uno de los sectores bajo estudio.

Cuadro 5

EDT I: Prueba de Kruskal-Wallis de comparación

Capacidad/desempeño	Valor de p de Kruskal-Wallis
Capacidad de planeación estratégica	0,296
Capacidad de I+D	0,030*
Capacidad de gestión de recursos	0,055
Capacidad de producción	0,000*
Capacidad de aprendizaje organizacional	0,005*
Desempeño innovador	0,000*

*Significatividad del 5%.

Fuente: elaboración propia.

Un paso posterior al análisis de comparación de medias es conocer entre cuáles medias hay diferencias y entre cuáles no, es decir, comparaciones 2 a 2. Cuando se compara medias a través de la prueba de Kruskal-Wallis, las comparaciones 2 a 2 no suelen estar implementadas en los paquetes estadísticos, aunque se puede utilizar el método de Dunnett T3, por su sencillez de aplicación, para los casos en que el supuesto de igualdad de varianzas no se cumple (lo cual aplica en este caso). A continuación se presentan los resultados arrojados por el *software* SPSS para el análisis 2 a 2 de cada sector económico y para las diferentes CIT que dieron resultados significantes en la diferencia de medias.

Capacidad de aprendizaje organizacional

El Cuadro 6 muestra los resultados con un nivel de significancia del 5%. La hipótesis nula es que las dos medias de los dos sectores industriales son iguales, en contraste con la hipótesis alternativa, que se basa en la diferencia estadísticamente significativa de las dos medias. Aquellas celdas que poseen un signo positivo o negativo se refieren a que existe una diferencia entre el par de sectores en consideración y se lee de la siguiente forma: el sector que aparece en la fila tiene mayor (signo +) o menor (signo -) capacidad de aprendizaje que el sector ubicado en la columna correspondiente a la celda de análisis.

Cuadro 6**EDT I, capacidad de aprendizaje organizacional: Comparación sectorial 2 a 2 con la prueba de Dunnett**

	313	332	342	356	381
313		+			
332	-		-	-	-
342		+			
356		+			
381		+			

Nota: diferencias con significatividad del 5%.

Fuente: elaboración propia.

El Cuadro 6 muestra que el sector más rezagado en acumulación de capacidad de aprendizaje es el 332 (fabricación de muebles y accesorios) frente a los sectores líderes (313, 342, 356 y 381), que corresponden, respectivamente, a la industria de las bebidas, imprentas y editoriales, fabricación de productos plásticos y fabricación de productos metálicos (excepto maquinaria y equipo). Se entiende que los demás sectores están en una posición intermedia, pues no alcanzan a diferenciarse de los demás.

Capacidad de producción

A diferencia de lo que sucede con la capacidad de aprendizaje, los sectores presentan más diferencias respecto a la capacidad de producción (Cuadro 7). En este caso, la diferencia es significativa para 11 sectores de los 27 que se estudiaron.

Cuadro 7**EDT I, capacidad de producción: comparación sectorial 2 a 2 con la prueba de Dunnett**

	313	321	322	323	324	331	332	342	351	352	390
313					+						
321										-	
322									-	-	
323									-	-	

Continúa

	313	321	322	323	324	331	332	342	351	352	390
324	-								-	-	
331									-	-	
332										-	
342										-	
351			+	+	+	+					
352		+	+	+	+	+	+	+			+
390										-	

Nota: diferencias con significatividad del 5%.

Fuente: elaboración propia.

Cabe destacar la posición de liderazgo de los sectores 351 y 352 (fabricación de sustancias químicas industriales básicas y fabricación de otros productos químicos, respectivamente) en la puntuación obtenida en la capacidad de producción sobre los demás sectores. Como sectores más rezagados figuran la fabricación de textiles (321), la fabricación de prendas de vestir (322), la industria del cuero (323), la fabricación de calzado (324), la industria de la madera (331), la fabricación de muebles y accesorios (332), las imprentas y editoriales (342) y aquellas catalogadas como otras industrias manufactureras (390) (deportes, fabricación de plumas y otros instrumentos de oficina o para la pintura, agujas alfileres, joyería falsa).

Desempeño innovador

En la evaluación realizada al desempeño innovador de los diferentes sectores industriales, el predominio del sector de fabricación de otros productos químicos (352) es evidente (Cuadro 8). Este sector posee aquellas industrias que se dedican principalmente a la producción de productos farmacéuticos y medicamentos, pinturas y barnices, productos biológicos a base de materias naturales y producidos sintéticamente, vitaminas y provitaminas, adhesivos, desinfectantes para el hogar, desodorantes de tocador y tintas para imprenta y escribir. Por el contrario, los sectores que se destacan por su bajo desempeño innovador respecto a los líderes

son los de fabricación de alimentos (311), fabricación de textiles (321) y de prendas de vestir (322), la industria del cuero (323) y las imprentas y editoriales (342).

Cuadro 8

EDT I, desempeño innovador: comparación sectorial 2 a 2 con la prueba de Dunnett

	311	321	322	323	342	352
311						-
321						-
322						-
323						-
342						-
352	+	+	+	+	+	

Nota: diferencias con significatividad del 5%.

Fuente: elaboración propia.

ANÁLISIS SECTORIAL: RESULTADOS A PARTIR DE LA SEGUNDA ENCUESTA DE INNOVACIÓN

Puesto que hubo una revisión del CIIU (de la revisión 2 a la 3) entre la EDT I, realizada en 1996, y la EDIT II, realizada en 2005, es importante aclarar que para los análisis relacionados con la Segunda Encuesta cambió el código, el número de dígitos (de tres a dos) y levemente la composición de los sectores. Sin embargo, los análisis comparativos aún pueden ser realizados para las dos encuestas, dado que la mayoría de los sectores permanecen con modificación nula o muy leve.

ANÁLISIS HORIZONTAL: CIT VS. DESEMPEÑO INNOVADOR POR SECTOR

En el Cuadro 9 se resaltan las pruebas con significatividad estadística del 5% correspondientes a la asociación entre las CIT y el desempeño inno-

vador. En primer lugar, es evidente la relación para todos los sectores de la capacidad estratégica con el desempeño innovador de las empresas; siguen en su orden, como variables significativas, la capacidad de producción, la capacidad de aprendizaje y la capacidad de I+D. Al igual que para la EDT I, la capacidad de producción y la capacidad de aprendizaje marcan altos valores de asociación con la innovación.

También es relevante la participación en la EDIT II de la capacidad de I+D, variable que para el análisis de la EDT I no fue relevante. Es posible un síntoma del aumento de los esfuerzos que realizan los establecimientos colombianos alrededor de la I+D, a través de la contratación de personal y la inversión para actividades de I+D. Los sectores en los que la capacidad de I+D está más asociada con el desempeño innovador son la elaboración de productos alimenticios y bebidas (15), la fabricación de papel y cartón y productos de papel y cartón (21), la fabricación de productos de caucho y de plástico (25), la fabricación de otros productos minerales no metálicos (26), la fabricación de productos metalúrgicos básicos (27), la fabricación de maquinaria y equipo no clasificados en otra parte (NCP) (29) y la fabricación de vehículos automotores, remolques y semirremolques (34). Los sectores 15, 21, 26 y 27 poseen, por lo menos, cinco capacidades asociadas al desempeño innovador.

Cuadro 9

EDIT II: coeficiente gamma del análisis CIT vs. el desempeño innovador

CIU	Mercadeo		I+D		Planeación estratégica		Aprendizaje organizacional		Producción		Gestión de recursos	
	γ	P	γ	p	γ	P	γ	p	γ	p	γ	p
15	0,44	0,00	0,45	0,00	0,87	0,00	0,30	0,00	0,45	0,00	0,05	0,40
17	-0,06	0,63	0,33	0,07	0,78	0,00	0,26	0,04	0,18	0,12	0,43	0,11
18	-0,03	0,59	-0,08	0,61	0,91	0,00	0,03	0,39	0,43	0,00	-0,61	0,94
19	0,33	0,02	0,31	0,13	0,84	0,00	0,26	0,03	0,03	0,44	0,34	0,06
20	0,26	0,28	-1,00	0,67	0,89	0,00	-0,26	0,73	0,77	0,05	0,29	0,30
21	0,65	0,01	0,68	0,01	0,92	0,00	0,45	0,05	0,50	0,04	0,42	0,11
22	-0,14	0,84	0,04	0,44	0,79	0,00	-0,16	0,88	0,04	0,40	-0,24	0,75

Continúa

CIU	Mercadeo		I+D		Planeación estratégica		Aprendizaje organizacional		Producción		Gestión de recursos	
	γ	P	γ	p	γ	P	γ	p	γ	p	γ	p
24	0,16	0,07	0,18	0,07	0,78	0,00	0,15	0,07	0,27	0,00	0,33	0,04
25	0,17	0,09	0,32	0,04	0,86	0,00	0,44	0,00	0,34	0,00	0,01	0,48
26	0,56	0,00	0,56	0,00	0,70	0,00	0,19	0,11	0,56	0,00	-0,09	0,57
27	0,76	0,01	0,76	0,00	0,96	0,00	0,63	0,02	0,80	0,00	0,68	0,02
28	0,09	0,33	-1,00	0,92	0,83	0,00	0,20	0,13	0,31	0,05	0,27	0,23
29	0,22	0,06	0,38	0,03	0,82	0,00	0,31	0,01	0,10	0,24	-0,12	0,63
31	0,17	0,21	0,00	0,50	0,84	0,00	0,12	0,27	0,06	0,39	-0,23	0,68
34	0,09	0,34	0,53	0,04	0,77	0,00	0,05	0,41	0,19	0,17	0,18	0,28
36	0,00	0,40	0,11	0,32	0,89	0,00	0,21	0,03	0,17	0,10	-0,03	0,54

Fuente: elaboración propia.

ANÁLISIS VERTICAL: CIT Y DESEMPEÑO INNOVADOR ENTRE SECTORES

El Cuadro 10 muestra los resultados del test de Kruskal-Wallis para las medias de las capacidades. Todas las capacidades, a excepción de la capacidad de gestión de recursos, poseen diferencias estadísticamente significantes, por lo menos, en uno de los sectores bajo estudio. En el Cuadro 11 se presentan, para cada CIT y el desempeño innovador, los sectores líderes y los sectores rezagados. De este cuadro se desprende que los sectores 24 y 15 sobresalen como sectores líderes en la acumulación de capacidades, mientras que los sectores 18, 19, 22 y 28 presentan un rezago significativo frente a los líderes. El desempeño innovador no presentó demasiadas diferencias significantes. Sin embargo, se destaca claramente el sector de fabricación de sustancias y productos químicos (24), tanto en la acumulación de CIT como en el desempeño innovador, igual que en el análisis de la EDT I. Algo semejante ocurre con el sector de elaboración de productos alimenticios (15), que en ambos análisis se destaca por su bajo desempeño innovador.

Cuadro 10

EDIT II: prueba de Kruskal-Wallis de comparación de medias

Capacidad/desempeño	Valor de p de Kruskal-Wallis
Capacidad de mercadeo	0,000
Capacidad de I+D	0,000
Capacidad de planeación estratégica	0,000
Capacidad de aprendizaje organizacional	0,000
Capacidad de producción	0,000
Capacidad de gestión de recursos	0,090
Desempeño innovador	0,000

Fuente: elaboración propia.

Cuadro 11

EDIT II: sectores líderes y rezagados

Capacidad/desempeño	Sectores líderes	Sectores rezagados
Capacidad de mercadeo	24, 21, 15, 22, 25, 26	18, 19
Capacidad de I+D	24, 15	18, 19, 20, 22, 28, 34, 36
Capacidad de planeación estratégica	24, 26	18, 20, 27, 28
Capacidad de aprendizaje organizacional	24	18, 22, 28
Capacidad de producción	24, 15, 25, 26	18, 19, 22, 36
Desempeño innovador	24, 36	15

Fuente: elaboración propia.

CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

En el ámbito empresarial, los resultados del análisis estadístico de las encuestas colombianas de innovación constituyen una evidencia empírica significativa sobre la validez del modelo teórico que explica el desempeño innovador empresarial sobre la base de un conjunto de CIT que, con apoyo en la literatura especializada, en este trabajo se han identificado como capacidad de I+D, capacidad de gestión de recursos, capacidad de aprendizaje organizacional, capacidad de planeación estratégica, capacidad de producción y capacidad de mercadeo.

Por otro lado, es sabido que el trabajo empírico sobre las CIT tiene dificultades no sólo conceptuales, sino de medición, dirección en la que el trabajo propone una aproximación metodológica que hace posible medir las CIT de manera que las variables resultantes sean: (a) consistentes con las propuestas teóricas, (b) medibles con base en los datos de las encuestas de innovación y (c) apropiadas para ser trabajadas con técnicas estadísticas para determinar grado y significancia de las asociaciones entre ellas y el desempeño empresarial.

Es muy relevante el hecho de que tanto el marco teórico como la métrica propuesta para las CIT arrojen resultados fundamentalmente consistentes al ser aplicados a las encuestas colombianas de innovación, que presentan un diseño diferente y cuya aplicación está separada por casi una década. Este resultado es una clara indicación de que las CIT son componentes estructurales de la organización productiva cuando se trata de explicar su comportamiento innovador.

Ciertamente los resultados aquí presentados deberán consolidarse en lo analítico y procedimental, pero los indicios arrojados por la exploración realizada hasta el momento son alentadores y sugieren que el trabajo apunta en la dirección correcta. Por una parte, el trabajo avanza en la verificación de la hipótesis de una asociación positiva y estadísticamente significativa entre las CIT y el desempeño innovador, principalmente como resultado del análisis a nivel de empresa. Por la otra, queda abierta la pregunta sobre la relación entre innovación (en términos de capacidades y desempeño) y desempeño del negocio, pregunta sobre la cual los resultados aquí obtenidos no son concluyentes, debido en parte a la dificultad de contar con los datos necesarios para trabajar válidamente desde el marco teórico propuesto. Además, es razonable suponer que habiendo una asociación positiva y significativa entre la acumulación de capacidades de innovación y el desempeño innovador, esta se manifieste con un retardo temporal más o menos importante, lo cual no es capturado por los datos de las encuestas.

Respecto a la industria colombiana representada en las empresas encuestadas, es notable su debilidad en términos tanto de CIT como de desempeño innovador. Los resultados muestran, de manera permanente, una gran concentración de las empresas en la zona de bajas/media-bajas

CIT y bajo/medio-bajo desempeño innovador, con tendencia a mejorar el desempeño cuando las capacidades aumentan. Esto señala una asociación positiva y estadísticamente significativa que es muy importante en términos de política industrial y gestión empresarial.

En efecto, estos resultados confirman empíricamente, para el contexto empresarial manufacturero colombiano, la importancia de generar procesos de aprendizaje organizacional que propicien la acumulación de CIT con miras a mejorar el desempeño innovador. Teóricamente, esta mejora se vería reflejada en una posición competitiva más ventajosa para las empresas y unas condiciones más propicias para la generación de riqueza, bienestar y desarrollo.

Por otra parte, los resultados del análisis sectorial, a diferencia de los empresariales, no son tan consistentes en cuanto a la asociación entre capacidades y desempeño. Para la Primera Encuesta, las dos capacidades que mejor se relacionan con un mejor desempeño innovador de los diferentes sectores industriales son la capacidad de producción y la capacidad de aprendizaje. Estos resultados son consistentes con el análisis en el ámbito empresarial; sin embargo, llama la atención la relación negativa entre la capacidad de aprendizaje y el desempeño innovador del sector 332 (fabricación de muebles y accesorios), lo que puede indicar una inversión y unas actividades de capacitación con una muy pobre orientación a la innovación y el desarrollo tecnológico.

Exceptuando lo dicho sobre la tendencia positiva de la relación entre las capacidades de aprendizaje y producción y el desempeño innovador sectorial, el análisis de la Primera Encuesta muestra que la asociación entre CIT y desempeño innovador no es significativa para todos los sectores. En particular, se identifican tres sectores cuyo desempeño innovador no está asociado con las CIT: 381 (fabricación de productos metálicos), 362 (fabricación de productos de vidrio) y 312 (fabricación de productos alimenticios). Además, es relevante mencionar la poca relación que posee la capacidad de I+D con el desempeño innovador de los diferentes sectores estudiados, lo cual también es consistente con los resultados del análisis de empresas, en los cuales esta capacidad está más débilmente asociada con el desempeño innovador que las capacidades de producción y aprendizaje.

El análisis de la Segunda Encuesta confirma la tendencia de asociación positiva entre las capacidades de producción y aprendizaje y el desempeño innovador sectorial; sin embargo, en este caso, emerge la capacidad de planeación estratégica como una capacidad mejor asociada al desempeño innovador, de forma consistente con los resultados encontrados en el análisis empresarial.

En cuanto al análisis de sectores específicos, la Primera Encuesta revela que el sector 332 (fabricación de muebles y accesorios) es el más rezagado en la acumulación de capacidades de aprendizaje y producción; en cambio, el sector 352 (fabricación de otros productos químicos) se destaca no sólo en la acumulación de capacidades de producción, sino en su desempeño innovador, en lo cual los sectores 321 (fabricación de textiles), 322 (fabricación de prendas de vestir) y 342 (imprentas y editoriales) son los más rezagados.

Por su parte, el análisis sectorial de la Segunda Encuesta destaca claramente el liderazgo del sector 24 (fabricación de sustancias y productos químicos), tanto en la acumulación de CIT como en el desempeño innovador, lo que tiende a ser consistente con los resultados del análisis de la EDT I. Algo semejante ocurre con el sector 15 (elaboración de productos alimenticios), que en ambas encuestas se destaca por su mínimo desempeño innovador. Estos resultados son interesantes, en la medida en que proponen la existencia de un grupo líder de empresas químicas, al igual que un grupo rezagado de empresas alimenticias, cuyos resultados son estadísticamente significantes en el orden nacional, desde la perspectiva de la innovación y en dos momentos diferentes del desarrollo industrial colombiano.

Finalmente, es importante señalar la necesidad de mejorar el diseño, la frecuencia y el acceso a los datos de las encuestas de innovación en Colombia. Cuatro recomendaciones específicas surgen en esta dirección, a la luz de la experiencia adquirida en el presente trabajo: (1) incluir en la encuesta variables que permitan una mejor medición de las CIT, tanto en el sentido de la completitud (que todos los ámbitos de las capacidades estén cubiertos) como de la consistencia teórica (que las variables incluidas

midan dimensiones relevantes de las capacidades). (2) Incluir y posibilitar el acceso a variables de desempeño empresarial relacionadas con las ventas, las utilidades (pérdidas), las exportaciones y el mercado. (3) Definir y caracterizar claramente las innovaciones logradas por la empresa durante la ventana de observación de la Encuesta. (4) Contar con la posibilidad de cruzar los datos de las encuestas de innovación y de otras encuestas nacionales (por ejemplo, la Encuesta Anual Manufacturera) en el ámbito de la empresa.

Como trabajo futuro se pueden identificar al menos cuatro líneas de indagación y profundización por seguir en el corto plazo: (1) consolidar los resultados desde la doble perspectiva analítica y procedimental, sometiendo a una revisión crítica las definiciones de los conceptos relativos a las capacidades y los desempeños, la selección de las variables observables que los componen, las métricas establecidas para estas y las técnicas estadísticas aplicadas. (2) Reproducir los análisis para empresas de una misma división territorial, de un mismo grupo industrial y de un mismo tamaño, y comparar los resultados. (3) Ampliar y profundizar los resultados del análisis sectorial, buscando la forma de integrar los datos de las encuestas de innovación con datos complementarios obtenidos de la Encuesta Anual Manufacturera. Y (4) reelaborar el análisis sectorial, explorando distintas clasificaciones de sectores que mejor reflejen las dinámicas de la innovación, por ejemplo, la taxonomía propuesta por Pavitt (1984) o la extendida de Kristensen (1999).

REFERENCIAS

- Christesen, J. F. (1995). Asset profiles for technological innovation. *Research Policy*, 24 (5), 727-745.
- DiBella, A.; Nevis, E. y Gould, J. (1996). Understanding organizational learning capability. *Journal of Management Studies*, 33, 361-379.
- Goodman, L. A. y Kruskal, W. H. (1954). Measures of association for cross-classification. *Journal of the American Statistical Association*, 49, 732-764.

- Guan, J. y Ma, N. (2003). Innovative capability and export performance of Chinese firms. *Technovation*, 23 (9), 737-747.
- Hitt, M. A.; Hoskisson, R. E. y Kim, H. (1997). International diversification: Effects on innovation and firm performance in product-diversified firms. *Academy of Management Journal*, 40 (4), 767-798.
- Keizer, J. A.; Dijkstra, L. y Halman, J. (2002). Explaining innovative efforts of SMEs: An exploratory survey among SMEs in the mechanical and electrical engineering sector in Netherlands. *Technovation*, 22 (1), 1-13.
- Kristensen, F. S. (1999). *Towards a taxonomy and theory of the interdependence between learning regimes and sectorial patterns of innovation and competition: An empirical analysis of an elaborated Pavitt taxonomy applying Danish data*. Recuperado el 6 de noviembre de 2009, de <http://www.druid.dk/conferences/winter1999/conf-papers/skov.pdf>.
- Liebetrau, A. M. (1983). *Measures of association*. Newbury Park, CA: Sage Publications.
- Organización para la Cooperación y el Desarrollo Económico (OCDE), (2002). *Manual de Frascati: propuesta de norma práctica para encuestas de investigación y desarrollo experimental*. Madrid: FECYT.
- (2005). *Manual de Oslo: guía para la recogida e interpretación de datos sobre innovación* (3a ed.). Paris: autor.
- Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13, 343-373.
- Robledo, J. y Aguilar, J. J. (2007). Naturaleza y características de las capacidades de mejoramiento continuo e innovación en pymes de alimentos. En *Memorias del XII Seminario Latino-Iberoamericano de Gestión Tecnológica, ALTEC 2007*, Buenos Aires, Argentina.
- Sher, P. J. y Yang, P. Y. (2005). The effects of innovative capabilities and R&D clustering on firm performance: The evidence of Taiwan's semiconductor industry. *Technovation*, 25, 33-43.
- Teece, D.; Pisano, G. y Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18 (7), 509-533.
- Vargas, M.; Malaver, F. y Zerda, A. (Eds.), (2003). *La innovación tecnológica en la industria colombiana: un estudio en dos cadenas industriales*. Bogotá: CEJA.
- Wang, C.; Lu, I. y Chen, C. (2008). Evaluating firm technological innovation capability under uncertainty. *Technovation*, 28 (6), 3489-363.
- Yam, R. C. M.; Guan, J. C.; Pun, K. F. y Tang, E. P. Y. (2004). An audit of technological innovation capabilities in Chinese firms: some empirical findings in Beijing, China. *Research Policy*, 33, 1123-1140.

Capítulo 5

Caracterización del relacionamiento empresas-instituciones de educación superior según la Segunda Encuesta Colombiana de Innovación

*Manuel Santiago Arbeláez Cuartas
Fredy Alexander Gómez Jiménez
Lida Isabel Tamayo Giraldo*

La economía global ha cambiado radicalmente en las últimas décadas. Hoy el ciclo de vida de los productos tiende a ser menor, y la competencia, a raíz de la apertura de los mercados, mucho mayor. El cambio tecnológico se ha acelerado considerablemente y se ha convertido en una amenaza a los patrones tradicionales de competencia exitosa, así como en una oportunidad de creación de ventajas competitivas a través de la innovación (Bettis y Hitt, 1995).

La economía del conocimiento avanza como una realidad incontestable y se difunde de una manera insospechada en los diferentes continentes. Las exigencias de responsabilidad social empresarial son cada vez más apremiantes, con sus desafíos de sostenibilidad y compromiso social. Por ello, ante este panorama mundial de cambios acelerados, las empresas necesitan ser mucho más veloces y asertivas en la formulación e implementación de

decisiones estratégicas. Sólo aquellas empresas que se adaptan a las nuevas condiciones del mercado y la tecnología sobreviven y trascienden en el tiempo; las demás, simplemente desaparecen.

Como mecanismo de respuesta a estas circunstancias altamente cambiantes, las empresas se ven hoy en la necesidad de innovar para así poder cautivar un mayor número de clientes con nuevas o mejores propuestas o aumentar la productividad y sostenibilidad de su cadena de producción por medio de la innovación en procesos. En este orden de ideas, las empresas necesitan encontrar la manera más adecuada de hacer innovación, al tiempo que vislumbran en la colaboración una puerta para incrementar sus capacidades de innovación, disminuir los riesgos que esta conlleva y acelerar los procesos de desarrollo de nuevos productos (Shilling, 2005).

La innovación puede lograrse a través de diferentes estrategias e iniciativas empresariales, y una de las más destacables fuentes conocidas es la relación que las empresas establecen con las instituciones de educación superior (IES). El estudio que se presenta en este capítulo explora las características de este relacionamiento y su incidencia en el desempeño innovador de las empresas, a partir de los datos de la Segunda Encuesta Colombiana de Innovación. Para ello, el estudio se enfoca en las *actividades de desarrollo tecnológico* (ADT) y, específicamente, en la *capacitación tecnológica* (CT), como objetos centrales de los diferentes tipos de relacionamiento que pueden establecer las empresas con las IES.

En la Segunda Encuesta Colombiana de Innovación, las ADT se definen como actividades tendientes a poner en práctica conceptos, ideas y métodos necesarios para la adquisición, asimilación e incorporación de nuevos conocimientos. El resultado de estas acciones es un cambio técnico en la empresa, sin que este sea necesariamente una innovación. Dichas actividades se clasifican en cinco grupos: tecnologías incorporadas al capital, tecnologías de gestión, tecnologías transversales, proyectos de investigación y desarrollo (I+D) y capacitación tecnológica (DANE, DNP y Colciencias, 2005).

COLABORACIÓN ENTRE IES Y EMPRESAS

La innovación en la empresa es un proceso que debe ser lo suficientemente ágil, a fin de que se llegue de manera oportuna al mercado. Frente a lo anterior, muchas veces, la empresa necesita recurrir a capacidades que no posee internamente y allí es donde surge el concepto de *innovación abierta*, como mecanismo de aproximación “abierta” a la innovación, que se apalanca en fuentes de ideas tanto internas como externas (Chesbroug, 2003).

Un buen ejemplo de innovación abierta se puede apreciar en la estrategia planteada por la British Telecom, la cual posee un fortalecido frente de colaboración con organizaciones universitarias, como la Universidad de Cambridge, que en sus laboratorios alberga equipos conformados por miembros de ambas instituciones que trabajan en conjunto por un objetivo común (BT, 2006). La innovación abierta se fundamenta en la colaboración, elemento que surge como una fuente llena de oportunidades que potencia, por medio de fuentes externas, las capacidades que posee la empresa en su interior.

A través de la colaboración, las empresas tienen rápidamente acceso a la experiencia y el conocimiento complementarios que requieren para innovar. Este trabajo se enfoca en la colaboración entre las empresas y las IES, concepto este que, además de las universidades, incluye también a los centros de estudio técnico, los centros de estudio tecnológico y el Servicio Nacional de Aprendizaje (Sena).

Al explorar el mundo de la colaboración entre las empresas y las IES, se encuentran tres tipos, como lo son las alianzas estratégicas, el licenciamiento y las organizaciones colectivas de investigación (Shilling, 2005). Otros autores hablan de ayudas para la investigación, investigación cooperativa, transferencia de conocimiento y transferencia de tecnología como las diferentes alternativas que se presentan en la colaboración y que varían en su uso según algunas condiciones de las partes (Santoro y Chakrabarti, 2002).

Según lo planteado por Santoro y Chakrabarti (2002), se encuentra que las ayudas para la investigación son ese tipo de contacto que la empresa ha mantenido de manera tradicional con las IES, por medio del cual la empresa, por ejemplo, hace alguna forma de donación a la IES para fo-

mentar la investigación. Esta clase de relacionamiento se ha caracterizado por presentar poca interacción entre las partes, pero ha ido tomando poco a poco un matiz diferente, ya que las empresas comienzan a pedir con mayor frecuencia algún tipo de contraprestación que puede consistir en la solución de algún problema puntual.

Con relación a la investigación cooperativa, se tiene que esta suele presentar un grado de interacción mucho mayor, que se caracteriza por contemplar contratos de consultoría de manera frecuente y porque de comienzo se plantean problemáticas que requieren una solución específica.

En cuanto a la transferencia de conocimiento, se presenta como un mecanismo de colaboración con mayor interacción entre las partes, la cual puede ser de tipo formal o informal, y puede darse por medio del desarrollo de un currículo, un programa de educación cooperativa, pasantías, contratación de graduados, publicaciones, reuniones y conferencias.

Finalmente, se encuentra la transferencia de tecnología como una forma que posee la empresa para resolver problemas puntuales por medio de un trabajo en conjunto que, por lo general, requiere un alto grado interacción y puede darse a través *joint ventures*, acuerdos de consultoría tecnológica y el uso por parte de la firma de los servicios de extensión de la IES.

La estabilidad y la eficiencia de los cuatro tipos de relacionamiento descritos han sido estudiadas desde un modelo que define cuatro dimensiones que son institucional, organizacional, ejecución y creación de valor (Bruno y Vasconcellos, 2003). Sus autores han aplicado este modelo a situaciones de relacionamiento de empresas brasileñas, y entre febrero de 2005 y marzo de 2006 fue aplicado en Colombia a ocho empresas de la región Caribe, que arrojó diferentes conclusiones en cada una de las dimensiones; no obstante, se resalta principalmente que las alianzas tecnológicas no pueden ser improvisadas, ya que son un proceso evolutivo y deben planearse muy bien, para que así permitan la construcción de relaciones (Abello Llanos, 2007).

Adicionalmente, se han realizado estudios que definen factores de tipo contextual, que se refieren a todas esas condiciones previas que se deben establecer para garantizar el relacionamiento IES-empresa y organiza-

cional, que explora todas esas actitudes que hacen que este se fortalezca mucho más (Mora Valentín, Montoro Sánchez y Guerras Martín, 2004).

De manera complementaria, el éxito en este relacionamiento también se asocia con la capacidad de absorción que posea la empresa, capacidad que radica en esa habilidad que la empresa posee para identificar, asimilar y explotar el conocimiento de su entorno, incluida también la habilidad de la firma para explotar conocimiento externo de un carácter algo más complejo, como lo son los hallazgos de investigación (Cohen y Levinthal, 1989).

En el relacionamiento IES-empresa también se presentan obstáculos que, sobre todo, giran en torno a las diferencias en los objetivos entre las partes, los tiempos que maneja la academia en el plano de la investigación, los resultados que esperan de la investigación ambas partes, las diferencias culturales, las dificultades frente a la apertura que generalmente se espera de la ciencia y la ignorancia que poseen los estudiantes sobre los desarrollos más recientes de la industria (Fontana, Geunab y Matt, 2006).

Durante el relacionamiento se presentan situaciones como: alto grado de incertidumbre, elevada asimetría de información entre las partes, altos costos transaccionales en el intercambio de conocimiento (lo cual requiere la disponibilidad de capacidades de absorción considerables), numerosas *externalidades* a otros actores del mercado y restricciones de financiación en las actividades de producción e intercambio de conocimiento, debido a la adversidad al riesgo y a la tendencia de los mercados a esperar resultados en corto plazo (Veugelers y Cassiman, 2005).

Las empresas, por medio de la colaboración con las IES, buscan desarrollar nuevas ideas o complementar capacidades frente a proyectos en curso, con el fin de innovar tanto en procesos como en productos (Kleivorick *et al.*, 1995). Frecuentemente, las grandes empresas enfocan esta relación en el desarrollo de tecnologías no medulares, mientras las pymes, por lo general, encuentran en este tipo de relacionamiento una forma de garantizar su subsistencia (Santoro y Chakrabarti, 2002).

Por lo tanto, al revisar casos aplicados, las empresas cada vez están más abiertas a involucrar entidades externas en el desarrollo de innovaciones en cada eslabón de toda su cadena de valor. Las universidades y sus grupos

de investigación son una de las posibilidades preferidas por las empresas para establecer este tipo de relaciones, debido a las garantías que brindan en lo que respecta a la protección del conocimiento.

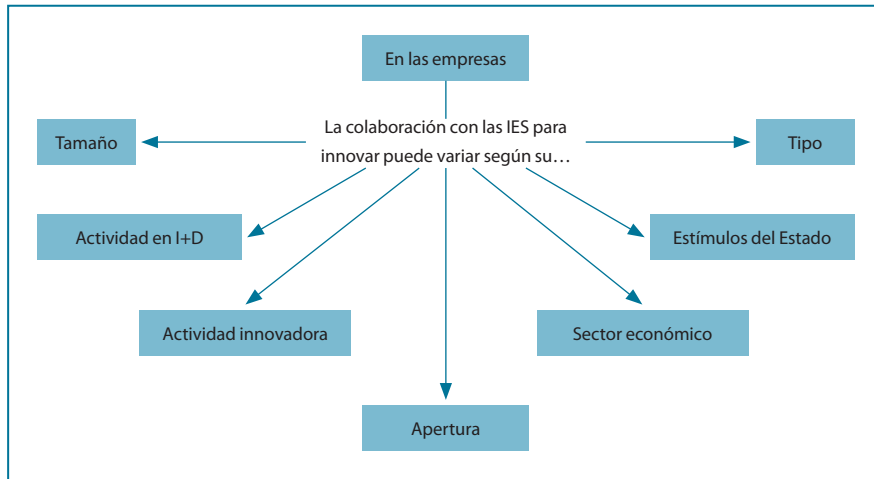
MARCO DE LAS RELACIONES

A partir del análisis de diferentes casos en los que similarmente se han evaluado encuestas de innovación en el mundo, como lo son la Eurostat Community Innovation Survey (CIS), en Europa, y la UK Innovation Survey, en el Reino Unido, se va concluyendo que siempre están presentes unos factores clave en este relacionamiento y que son determinantes para que este mismo se dé.

Desde la óptica de las empresas, se ha observado que para aquellos casos en los que estas se relacionan con la academia, los factores que encauzan el proceso de selección de la institución con la que se van a relacionar son la reputación de la institución, su dominio de las competencias en la temática de interés y el historial de colaboración que se tenga con esta. Por otro lado, se observa que hay unas características comunes en las empresas que colaboran con las IES (Gráfico 1) como son: el tamaño de la firma, que se puede medir por medio del número de empleados que esta posee; la actividad en I+D, que se puede medir como el número de empleados dedicados a I+D en la empresa, de la mano con el número de patentes que solicita la empresa año tras año; la actividad innovadora de la empresa, que se observa en los tipos de innovaciones llevadas a cabo por la empresa, sean de producto, de proceso, organizativa o de mercadotecnia (OCDE, 2005); la apertura de la firma, que se puede medir por medio de su colaboración con agentes externos; el sector económico donde se desempeña la firma, ya que el tamaño del mercado y su velocidad de cambio puede variar entre sectores; los estímulos del Estado, que se puede medir como el acceso que tienen las empresas a fuentes de financiación relacionadas con algún tipo de estímulo estatal, y, finalmente, el tipo de empresa según sea una empresa nacional o multinacional.

Gráfico 1

Variables asociadas al relacionamiento de las empresas con las IES



Fuente: elaboración propia.

Las siete variables anteriores han sido exploradas por diferentes autores alrededor del mundo. A continuación se detallan los resultados más relevantes que se han encontrado al respecto.

En cuanto a la variable tamaño, se encuentra que diferentes trabajos han revelado que las firmas grandes suelen tener un vínculo de colaboración más estrecho con la academia (Cohen, Nelson y Wals, 2002; Mohnen y Hoareau, 2002; Laursen y Salter, 2004; López, 2008; Segarra y Arauzo, 2008). Esta relación se debe a que en la medida en que una empresa es más grande, esta posee por lo general una mayor capacidad de absorción (López, 2008). También se ha encontrado que las empresas pequeñas, particularmente en el caso de los *start-ups*, tienen un estrecho vínculo de colaboración con las IES (Cohen, Nelson y Wals, 2002); este vínculo tan estrecho se debe, sobre todo, a que muchos de los *start-ups* son ideas que han emergido de alguna IES y están en proceso de consolidación en el mercado.

Por su actividad en I+D, Fontana, Geunab y Matt concluyeron en el 2006, con el análisis de la encuesta KNOW, realizada en siete países eu-

ropeos, que las empresas que poseen actividad en I+D en su interior son mucho más propensas a colaborar con la academia que aquellas que no.

En relación con la actividad innovadora, en varios escenarios ha surgido la inquietud de si hay una mayor tendencia a que esta colaboración se dé cuando se trata de innovaciones en productos o en procesos. En el caso del análisis de la encuesta KNOW, del 2000, a pesar de que se tenían algunas expectativas sobre esta variable, se encontró que no había una relación estrecha entre el tipo de innovación y el grado de relacionamiento (Fontana, Geunab y Matt, 2006).

Respecto a la apertura de la firma, se ha encontrado que en la medida en que una firma se dispone a colaborar o explorar la innovación a través de fuentes externas, esta es más propensa a colaborar con la academia (Laursen y Salter, 2004; Fontana, Geunab y Matt, 2006).

Por otra parte, el sector económico es determinante para medir la magnitud de este relacionamiento. En la UK Innovation Survey se encontró, por ejemplo, que sólo unos pocos sectores industriales muestran un marcado aporte de la academia a la innovación empresarial (Laursen y Salter, 2004). Quizás esto se explica porque algunos sectores industriales de gran avance tecnológico utilizan la ciencia que se produce en la academia para innovar (Klevorick *et al.*, 1995).

En relación con el impacto que tienen los estímulos del Estado en el fomento de este relacionamiento, se tiene que debido a la diferencia cultural entre las empresas y la academia, ambas partes están muy alejadas para llegar a la cooperación de una manera natural, por lo que se requiere que el gobierno fomente o encauce los vínculos (Mohnen y Hoareau, 2002).

Finalmente, respecto al tipo de empresa, según sea multinacional o nacional, se presenta el caso del análisis desde la perspectiva española de la CIS III en Europa, en la cual se encontró que las firmas españolas se inclinan a tener mayor colaboración con la academia española que aquellas multinacionales radicadas en España, donde las decisiones de colaboración se toman desde la casa matriz (Segarra y Arauzo, 2008).

METODOLOGÍA

RECOLECCIÓN DE DATOS

El principal insumo para este trabajo está conformado por los datos recolectados a través de la Segunda Encuesta Colombiana de Innovación —o Segunda Encuesta de Desarrollo e Innovación Tecnológica en el Establecimiento Industrial Colombiano (EDIT II), aplicada en 2005 para los años 2003 y 2004—. Los datos se obtuvieron por entrevista directa a 6.172 establecimientos industriales establecidos en Colombia, y tuvo cobertura nacional representativa de agrupaciones industriales según el Código de Clasificación Internacional Industrial Uniforme (CIIU).

La unidad de selección y observación fue el establecimiento industrial, definido como la unidad económica que, bajo una forma jurídica única o un solo propietario y en una sola ubicación física, se dedica a la producción del grupo más homogéneo posible de bienes manufacturados. Respecto a la muestra de la EDIT II, debe observarse que de los 6.172 establecimientos encuestados, un número significativo no respondió la totalidad de las preguntas que se seleccionaron para el estudio, por lo que el filtrado de la encuesta resultó en un conjunto final de 4.003 establecimientos, sobre cuyos datos se aplicaron las técnicas de análisis.

HERRAMIENTAS PARA EL ANÁLISIS DE DATOS

Dado que el estudio tiene un enfoque no supervisado, las técnicas estadísticas exploratorias de búsqueda de información fueron las que más se acomodaron al trabajo. Principalmente se utilizaron herramientas estadísticas tanto para la exploración y depuración de variables como para el posterior análisis de relacionamiento entre las empresas y las IES.

Adicionalmente, se aplicaron técnicas de análisis multivariante (TAM), sobre todo para la reducción y clasificación de variables (análisis de factor y análisis de *cluster*). Este procedimiento fue útil en el hallazgo de perfiles

y características similares que disminuyeran la complejidad de análisis del gran número de variables.

Finalmente, se hicieron comparaciones de medias entre los diferentes perfiles encontrados. Como la normalidad de los datos no se garantiza, se aplicó una prueba no paramétrica para la comparación de medias (Kruskal-Wallis), y la prueba de comparación *post hoc* dos a dos de Dunnett, la cual tampoco exige la normalidad de los datos ni la homogeneidad de las varianzas.

RESULTADOS

El Cuadro 1 muestra tres grupos de variables seleccionadas de la EDIT II: variables de relacionamiento, variable de desempeño y variables caracterizadoras. Las primeras poseen una característica especial: miden directamente la relación empresa-IES, por lo que si alguna empresa posee valores apreciables en una o varias variables de estas, al mismo tiempo dicha empresa poseerá un relacionamiento fuerte con la IES. Entre tanto, la variable de desempeño se encuentra representada por el número de innovaciones que la empresa logró obtener en los años 2003 y 2004, tanto en productos como en procesos.

Cuadro 1

Variables de la EDIT II seleccionadas para el estudio

Variable	Tipo de variable
Valor financiado de actividades de desarrollo tecnológico para el período 2003-2004 (Colciencias línea Universidad Empresa)	Cuantitativa
Valor financiado de actividades de desarrollo tecnológico por la universidad	Cuantitativa
Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (Sena)	Cuantitativa
Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (Universidad)	Cuantitativa

Continúa

Variable	Tipo de variable
Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (Instituto Tecnológico)	Cuantitativa
Variable de desempeño	
Número de innovaciones realizadas por el establecimiento en el período 2003-2004	Cuantitativa discreta
Variabes caracterizadoras	
Tamaño de la empresa	Nominal: micro, pequeña, mediana y grande
Número de empleados en actividades de I+D	Cuantitativa discreta
Número de patentes solicitadas	Cuantitativa discreta
Número de relaciones ejecutadas con entidades externas para el desarrollo de actividades de desarrollo tecnológico	Cuantitativa discreta
Sector industrial según el código CIU	Nominal código a 3 dígitos
Financiamiento de actividades de desarrollo tecnológico por parte de entidades públicas	Cuantitativa continua
Empresa nacional o extranjera	Nominal: nacional o extranjera
Tipo de innovación	Nominal: producto, procesos, organizativa y de mercadotecnia

Fuente: elaboración propia.

Finalmente, las variables caracterizadoras presentan las diferentes “características” que describen el establecimiento en cuanto a su tamaño, su actividad en I+D, su actividad innovadora, su grado de apertura, el sector económico al que pertenece, la financiación que recibe del Estado y el tipo de empresa, según sea multinacional o nacional, las cuales fueron identificadas como variables que se utilizan en la literatura para caracterizar la relación IES-empresa.

El tamaño de la empresa se calcula a partir del número de empleados del establecimiento de la siguiente forma: si posee menos de 10 empleados es micro; entre 10 y 50, pequeña; entre 51 y 200, mediana, y más de 200, grande. Respecto a si la empresa se considera nacional o extranjera, se adoptó el umbral del 50%; es decir, si una compañía extranjera tiene

más del 50% del capital social de una compañía establecida en Colombia, entonces se considera extranjera (véase Cuadro 1).

Como se presentó en la metodología del trabajo, se ha optado por realizar un análisis pasivo o no supervisado en el que se realizan diferentes pruebas estadísticas, específicamente sobre las variables de relacionamiento presentadas en el Cuadro 1, para después observar qué información y qué relacionamiento de interés emergen. Se han seleccionado las variables de relacionamiento (cinco en total) para aplicar técnicas de estadística multivariante, con el fin de crear perfiles sobre los establecimientos estudiados, esperando obtener *clusters* con características particulares y diferenciables que permitan una comparación desde el punto de vista innovador.

Al realizar una prueba de KMO (medida de adecuación de Kaiser-Meyer-Olkin) y un test de esfericidad de Barlett, con su significancia sobre las cinco variables, se obtienen los valores de 0,667 y 0,000 para cada prueba. Ello permite concluir que los datos son adecuados para realizar un análisis de factor (valores recomendados: $KMO > 0,5$ y valor de $p < 0,05$). Con la extracción de los factores por medio del *software* SPSS se obtuvieron dos factores que explican el 59,3% de la varianza, por medio de la metodología de los componentes principales. Como lo muestra el Cuadro 2, el factor 1 tiene afinidad con las variables relacionadas con los esfuerzos de capacitación que realizan las empresas con las IES, mientras el factor 2 tiende a representar las variables asociadas con la realización de ADT entre las empresas y las IES. El factor 1 se etiqueta como “Capacitación”, y el factor 2, como “ADT”.

Cuadro 2
Análisis de componentes principales

Código	Variable	Factor 1	Factor 2
CAPUNI	Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (universidad)	0,814	0,099
CAPTECN	Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (instituto tecnológico)	0,801	0,099
CAPSENA	Valor invertido en capacitación tecnológica y gestión en el período 2003-2004 (Sena)	0,775	-0,099

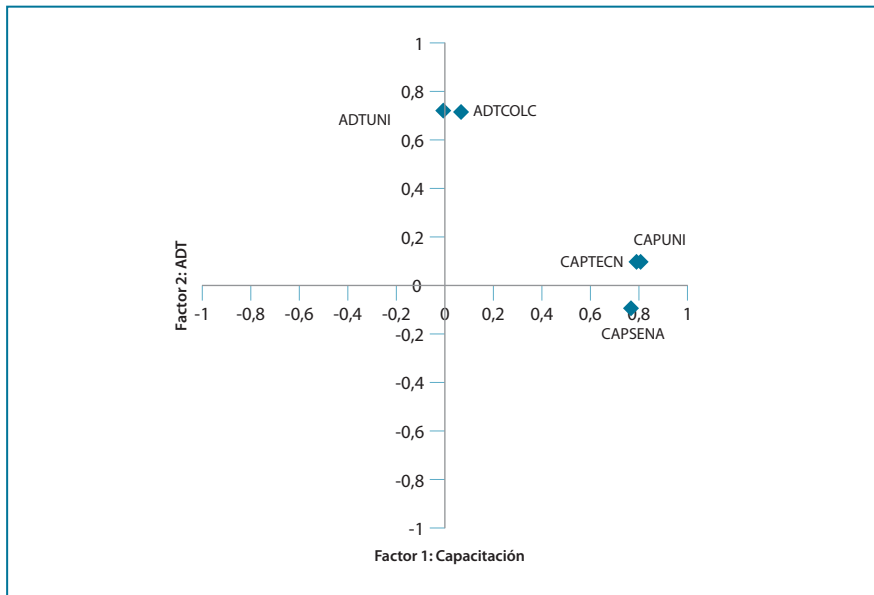
Continúa

Código	Variable	Factor 1	Factor 2
ADTCOLC	Valor financiado de actividades de desarrollo tecnológico para el período 2003-2004 (Colciencias línea Universidad Empresa)	-0,003	0,719
ADTUNI	Valor financiado de actividades de desarrollo tecnológico para el período 2003-2004 (universidad)	0,06	0,715

Método de rotación: Varimax con normalización de Kaiser (converge con tres iteraciones).
 Fuente: elaboración propia.

El Gráfico 2 corresponde a los valores obtenidos en el Cuadro 2, que muestra de forma más clara el perfil de la agrupación de las variables. Existe de manera significativa una correlación entre las variables que suelen dirigirse a la realización de ADT (eje vertical) y entre las variables tendientes a describir el grado capacitación realizada por el establecimiento (eje horizontal).

Gráfico 2
Construcción de los clusters: gráfico de factores



Fuente: elaboración propia.

Al realizar un análisis de *cluster* sobre las puntuaciones obtenidas de los dos factores encontrados, se obtiene el agrupamiento descrito en el Cuadro 3, que revela que el procedimiento ha logrado determinar tres agrupamientos diferentes, los cuales, de acuerdo con sus características particulares, se llaman respectivamente establecimientos aislados, establecimientos capacitados y establecimientos relacionados. Como era de esperarse por la experiencia empírica, en el primer *cluster*, que se denomina aislados, nombre dado por su poco relacionamiento con las IES, se agrupa la gran mayoría de los establecimientos bajo estudio.

Cuadro 3

Clusters y número de establecimientos por cluster

Nombre del cluster	Número de establecimientos
Aislados	3.933
Capacitados	58
Relacionados	12

Fuente: elaboración propia.

Los resultados de comparación de medias de las variables caracterizadoras cuantitativas continuas entre los tres *clusters* encontrados se muestran en el Cuadro 4 y para las variables caracterizadoras nominales se muestra en el Cuadro 5. Para el caso de las variables cuantitativas continuas, se encuentran los promedios para cada cluster y se presenta el resultado de la prueba no paramétrica de Kruskal-Wallis y de la prueba Dunnett de comparación dos a dos. Para el caso de las variables nominales, el Cuadro 5 muestra los resultados agrupados a través de una tabla de contingencia.

Cuadro 4

Resultados de comparación de medias de las variables caracterizadoras cuantitativas continuas

Variables	Prueba de igualdad de medias	Aislados (1)	Capacitados (2)	Relacionados (3)	Prueba dos a dos de Dunnett
Número de empleados en actividades de I+D	0,000	1,1	4,7	2,4	2>1*
Número de patentes solicitadas	0,000	0	0,2	0,3	1=2=3*
Número de relaciones con entidades externas para ADT	0,000	0,4	1,2	2	2>1**
Financiamiento de ADT sector público (miles de pesos)	0,000	60.262	343.507	60.049	1=2=3**

* Resultados obtenidos con un 5% de significancia; ** resultados obtenidos con un 10% de significancia.
Fuente: elaboración propia.

Cuadro 5

Variables caracterizadoras: contingencia para las variables tamaño y tipo vs. cluster

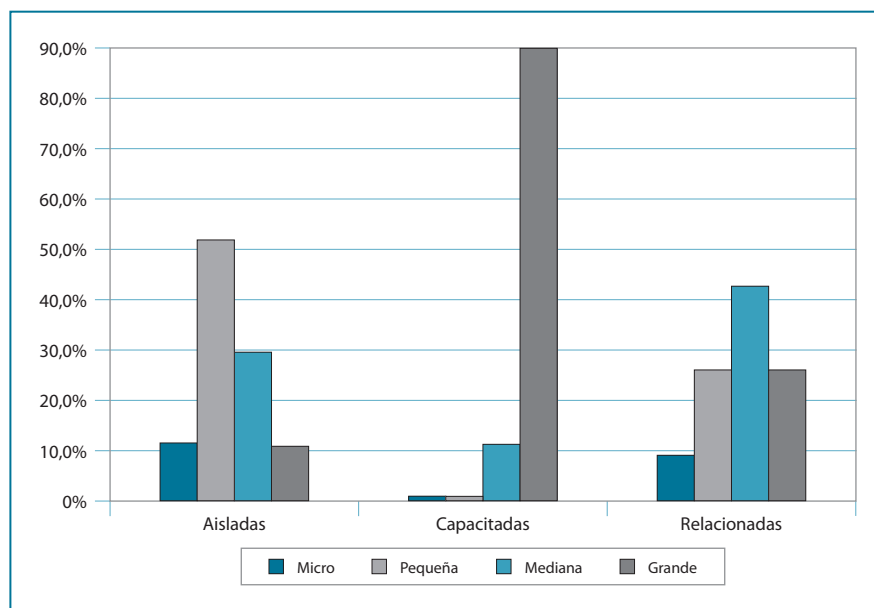
Tamaño de la empresa			
Variables	Aislados	Capacitados	Relacionados
Micro	419	0	1
Pequeña	1.996	0	3
Mediana	1.128	6	5
Grande	390	52	3
Tipo de empresa			
Variables	Aislados	Capacitados	Relacionados
Nacional	3.732	36	11
Extranjera	201	22	1

Fuente: elaboración propia.

El Cuadro 5 se puede observar de manera gráfica y en porcentajes a partir de los gráficos 3 y 4, para el tamaño de la empresa y tipo de empresa, respectivamente.

Gráfico 3

Variables caracterizadoras: distribución de los *clusters* por tamaño

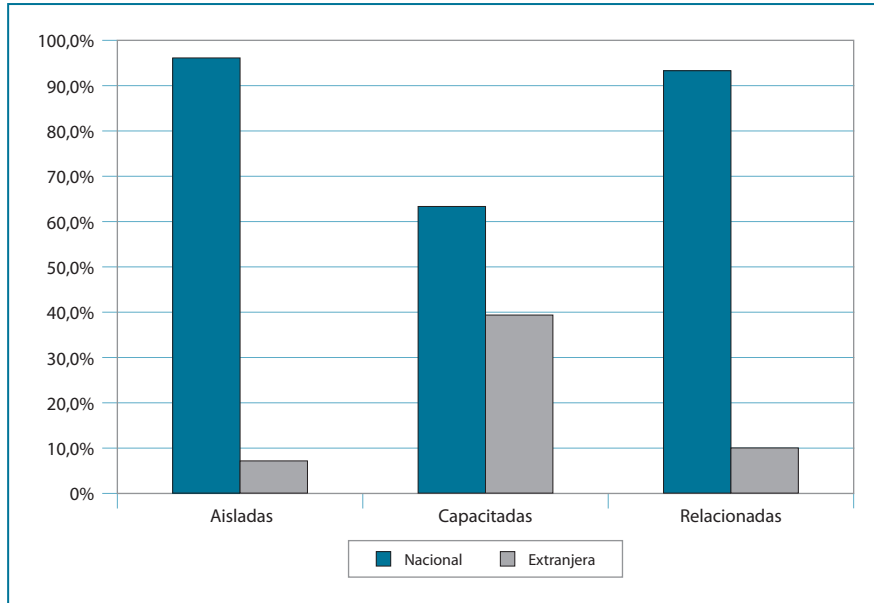


Fuente: elaboración propia.

CLUSTER 1: LOS ESTABLECIMIENTOS AISLADOS

El primer *cluster*, llamado los aislados, agrupa a la gran mayoría de las compañías tomadas en el estudio (3.933 de 4.003), lo que representa el caso típico de la empresa colombiana. Se llaman aislados porque su relacionamiento con las IES se limita únicamente, en caso de que exista, a realizar pocas actividades de capacitación; es decir, no se encuentra en este *cluster* un relacionamiento fuerte, bien sea en capacitación o en ejecución de ADT.

Gráfico 4

Variables caracterizadoras: distribución de los *clusters* por tipo

Fuente: elaboración propia.

El Cuadro 4 permite concluir que los establecimientos que fueron considerados aislados son aquellos que poseen el menor número de empleados en actividades de I+D, por lo que sus esfuerzos en esta rama suelen ser muy limitados. Al mismo tiempo, su relacionamiento externo para realizar ADT fue menor que el del *cluster* llamado capacitados. También es particular el hecho de que los aislados están conformados principalmente por pymes y su capital es, en su gran mayoría, nacional (gráficos 3 y 4). Finalmente, este *cluster* tiene una alta representatividad de las empresas que pertenecen al sector industrial con código CIIU 181 (fabricación de prendas de vestir, excepto prendas de piel), con una representación de 451 de los 3.933 establecimientos del *cluster*.

CLUSTER 2: LOS ESTABLECIMIENTOS CAPACITADOS

Los establecimientos capacitados poseen unos resultados muy destacables en las variables relacionadas con la capacitación, bien sea a través de una institución universitaria, institución tecnológica, institución técnica o el Sena. Predominan las relativamente altas inversiones realizadas en actividades de capacitación con las IES y de allí viene su nombre. Son, sin ninguna discusión, el *cluster* más dinámico de los tres encontrados, dado que de acuerdo con los resultados presentados en el Cuadro 4, poseen los mayores esfuerzos en actividades de I+D, a través de la contratación de empleados para dichas actividades, y al mismo tiempo poseen el mayor número de relacionamientos con entidades externas.

Respecto a las variables caracterizadoras nominales, el Gráfico 4 muestra que aunque un buen porcentaje de empresas son nacionales, aproximadamente un 38% lo conforman empresas extranjeras. En cuanto al tamaño, el Gráfico 3 muestra que pertenecen principalmente a las grandes empresas, con un porcentaje cercano al 90%, es decir, a empresas con un número mayor a 200 empleados. No existe una fuerte representatividad de algún sector industrial en especial, pero vale la pena mencionar la figuración de los sectores 242 (fabricación de otros productos químicos) y 269 (fabricación de productos minerales no metálicos) con siete y cinco establecimientos, respectivamente.

CLUSTER 3: LOS ESTABLECIMIENTOS RELACIONADOS

Lo más relevante de las empresas agrupadas en este *cluster* es su interacción con las IES a través de ADT. Por lo tanto, a diferencia de los establecimientos aislados y los capacitados, este exhibe fuertes relaciones, principalmente en ADT; por ello es el *cluster* más pequeño, pues agrupa sólo a 12 establecimientos.

Un resultado particular es que este *cluster* no muestra un predominio de alguno de los diferentes tamaños de empresa, es decir, sus empresas se distribuyen en casi iguales porcentajes entre los distintos tamaños definidos, lo que lleva a concluir que esta variable no contribuye a caracterizar las

relaciones colaborativas para llevar a cabo ADT. Lo que sí es predominante es el origen del capital de la empresa, pues 11 de las 12 empresas estudiadas son de capital nacional (Gráfico 4). Además, no existe un sector relevante de la economía que pertenezca a este *cluster*; sin embargo, el sector 242 (fabricación de otros productos químicos) posee tres establecimientos de los 12 incluidos.

DESEMPEÑO INNOVADOR DE LOS *CLUSTERS*: RESULTADOS DE INNOVACIÓN

Además de estudiar de manera exploratoria el relacionamiento entre empresas e IES a partir de los datos de la EDIT II, el presente estudio buscó identificar evidencias de correlación entre dicho relacionamiento y el desempeño innovador empresarial, entendido este como los resultados obtenidos en innovación. El Cuadro 6 muestra los resultados de comparar las medias de número de innovaciones obtenidas por los establecimientos para cada *cluster* encontrado. Los resultados se inclinan a que el *cluster* más dinámico de los tres encontrados, el de los establecimientos capacitados, es el que posee al mismo tiempo los mejores resultados en número de innovaciones obtenidas.

Cuadro 6

Comparación de medias: número de innovaciones para cada *cluster*

Variables	Prueba de igualdad de medias	Aislados (1)	Capacitados (2)	Relacionados (3)	Prueba dos a dos de Dunnett
Número promedio de innovaciones en productos y procesos	0,000	0,38	1,34	0,5	2>1

Fuente: elaboración propia.

Particularmente se observa que aquellas compañías agrupadas en el *cluster* de establecimientos relacionados no presentan un comportamiento igual en el número de innovaciones realizadas, a pesar del fuerte enfoque en actividades de ADT. Cabe resaltar que en el *cluster* de los establecimientos

aislados, 3.339 de los 3.933 establecimientos (85%) no tienen ningún tipo de innovación entre 2003 y 2004, lo que demuestra el poco dinamismo existente en este *cluster* para la innovación. En los demás (15%) prima la innovación en productos y una combinación de innovación en productos y procesos.

Para el *cluster* de los establecimientos relacionados, su innovación es bastante pobre (9 de 12 establecimientos no realizaron ningún tipo de innovación), y no hay tendencia hacia algún tipo específico de innovación, lo que abre interrogantes en dos direcciones: primero, el pequeño número de establecimientos encontrados puede ser insuficiente para llegar a conclusiones confiables (12 de 4.003); segundo, los resultados pueden sugerir que las empresas requieren plazos relativamente amplios para desarrollar el potencial innovador de sus relaciones de colaboración en ADT.

Finalmente, el *cluster* de los establecimientos capacitados posee una fuerte participación de las innovaciones de producto y de proceso. El Cuadro 6 muestra una diferencia significativa entre los capacitados y los aislados en lo que respecta al número promedio de innovaciones producidas en los años estudiados.

CONCLUSIONES

El trabajo identifica tres *clusters* que contribuyen a caracterizar el relacionamiento entre empresas e IES a partir de la EDIT II. El primero de ellos (establecimientos aislados) agrupa a la gran mayoría de empresas y representa el caso típico de la empresa colombiana. Su relacionamiento con las IES se limita, en caso de existir, a realizar algunas actividades de capacitación. Estas empresas muestran el menor número de empleados en actividades de I+D, tienden a ser pymes, su capital es fundamentalmente de origen nacional y en su gran mayoría (85%) son empresas no innovadoras. Las empresas del sector CIIU 181 (fabricación de prendas de vestir, excepto prendas de piel) constituyen un buen ejemplo de este *cluster*.

El segundo es el *cluster* de los establecimientos capacitados. Estas empresas exhiben inversiones relativamente significativas en actividades de capacitación, muestran los mayores esfuerzos en actividades de I+D (a

través de la contratación de empleados para dichas actividades) y poseen el mayor número de relacionamientos con entidades externas. Un porcentaje relativamente alto son de capital extranjero y tienden a ser grandes empresas. Además, es el *cluster* que exhibe el mejor desempeño innovador en términos de número de innovaciones obtenidas. Los sectores CIIU 242 (fabricación de otros productos químicos) y 269 (fabricación de productos minerales no metálicos) proporcionan un buen ejemplo del tipo de actividad productiva de estas empresas.

La característica más relevante del tercer *cluster*, el de los establecimientos relacionados, es la de presentar la mayor intensidad en el relacionamiento con IES a través de ADT, donde las actividades de capacitación también son relevantes. No obstante, es el *cluster* más pequeño, pues agrupa solamente a 12 establecimientos. Es claro cómo el relacionamiento de las empresas con las IES para llevar a cabo ADT es algo excepcional en Colombia. Curiosamente, este *cluster* no muestra ningún predominio de determinado tamaño de empresas, pero sí una tendencia a ser empresas de capital nacional. Su desempeño innovador no es significativo, lo cual llama a examinar con mayor detalle las dinámicas de la innovación empresarial, bajo la existencia de vínculos colaborativos entre empresas e IES en ADT.

REFERENCIAS

- Abello Llanos, R. (2007). Factores claves en las alianzas universidad-industria como soporte de la productividad en la industria local: hacia un modelo de desarrollo económico y social sostenible. *Investigación y Desarrollo*, 15, 208-225.
- Banco de la República (2003). *Código Industrial Internacional Uniforme: revisión 3*. Recuperado el 28 de junio de 2009, de <http://quimbaya.banrep.gov.co/servicios/saf2/BRCodigosCIIU.html>.
- Bettis, R. y Hitt, M. (1995). The new competitive landscape. *Strategic Management Journal*, 16, 7-19.
- Bruno, M. A. y Vasconcellos, E. (2003). Applying a management framework to the three high-sharing technological alliances. *Tecnología, Produção e Logística*, 28-47.
- BT (2006). *Embracing open innovation: A new approach to creating sustainable value*. Recuperado el 15 de octubre de 2008, de http://globalservices.bt.com/LeafAction.do?Record=Embracing_open_innovation_whitepaper_all_en-gb

- Cassiman, B. y Veugelers, R. (2003). *R&D cooperation between firms and universities: Some empirical evidence from belgian manufacturing*. Recuperado el 15 de mayo de 2009, de <http://www.cepr.org/pubs/new-dps/dplist.asp?dpno=3951.asp>.
- Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School.
- Cohen, W. M. y Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *The Economic Journal*, 99, 569-596.
- Cohen, W. M., Nelson, R. R., & Wals, J. P. (2002). Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science* 48, 1-23.
- DANE, DNP y Colciencias (2005). *Manual de diligenciamiento de la Segunda Encuesta de Desarrollo e Innovación Tecnológica 2005*. Bogotá.
- Fontana, R.; Geunab, A. y Matt, M. (2006). Factors affecting university-industry R&D projects: The importance of searching, screening and signalling. *Research Policy*, 35, 309-323.
- Klevorick, A. K.; Levin, R. C.; Nelson, R. R. y Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research Policy*, 24, 185-205.
- Laursen, K. y Salter, A. (2004). Searching high and low: What types of firms use universities as a source of innovation? *Research Policy*, 33, 1201-1215.
- López, A. (2008). Determinants of R&D cooperation: Evidence from Spanish manufacturing firms. *International Journal of Industrial Organization*, 26 (1), 113-136.
- Miotti, L. y Sachwald, F. (2003). Co-operative R&D: Why and with whom? An integrated framework of analysis. *Research Policy*, 32, 1481-1499.
- Mohnen, P. y Hoareau, C. (2002). *What type of enterprise forges close links with universities and government labs?: Evidence from CIS 2*. Montreal: CIRANO. Recuperado el 19 de octubre de 2009, de <http://www.cirano.qc.ca/pdf/publication/2002s-25.pdf>.
- Mora Valentin, E. M.; Montoro Sánchez, A. y Guerras Martin, L. A. (2004). Determining factors in the success of R&D cooperative agreements between firms and research organizations. *Research Policy*, 33, 17-40.
- Organización para la Cooperación y el Desarrollo Económico (OECD), (2005). *Manual de Oslo* (3a ed.). Madrid: Grupo Tragsa.
- Santoro, M. D. y Chakrabarti, A. K. (2002). Firm size and technology centrality in industry-university interactions. *Research Policy*, 31, 1163-1180.
- Segarra Blasco, A. y Arauzo Carod, J. M. (2008). Sources of innovation and industry-university interaction: Evidence from Spanish firms. *Research Policy*, 37, 1283-1295.
- Shilling, M. (2005). *Strategic management of technological innovation*. New York: McGraw-Hill.
- Veugelers, R. y Cassiman, B. (2005). R&D cooperation between firms and universities: Some empirical evidence from Belgian manufacturing. *International Journal of Industrial Organization*, 23, 355-379.

Capítulo 6

Las capacidades de absorción tecnológica: una mirada centrada en la adquisición de la tecnología¹

Florentino Malaver Rodríguez
Marisela Vargas Pérez
Felipe Ardila Camelo

En este capítulo se acude a técnicas de descubrimiento de conocimiento para indagar de manera exploratoria por las capacidades de absorción desplegadas por las empresas industriales de Bogotá y Cundinamarca (ByC) cuando adquieren tecnologías exógenas. Tales capacidades son analizadas a través de los recursos humanos y las prácticas utilizadas para dicha incorporación. Para ello se estructura un marco analítico y metodológico, que articula el marco conceptual y algunas técnicas de visualización de asociación entre variables y de búsqueda de patrones ocultos en la información.

En el estudio se utiliza la información arrojada por la Encuesta de Desarrollo Tecnológico e Innovación de Bogotá y Cundinamarca (EIByC),

¹ Los autores agradecen la colaboración y los valiosos aportes de Nelson Obregón (PhD), director del Doctorado de Ingeniería de la Pontificia Universidad Javeriana.

realizada en el 2005, en la cual se indaga de manera explícita por esta problemática. El trabajo consta de cuatro secciones. En la primera se presenta el enfoque analítico que orienta el ejercicio; en la segunda, el marco metodológico; en la tercera, el análisis de los resultados obtenidos, y, por último, las conclusiones obtenidas, leídas desde la perspectiva de la política.

ENFOQUE ANALÍTICO

Para algunos autores, las brechas comerciales se explican por las brechas tecnológicas (Hurtienne y Messner, 1999). Para otros (Pérez, 2001), el éxito de cualquier estrategia de desarrollo de los países menos avanzados dependerá de su capacidad para aprovechar las oportunidades que abren las revoluciones tecnológicas generadas en los países desarrollados, esto es, dependerá de sus capacidades para adquirir, adaptar y, ojalá, desarrollar las nuevas tecnologías.

Experiencias recientes de industrialización tardía muestran que el desarrollo de las capacidades tecnológicas es el resultado de un proceso de aprendizaje que corre paralelo a los de industrialización y desarrollo productivo (Kim y Nelson, 2000; Lee, 2000). Ese proceso transita por varias etapas: primero, por una fase duplicativa, en la que predomina la copia de productos y procesos desarrollados en los países avanzados; luego, por una fase de imitación creativa, en la que se adaptan y mejoran esas tecnologías, procesos y productos; por último, por una fase generativa, en la cual se crean nuevos procesos y productos que están situados en la frontera tecnológica y competitiva.²

En países en vía de desarrollo que, como Colombia, todavía no han alcanzado la tercera etapa, las capacidades para incorporar y usar de manera creativa las tecnologías desarrolladas en los países más avanzados resultan fundamentales para su desarrollo tecnológico y su competitividad. Esto se debe a que el grueso del cambio técnico que en ellos se registra es exó-

² Esta taxonomía corresponde a una adaptación hecha por Malaver y Vargas (2006), y que resulta de combinar las desarrolladas por Kim y Nelson (2000) y por Lee (2000).

geno. El desarrollo de esas capacidades de absorción tecnológica depende en buena medida de los aprendizajes generados por arduos procesos de transferencia tecnológica, que trascienden las compras pasivas —producto de la confianza en que el cambio técnico es transparente y viene incorporado en las maquinarias y equipos adquiridos—.

Pese a la importancia estratégica de las capacidades de absorción para acelerar el tránsito hacia estadios más avanzados de desarrollo tecnológico, este tema no es abordado de manera explícita en las encuestas de innovación y desarrollo efectuadas en América Latina. Un intento de aproximación al tema se efectuó, aunque de manera parcial, en la EIByC, realizada en el 2005 (Malaver y Vargas, 2006). Por ello, este trabajo se centra en los resultados arrojados por dicha encuesta.

La EIByC permite indagar por las capacidades de absorción en las empresas de ByC desde dos de las perspectivas analíticas más importantes: la neoclásica y la evolutiva. En el primer caso, tal como lo señala Arundel (2007), la mirada se centra en los *inputs* y los *outputs* de los procesos de innovación. Para nuestro caso, los primeros aluden a los recursos tecnológicos, esto es, los recursos humanos, financieros, infraestructuras físicas y tecnológicas (departamentos de I+D, etc.); los segundos, a los resultados en materia de incorporación de maquinaria y equipo, como una variable *proxy* de la absorción tecnológica.³

En el segundo caso, la mirada se centra en los procesos de aprendizaje generados por la adquisición, asimilación y explotación creativa de las tecnologías incorporadas a las empresas. Esos aprendizajes se acumulan y se constituyen en la base para los nuevos desarrollos de las capacidades de absorción de las tecnologías adquiridas, que son *path dependent* (Dahlman y Nelson, 1993).

Esos aprendizajes son, además, indispensables para el éxito de la absorción pues, como lo señalan Nelson y Sampat (2001), una implantación exitosa de nuevas tecnologías físicas requiere tecnologías sociales congruentes con ellas, y esto demanda la mayoría de las veces cambios en

³ Asunto sobre el cual se ahondará más adelante.

las rutinas organizacionales, es decir, en las prácticas que se despliegan cuando se hacen las cosas. Esos aprendizajes a su vez, se acumulan, y de esta forma se van desarrollando las capacidades tecnológicas, expresadas en nuevas o mejores prácticas para absorber el cambio técnico.

MARCO METODOLÓGICO

En concordancia con el enfoque analítico, se establece un marco conceptual centrado en las capacidades de absorción tecnológica de la firma, a partir del cual, en un segundo momento, se definen las variables y relaciones entre ellas; luego se describen las técnicas de descubrimiento que se utilizan en el trabajo, y, finalmente, este marco conceptual y las técnicas se articulan en un protocolo, que define las actividades seguidas para el logro de los objetivos buscados.

MARCO CONCEPTUAL

Por capacidades se entiende lo que una empresa es capaz de hacer (Grant, 2005). Estas se expresan en la forma como se hacen las cosas, en este caso, en las prácticas para absorber las tecnologías. Las capacidades de absorción se definieron originalmente con Cohen y Levinthal (1990) como la habilidad de una firma para reconocer el valor de la nueva información externa, para asimilarla y aplicarla con fines comerciales.

Múltiples estudios se han nutrido de este trabajo seminal. Entre ellos se destaca el realizado por Zahra y George, quienes ponen la noción de capacidad de absorción en una perspectiva dinámica y la definen como el “conjunto de rutinas organizacionales y de procesos estratégicos por los cuales las empresas adquieren, asimilan, transforman y explotan conocimiento con el propósito de crear valor” (2002: 198).

Zahra y George (2002) también plantean un “modelo extendido de las capacidades de absorción” (Vega *et al.*, 2008: 4), conformado por cuatro dimensiones: la *adquisición*, referida a las capacidades de una firma para

identificar y adquirir conocimiento externo que es crítico para su operación; la *asimilación*, relacionada con los procesos que le permiten a la firma analizar, procesar, interpretar y entender esos conocimientos (tecnologías); la *transformación*, referente a la capacidad, para combinar el conocimiento existente en la empresa con el nuevo conocimiento adquirido y asimilado, y la *explotación*, concerniente a las rutinas, que le permiten extender, refinar y apalancar las capacidades existentes o crear unas nuevas por la incorporación en las operaciones de la empresa del conocimiento adquirido y transformado. A las dos primeras las consideran capacidades potenciales, y las dos últimas, realizadas.

Dadas las limitaciones de la fuente de información empleada —y de las encuestas en general— para captar fenómenos organizacionales tan complejos y dinámicos como este, aquí se trabaja con la primera dimensión, esto es, la de adquisición (o incorporación) de los conocimientos (tecnologías) externos; pero se profundiza en la indagación por los procesos de adquisición, descomponiéndolos en las actividades de identificación, selección y negociación de las tecnologías que se van a incorporar a la empresa. Allí se despliegan prácticas (rutinas) que reflejan distintos niveles de capacidades acumuladas. Las más desarrolladas, se considera, al igual que en Zahra y George (2002), tendrán un potencial de contribución mayor a la asimilación, transformación y explotación, es decir, a la generación de valor agregado y competitividad de la empresa.

DEFINICIÓN DE VARIABLES

De acuerdo con el enfoque analítico y el marco conceptual definidos, se seleccionaron las variables relacionadas con los recursos, con las prácticas (en las que se despliega el núcleo de las capacidades) para adquirir la tecnología y con los desempeños tecnológicos, en los que se expresan los resultados de la capacidad de absorción tecnológica, en la dimensión aquí estudiada. De esta forma, se puede diferenciar una lógica más cercana a la perspectiva neoclásica, centrada en la relación *input-output*, esto es, entre los recursos y los desempeños, y otra más cercana a la evolutiva, en la que

entran juego los procesos que median entre los *inputs* y los *outputs*, a través de las prácticas (rutinas) donde se generan el grueso de los aprendizajes.

En lo referente a los recursos, se optó por centrar la atención en los *recursos humanos*, pues existe un amplio consenso en que los conocimientos encarnados en las personas que laboran en las firmas son fundamentales para las capacidades de absorción. En este trabajo se considera el componente del conocimiento individual más fácilmente mensurable, esto es, el nivel de formación formal del personal ocupado —pregrado, tecnológica, posgrado, etc.— (Lall, 1994; Vega *et al.*, 2008).⁴ Además, se considera su ubicación dentro de las áreas funcionales de la empresa y los departamentos de investigación y desarrollo (I+D) o sus sustitutos. Tal información resulta útil para establecer la relación existente entre el desarrollo relativo de los recursos humanos y los desempeños en materia de las tecnologías incorporadas.

Respecto a las *prácticas (o rutinas) para adquirir la tecnología*, se consideraron las desplegadas en los procesos de identificación, selección y negociación de tecnologías (incorporadas en la maquinaria y equipo importadas). Estas prácticas pueden posibilitar una difusión acelerada y creativa de las nuevas tecnologías o, por el contrario, contribuir a acrecentar la dependencia y el atraso tecnológico (Malaver y Vargas, 2006), dependiendo tanto del carácter estratégico de las tecnologías adquiridas como de las capacidades para identificarlas y negociar su transferencia.⁵

En el proceso de *identificación*, las prácticas pueden revelar desde las capacidades más desarrolladas —por ejemplo, cuando la empresa adelanta ejercicios de prospectiva o de vigilancia tecnológica, con sus análisis cuantitativos derivados de consultas de bases bibliográficas o de pa-

⁴ Los otros componentes, como las habilidades, resultantes de los aprendizajes (en su mayoría tácitos), producto de la experiencia, y la creatividad, son muy difícilmente medibles, sobre todo en encuestas.

⁵ Cabe recordar que la contribución de esas tecnologías exógenas incorporadas a la competitividad dependerá también de la capacidad de combinarlas con las existentes en la empresa, de la capacidad de usarlas creativamente y de generar nuevos conocimientos (Nonaka y Takeuchi, 1994; Zahra y George, 2002). Aquí, tal como se dijo, el análisis sólo contemplará su incorporación.

tentes— hasta las más básicas —cuando acude solamente a los proveedores—. En el caso de la *selección*, las prácticas pueden cubrir desde estudios de factibilidad y análisis de riesgo tecnológico hasta el simple chequeo de las características de las tecnologías frente a las especificaciones técnicas requeridas. Los procesos de *negociación*, entre tanto, pueden abarcar desde la transferencia de conocimientos, que permitan el dominio de las tecnologías, su explotación comercial y su uso creativo o la celebración de contratos para la investigación conjunta con el proveedor, hasta la compra pasiva de la tecnología como un artefacto (Malaver y Vargas, 2006).

En lo referente a los desempeños en que se expresan las *capacidades de absorción*, como ya se mencionó, dada la naturaleza especialmente exógena del cambio técnico en nuestros países (Jaramillo *et al.*, 2000), la mirada en este trabajo se centra en el cambio técnico incorporado en las maquinarias y equipos adquiridos por las empresas.⁶

Al respecto, es útil la distinción que establecen Bell y Pavitt (1993), quienes dividen las capacidades en productivas y en tecnológicas. Las capacidades productivas se refieren a los bienes de capital, al conocimiento y a las destrezas laborales que sirven para producir con una tecnología determinada. Por el contrario, las capacidades tecnológicas se refieren a los recursos y habilidades para gestionar y generar el cambio técnico, incluidos los conocimientos, las experiencias y los vínculos institucionales.

En el caso de las capacidades productivas, la adquisición de maquinaria y equipo, que es donde aquí está centrada la indagación, se traduce en incorporaciones que no generan cambio técnico; en el segundo caso, la adquisición genera procesos de aprendizaje, que se traducen en cambio técnico. Cuando esas adquisiciones surgen en medio de procesos de innovación, el cambio técnico no sólo es mayor, sino que es generado endógenamente. Por ello, aunque en general la adquisición de maquinaria y

⁶ Por las limitaciones de la EIByC no se consideraron las tecnologías desincorporadas al capital (licencias), pues en esta, como en la gran mayoría de las encuestas, incluida la Community Innovation Survey (CIS), esas tecnologías hacen parte de las que están vinculadas a los procesos y capacidades de innovación. La distinción entre capacidades de absorción y de innovación desborda los alcances de este trabajo.

equipo conlleva modernización para la empresa, a efectos de establecer su contribución al desarrollo tecnológico (por el cambio tecnológico incorporado), es preciso diferenciar entre la maquinaria y equipo que (Malaver y Vargas, 2009):

- No representa desarrollo tecnológico, esto es, maquinaria y equipo (nuevo o usado) destinados a la producción de bienes o servicios, con características similares a las que tienen los existentes en la empresa. Por lo tanto, son tecnologías convencionales, en el sentido de que ya existen en la empresa. En consecuencia, pueden mejorar las capacidades productivas, pero no las tecnológicas.⁷
- Produce un desarrollo tecnológico, esto es, maquinaria y equipo (nuevo o usado) destinados a la producción de bienes o servicios con características significativamente diferentes o sustancialmente mejoradas frente a las que tienen los existentes en la empresa. De este modo, contienen una nueva tecnología y provocan procesos de aprendizaje y, por lo tanto, se traducen en cambio técnico.
- Está vinculada de manera directa con un proceso de innovación, entendido este como aquel que convierte ideas en productos, procesos, técnicas organizacionales o de *marketing* nuevas o significativamente mejoradas. Dicho proceso de innovación puede deberse tanto a la generación de conocimiento tecnológico como a una aplicación creativa del existente.

En consecuencia, no toda maquinaria y equipo trae consigo cambio técnico incorporado, y esto puede acontecer incluso cuando son nuevos. Para establecer en qué medida lo ha generado puede resultar útil la anterior clasificación.⁸

⁷ En el sentido dado a estas por Bell y Pavitt (1993).

⁸ Estos matices sobre la relación entre la incorporación de maquinaria y equipo y el cambio técnico representa un avance de la Tercera Encuesta colombiana frente a otras encuestas internacionales, que no lo desarrollan (Malaver y Vargas, 2009).

TÉCNICAS DE MINERÍA DE DATOS

El *Knowledge Discovery in Database* (KDD), o minería de datos, es una herramienta para el análisis de grandes volúmenes de datos, de los cuales es necesario extraer o “minar” información que genere conocimiento. Esta permite transformar los datos (materia prima) en información útil (producto), y luego esta se transforma en conocimiento (Polanco, 2009). Para ello, esta se apoya en una serie de técnicas estadísticas avanzadas y de inteligencia artificial, con el fin de descubrir patrones ocultos en la información. El uso de estas herramientas resulta más robusto que el análisis estadístico clásico para describir fenómenos complejos (Morgan y Kaufmann, 2005).

En general, la minería de datos se puede catalogar en dos categorías: minería descriptiva y minería predictiva (Jiawei y Kamber, 2006). La primera caracteriza las propiedades generales inherentes a los datos (patrones estructurales), generando modelos de caja blanca donde la estructura del proceso puede interpretarse; la segunda puede plantear modelos de caja negra para realizar predicciones inferidas de los datos. Estos modelos proponen una solución eficiente interpolando o extrapolando, pero poseen una baja capacidad interpretativa.

Para el análisis de las capacidades de absorción tecnológica, en particular las activadas en los procesos de adquisición de tecnologías externas a las empresas industriales de ByC, aquí se utiliza el KDD con el propósito de extraer reglas de asociación entre las variables y su clasificación. Esto se hace desde una perspectiva descriptiva y con modelos de caja blanca, que muestran patrones estructurales.

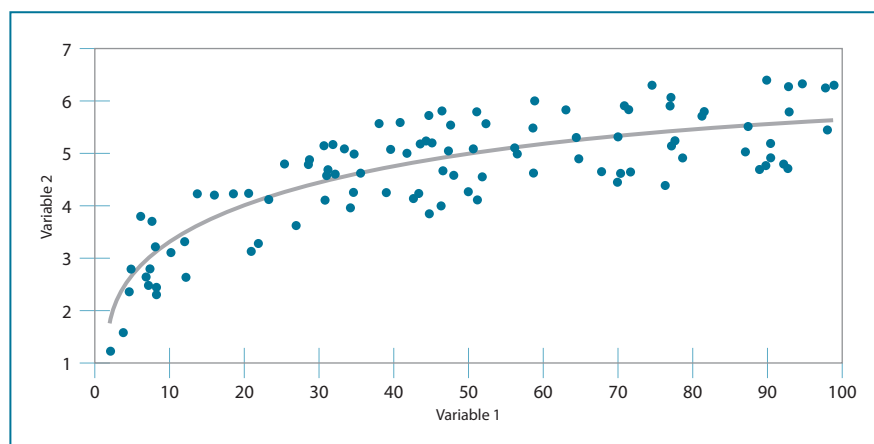
En concordancia con el carácter introductorio al uso de las herramientas del KDD, aquí se realiza un ejercicio que parte de un estudio exploratorio básico, basado en análisis de correlación no lineal y técnicas de visualización, que permiten establecer correlaciones entre las variables y visualizar la textura de los patrones. Luego se intenta identificar patrones y reglas de asociación entre las variables definidas, usando como técnica de descubrimiento los árboles de decisión, que ofrecen la posibilidad de determinar una estructura de reglas común a los datos (Quinlan, 1986). Enseguida se efectúa una breve presentación de esas técnicas.

Análisis exploratorio básico

Para el primer análisis exploratorio de los datos se emplean diagramas de dispersión y de correlación de las variables. La visualización de las variables, aunque parezca trivial, aporta una noción básica de la estructura de los datos. Además, es importante tener un estadístico de la correlación entre las variables. Correlación que, por lo general, no es lineal, y por ello las estadísticas de correlaciones lineales no siempre son un estimativo adecuado. A fin de describir ese tipo de relación, existen estadísticos no lineales como la información mutua promedio (AMI, por su sigla en inglés⁹), basada en la teoría de la información (Shannon, 1948), que cuantifica la información que genera el conocimiento de una variable X_1 , cuando se tiene incertidumbre respecto a otra variable X_2 ; en otras palabras, señala cuánta incertidumbre acerca de la variable X_2 se reduce al conocer X_1 . Así no necesariamente tiene que existir una relación lineal entre las variables para que se vislumbre su asociación (Gráfico 1).

Gráfico 1

Relación no lineal entre dos variables



Fuente: elaboración propia.

⁹ *Average Mutual Information.*

La información mutua entre dos variables (AMI) X_1 y X_2 , notada por $I(X_1, X_2)$, se define como:

$$I(x_1, x_2) = H(x_1) - H(x_1 | x_2)$$

$$I(x_1, x_2) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1)p(x_2)}$$

Donde $H(X_1)$ es la entropía de X_1 ; $H(X_1|X_2)$ es la entropía de X_1 , dado X_2 ; $p(X_1, X_2)$ es la función de probabilidad conjunta entre las variables X_1 y X_2 , $p(X_1)$ y $p(X_2)$ son las funciones de probabilidad marginales.

Ahora bien, dado que la información se calcula en bits, es necesario reescalarla entre 0 y 1, dividiendo la información de las dos variables entre la información de la variable que se conoce, así:

$$\rho = \frac{I(X_1, X_2)}{H(X_1)}$$

La interpretación de este estadístico es similar a la del coeficiente de correlación lineal: entre mayor sea el AMI mayor correlación existe; pero con la diferencia de que un AMI desde 0,1 puede ser significativo (Yao, 2003), es decir, se reduce la incertidumbre de la variable desconocida.

Búsqueda de reglas de asociación

Dado que las bases de datos contienen registros de los fenómenos que pueden asociarse a conceptos o clases, cabe preguntarse por la existencia de reglas estructurales ocultas, expresadas en los patrones generados por el sistema objeto de análisis, el cual puede analizarse como una máquina de inferencia lógica, en la que de acuerdo con una entrada determinada se genera una salida.¹⁰

¹⁰ Un ejemplo se encuentra en el Anexo 1.

Los árboles de decisión¹¹

Un árbol de decisión es un diagrama de flujo con estructura de árbol, donde cada nodo representa un atributo; cada rama, el valor que toma el atributo, y cada hoja, las clases de salida. Los árboles de decisión pueden interpretarse como las reglas de clasificación de un conjunto de datos. También pueden interpretarse como una técnica para reducir la dimensionalidad de un conjunto de datos, basada en la métrica de entropía entre las variables.¹² Algoritmos como el ID3 y J48 buscan reducir el número de inferencias lógicas, midiendo la ganancia de información entre las variables de entrada y la de salida. Esto en contraposición con los métodos lineales, como el de componentes principales, que se basan en la explicación de la variabilidad de los datos.

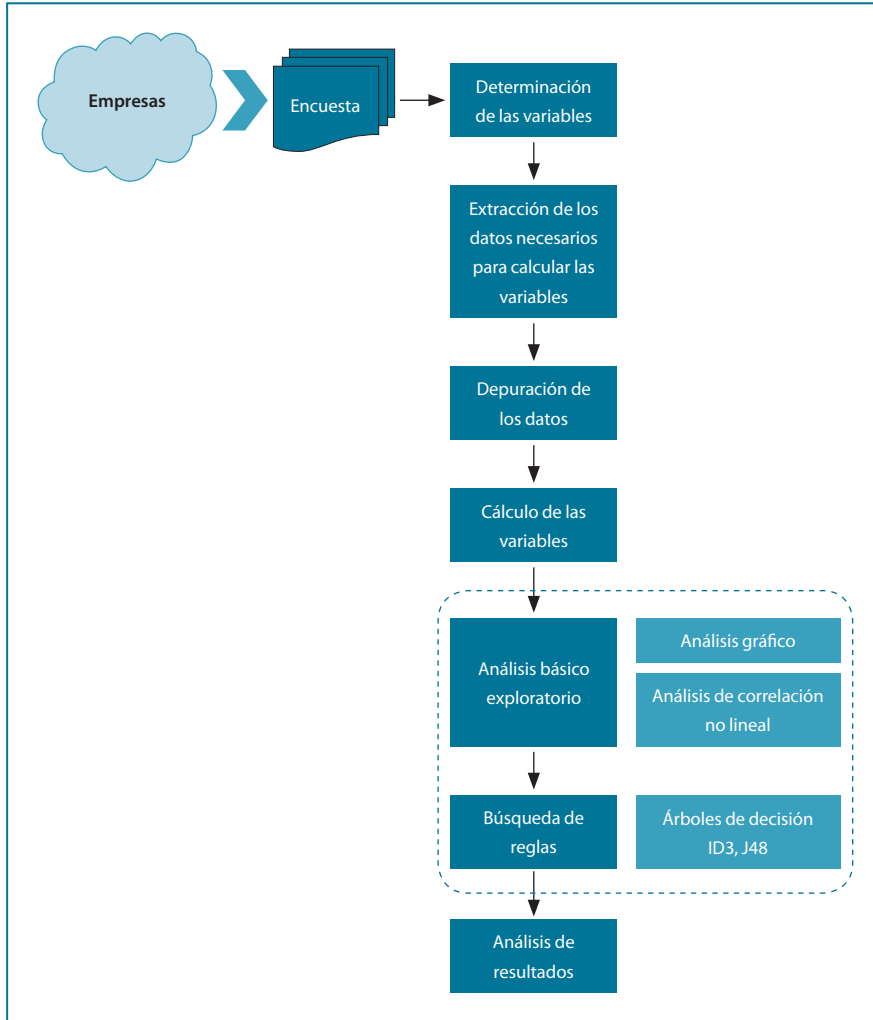
ARTICULACIÓN DEL MARCO CONCEPTUAL Y LAS TÉCNICAS DE KDD

Para realizar el ejercicio de identificación de las capacidades de absorción tecnológica, se articula el marco conceptual y las técnicas de KDD, señaladas en páginas anteriores, siguiendo el protocolo de modelación propuesto por Jiawei y Kamber (2006). En consecuencia, el ejercicio de indagación consta de las actividades señaladas en el Gráfico 2.

¹¹ Véase Anexo 1.

¹² La entropía indica el límite teórico para la compresión de los datos o la información contenida en el mensaje.

Gráfico 2
Metodología de modelación



Fuente: adaptado de Jiawei y Kamber (2006).

Determinación y extracción de las variables

La información utilizada proviene de la EIByC, encuesta realizada en el 2005 por la Cámara de Comercio de Bogotá y el Observatorio Colombiano de Ciencia y Tecnología, cuyos resultados se publicaron en Malaver y

Vargas (2006).¹³ Enseguida se describen las variables construidas a través de las cuales se busca captar los grados de desarrollo de los recursos humanos, de las prácticas para la adquisición de la tecnología (incorporada al capital) y el desempeño en materia de absorción asociado a la incorporación de tecnología (Cuadro 1).

Cuadro 1
Descripción de las variables construidas

Nombre de la variable	Valores	
Recursos humanos		
NivelRH1	1	El máximo grado de formación en la empresa es tecnólogo o educación secundaria
	2	El máximo grado de formación en la empresa es especialización o profesional
	3	El máximo grado de formación en la empresa es doctorado o maestría
NivelRH2	1	El máximo grado de formación en producción o I+D es tecnólogo o educación secundaria
	2	El máximo grado de formación en producción o I+D es especialización y/o profesional
	3	El máximo grado de formación en producción o I+D es doctorado o maestría
Prácticas para adquirir tecnologías		
NivelFuentes (fuentes de identificación)	1	Proveedores/clientes y seguimiento a los competidores
	2	Asistencia a ferias tecnológicas, búsquedas electrónicas, referenciación de empresas internacionales, participación en redes de información tecnológica, consulta de publicaciones especializadas
	3	Ejercicios de vigilancia científica-tecnológica, ejercicios de prospectiva tecnológica

Continúa

¹³ El universo del estudio está conformado por 2.404 empresas manufactureras de Bogotá y Cundinamarca con más de 10 empleados en el 2002 —según los registros mercantiles de la CCB y los directorios industriales del Departamento Administrativo Nacional de Estadística (DANE)—, a partir del cual se realizó un muestreo aleatorio estratificado. El tamaño de muestra fue de 400 empresas, que es representativa para el Código Industrial Internacional Uniforme (CIIU) revisión dos, a tres dígitos, con un error estándar relativo de 3,73%. Además, se calcularon factores de expansión para cada uno de los grupos establecidos de la CIIU, con el fin de estimar los indicadores para el universo estudiado.

Nombre de la variable	Valores	
NivelMétodo (para seleccionar la tecnología)	1	Chequeo frente a especificaciones requeridas
	2	Estudios de factibilidad tecnológica o económica
	3	Análisis del impacto competitivo de la tecnología o análisis del riesgo de la inversión tecnológica
NivelAcuerdos (negociación de la tecnología)	1	Contratos de <i>know how</i> , capacitación en la planta, contratos de servicios tecnológicos, contratos llave en mano
	2	Capacitación en la planta del proveedor, acuerdos para actualización tecnológica posventa, acuerdos para acceder a conocimientos para reparaciones o adaptaciones; para acceder a documentación técnica
	3	Contratos de asistencia técnica para hacer adaptaciones al medio; acuerdos para investigaciones puntuales
Desempeños en la adquisición de tecnologías		
NivelTecnologías (adquiridas)	1	Maquinaria y equipo usado, tecnologías convencionales (similares a las existentes en la empresa)
	2	Nuevas tecnologías que provocaron cambios significativos o saltos frente a los existentes o ventajas tecnológicas
	3	Incorporación para apoyar procesos de innovación

Fuente: elaboración propia a partir de ElByC (2005).

Además, se consideraron variables asociadas con el tamaño y el grado de innovación de las empresas. Para el tamaño (*tamañoEmpresa*) se definieron las siguientes categorías:

1: empresas con menos de 20 empleados y 2: empresas de 20 a 49 empleados, que se denominaron pequeñas empresas.

3: empresas de 50 a 99 empleados y 4: empresas de 100 a 199 empleados, que corresponden a las empresas medianas.

5: empresas con 200 o más empleados, o empresas grandes.

Para el grado de innovación (gradoInnovación) se establecieron las siguientes agrupaciones:¹⁴

NI: las empresas no innovadoras, es decir, aquellas que no han adelantado ningún esfuerzo o actividad para innovar.

PI: las empresas potencialmente innovadoras, las que han adelantado actividades de innovación, pero aún no han alcanzado desarrollos (en proceso o en producto) novedosos para el mercado donde compiten.

IA: las empresas innovadoras en sentido amplio.

IE: las empresas innovadoras en sentido estricto, es decir, aquellas que han obtenido innovaciones de producto o proceso, como resultado de actividades formales de I+D y que son nuevas para el mercado internacional o patentadas.

Depuración y cálculo de los datos

Para el análisis exploratorio básico se extrajo a un *Data Warehouse* la información de la base de datos de la EIByC necesaria para el cálculo de las variables. Y con esta se generaron las diferentes configuraciones de visualización de los datos. Con base en el criterio experto se calcularon los diferentes valores categoriales de las variables y se establecieron los diferentes patrones del modelo.

La aplicación de los árboles de decisión exige que los patrones de entrenamiento que se usen para calibrar un árbol ID3 o J48 sean congruentes, es decir, para una entrada sólo puede existir una salida. En los datos obtenidos de las encuestas se presentaron múltiples incongruencias, por lo que fue necesario generar un algoritmo para depurar los patrones, agregando etiquetas *dummy*, a fin de eliminar las ambigüedades. Con los patrones depurados se corrieron diferentes configuraciones de patrones de entrada.

¹⁴ Esta clasificación parte de la propuesta en Durán *et al.* (1998) y de los ajustes realizados por Malaver y Vargas (2006).

Análisis exploratorio básico

Inicialmente se indagó por la relación entre el nivel de formación de los recursos humanos y el desempeño en la adquisición. Luego, se auscultó por la asociación entre las prácticas de adquisición y los desempeños en términos de la tecnología incorporada. Y en ambos casos se trató de verificar si el tamaño y el grado de innovación son factores discriminantes en esas relaciones. Para ello se utilizaron diagramas de dispersión y análisis de la correlación no lineal entre las variables.

Búsqueda de reglas

Luego del análisis exploratorio básico, en un tercer momento, se indagó por la asociación entre los niveles de formación de los recursos humanos, el desarrollo de las prácticas de gestión para adquirir las tecnologías y los desempeños alcanzados. Para ello, se buscó identificar los patrones ocultos en la información sobre dichas variables, utilizando la técnica de los árboles de decisión.

ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

En esta sección se presentan los resultados obtenidos de la aplicación de las técnicas de minería de datos, descritas en la sección anterior, con el fin de ahondar en la identificación de las capacidades de absorción desplegadas en la adquisición de las tecnologías incorporadas por las empresas industriales de ByC.

RESULTADOS DEL ANÁLISIS EXPLORATORIO

Como es propio de un análisis exploratorio, la indagación por la asociación entre los recursos humanos y el desempeño alcanzado en materia de absorción de conocimiento externo por parte de las empresas, mediante

la adquisición de la tecnología, inició con una mirada a la relación entre el número de empleados con diferentes niveles de formación en toda la empresa (variable de entrada) y las distintas características de la tecnología adquirida (variable de salida), en términos del cambio tecnológico que pudo traer incorporado. Esto se hizo contrastando los resultados por tamaños de empresa y de acuerdo con los grados de innovación obtenidos por las empresas.

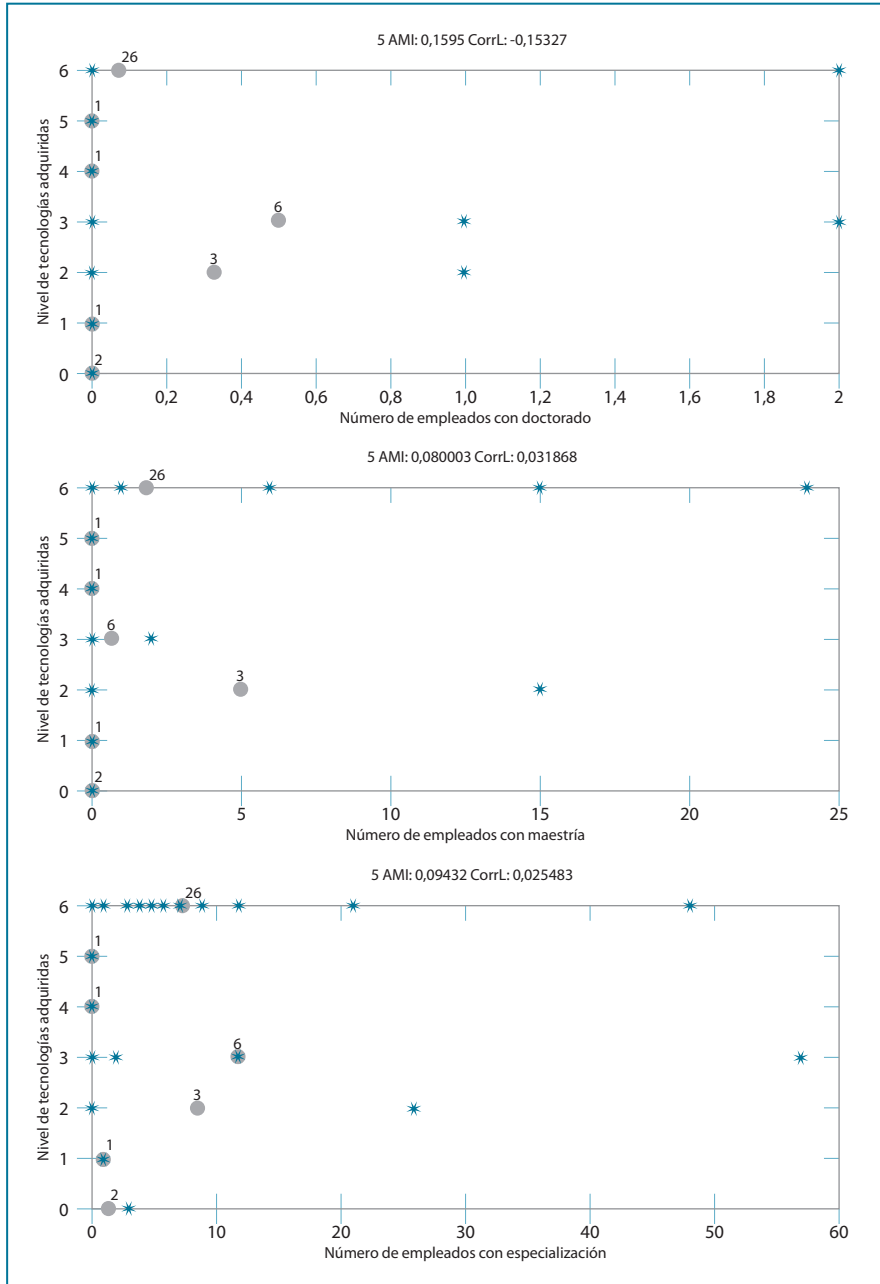
En ese primer ejercicio se utilizaron las variables correspondientes de la EIByC y, en ninguno de los dos casos, esto es, ni por tamaños o grados de innovación, se identificaron patrones claros en la relación. A título de ejemplo, en el Gráfico 3 se presentan los diagramas de dispersión para el caso de las empresas grandes.¹⁵ El gráfico muestra que, de acuerdo con el AMI, existe una correlación significativa para el caso de los empleados profesionales, tecnólogos y con doctorado. Pero según las medias, denotadas por los círculos, no es posible visualizar, en ninguno de los casos, un patrón claro entre los recursos humanos y los desempeños. Cabe advertir que esa situación se repite cuando el lente se enfoca en el personal ocupado en las áreas de producción y de I+D.

En un segundo momento se indaga por la correlación entre las prácticas usadas para la adquisición de la tecnología —en los procesos de identificación, selección y negociación—, que corresponden a las variables de entrada, y el desempeño como variable de salida. Entre estos tres procesos sólo se identifican algunos patrones de relación en el caso de los acuerdos producto de la negociación para la adquisición de la tecnología.

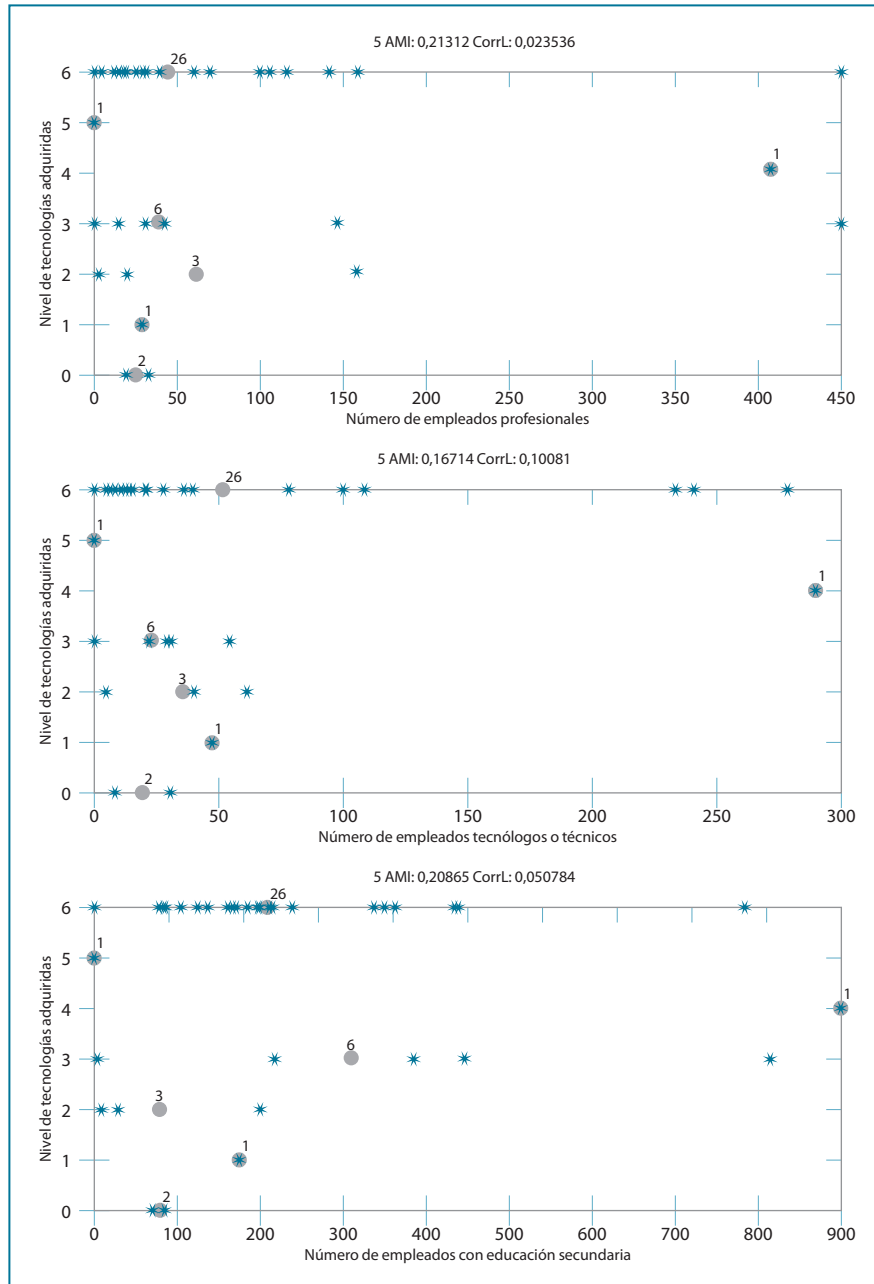
¹⁵ En el diagrama de dispersión, el número puesto en la parte superior izquierda (el 5) indica el tamaño de la empresa (grande en este caso); junto a este aparece el estadístico AMI, así como el de correlación lineal (CorrL), dentro del gráfico; los círculos indican la media del número de empleados y el nivel de tecnologías adquiridas, para el número de empresas que aparece junto al círculo. Por ejemplo, hay 6 empresas (grandes) que tienen en promedio 0,5 empleados con doctorado y un nivel de tecnologías de tres.

Gráfico 3

Relación entre recursos humanos y adquisición de tecnología, según diagramas de dispersión en empresas grandes (tipo 5)



Continúa

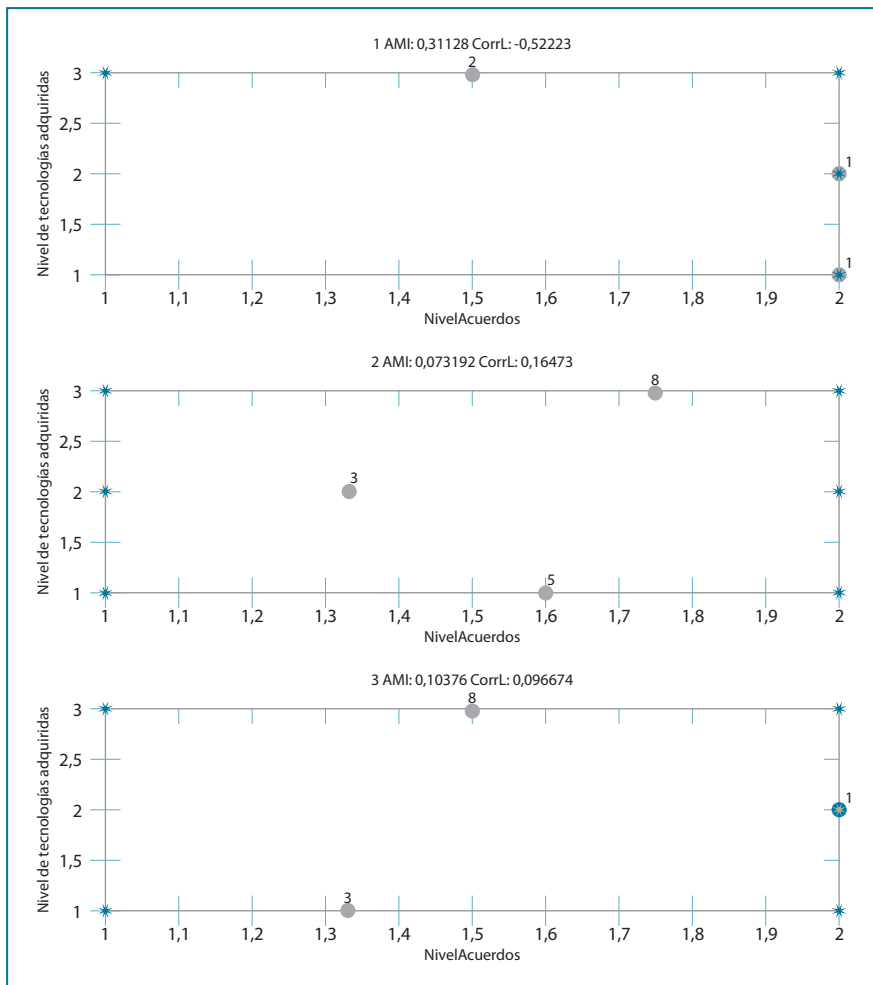


Fuente: elaboración propia con base en datos de la EIByC (2005).

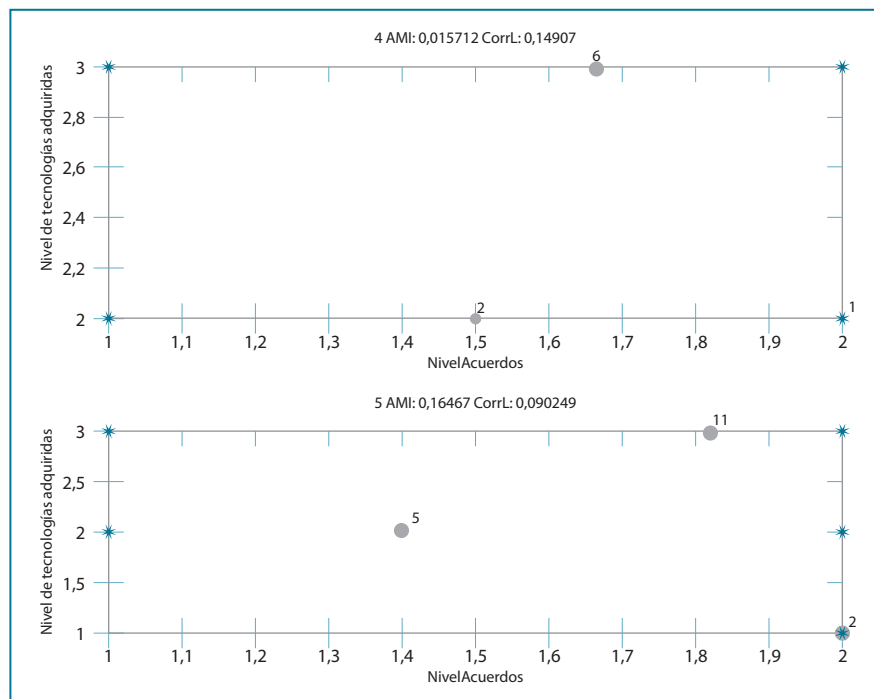
Tal como se puede visualizar en el Gráfico 4, excepto en el caso de las empresas más pequeñas (1 en el gráfico), en las que este tipo de acuerdos no se presentan, en las demás se visualiza una relación positiva entre el grado de complejidad de los acuerdos bajo los cuales se negocia la tecnología y el grado de absorción tecnológica alcanzado con las tecnologías adquiridas. Sin embargo, la estructura de correlación no es muy clara.

Gráfico 4

Relación entre prácticas de negociación y adquisición de tecnología, según diagramas de dispersión por tamaños de empresa



Continúa



Fuente: elaboración propia, a partir de los datos de la EIByC (2005).

El mismo ejercicio se hizo para las prácticas de identificación y selección de la tecnología, pero los resultados no fueron concluyentes. En el primer caso, las empresas grandes (tamaño 5) exhiben las prácticas más desarrolladas, aunque son mínimos los casos en que se realizan ejercicios de vigilancia o prospectiva tecnológica para identificar los nuevos desarrollos tecnológicos que se van a incorporar. No obstante, la correlación es baja. Para los otros tamaños no es posible visualizar patrones de tendencia en los datos.

En el segundo caso se puede constatar que las empresas grandes y medianas aplican los métodos más sofisticados de selección de la tecnología —como los análisis de impacto competitivo o de riesgo tecnológico asociado a la maquinaria y equipo incorporado—, que revelan mayores capacidades de absorción acumuladas. Sin embargo, la correlación es muy baja según el AMI.

En vista de que los resultados anteriores fueron poco concluyentes, se exploraron todas las posibles relaciones entre los recursos humanos, las prácticas y los desempeños en la incorporación de las tecnologías; pero en este caso se utilizaron las variables construidas (y presentadas en el Cuadro 1), que toman los valores 1, 2 y 3, según los grados de desarrollo que reflejan. Para ello se calculó la correlación de todas las posibles combinaciones entre las variables.

Los resultados arrojados por el ejercicio de exploración de la estructura de las correlaciones, que se sintetizan en el Cuadro 2, indican que en general existe una alta correlación entre las prácticas y el grado de absorción (desempeño) alcanzado mediante las tecnologías adquiridas. La mayor correlación (0,664) se presenta entre la complejidad de los métodos de selección y el desempeño; la segunda más alta se registra entre las prácticas para identificar las tecnologías y el grado de absorción representado por estas, y la tercera, entre las prácticas de negociación y el desempeño.

Cuadro 2
Correlación entre las variables de análisis

Variable	Variable	AMI
nivelFuentes	nivelMétodo	0,86797641
nivelMétodo	nivelTecnologías	0,66412007
nivelFuentes	nivelTecnologías	0,65393836
nivelMétodo	nivelAcuerdos	0,52714879
nivelFuentes	nivelAcuerdos	0,51749516
nivelAcuerdos	nivelTecnologías	0,45565517
nivelTecnologías	gradoInnovación	0,30914455
tamañoEmpresa	nivelTecnologías	0,09867053
nivelMétodo	gradoInnovación	0,09691862
nivelMétodo	tamañoEmpresa	0,08980256
nivelFuentes	tamañoEmpresa	0,07722069
nivelRH2	tamañoEmpresa	0,06632928
nivelRH1	tamañoEmpresa	0,06482818

Continúa

Variable	Variable	AMI
nivelAcuerdos	tamañoEmpresa	0,06057764
nivelFuentes	gradoInnovación	0,05966964
nivelRH1	nivelRH2	0,05533315
nivelRH2	nivelAcuerdos	0,05471586
nivelAcuerdos	gradoInnovación	0,053897
nivelRH1	nivelTecnologías	0,05331664
tamañoEmpresa	gradoInnovación	0,05142712
nivelRH1	nivelAcuerdos	0,04974498
nivelRH1	nivelMétodo	0,04674582
nivelRH2	gradoInnovación	0,04391337
nivelRH1	nivelFuentes	0,04186351
nivelRH1	gradoInnovación	0,03204323
nivelRH2	nivelMétodo	0,02809249
nivelRH2	nivelFuentes	0,02713568
nivelRH2	nivelTecnologías	0,02693407

Fuente: elaboración propia a partir de los datos de la EIByC (2005).

Cabe advertir que también existe una alta correlación entre los niveles de las distintas prácticas de adquisición, es decir, entre las fuentes de información para identificar las tecnologías, los métodos para seleccionarlas y los mecanismos de negociación y acuerdos para su incorporación, tal como lo ilustra el Cuadro 2.

En la misma dirección, el ejercicio muestra que existe una correlación relativamente alta entre el grado de absorción tecnológica y los grados de innovación alcanzados por las empresas. Hecho que, por lo demás, resulta consistente con lo señalado por la literatura, en cuanto a que los procesos de innovación son los mayores generadores de aprendizajes que se acumulan y permiten mejores capacidades de absorción de conocimiento, las cuales se traducen en mejores prácticas y desempeños en la adquisición, asimilación, transformación y explotación de las tecnologías exógenas. Pero, como se ha señalado, ese ámbito de análisis desborda los alcances de este estudio.

Por el contrario, los resultados muestran unas correlaciones bastante pequeñas entre los niveles de formación de los recursos humanos y el desempeño en materia de adquisición. Igual acontece con la correlación entre los recursos humanos y las prácticas para la adquisición de las tecnologías. Y esto sucede cuando se consideran los niveles de formación en toda la empresa (nivelRH1) o cuando se analiza sólo el personal ocupado en las áreas de producción y de I+D (nivelRH2). Esto, por lo demás, resulta consistente con los resultados arrojados por los ejercicios anteriores.

RESULTADOS DE LA BÚSQUEDA DE REGLAS

Luego del análisis exploratorio se hizo un ejercicio orientado a identificar los patrones ocultos en la información sobre las variables usadas para indagar por las tres dimensiones del análisis: recursos humanos, prácticas y desempeños de adquisición. Para ello, como se mencionó, se utilizó la técnica de los árboles de decisión.¹⁶

Inicialmente se generaron árboles mediante el algoritmo ID3 implementado en la herramienta Weka, el cual clasifica la totalidad de los patrones suministrados para la calibración. Dado que la dimensionalidad de los árboles generados por el conjunto de patrones de las empresas fue muy elevada, se recurrió al algoritmo J48, que realiza una poda selectiva de las ramas, facilitando la interpretación de estos. Entre los patrones mostrados por el árbol de decisión representado en el Gráfico 5 se destacan las siguientes reglas en las empresas industriales de ByC:

- Si el nivel del método de selección de la tecnología es alto, entonces el desempeño en la adquisición tecnológica es alto (en 30 casos).¹⁷

¹⁶ Algunas indicaciones —pertinentes dado el carácter introductorio del uso de estas técnicas— sobre la forma de leer la información arrojada por los árboles para identificar las reglas ocultas se presenta en el Anexo 2.

¹⁷ Frente a un total de 144 casos considerados en el ejercicio.

- Si el nivel del método de selección de la tecnología es medio y el nivel de formación de los recursos humanos en la empresa es medio, entonces el desempeño en la adquisición tecnológica es alto (en 36 casos).
- Si el nivel del método de selección de la tecnología es medio, el nivel de formación de los recursos humanos en la empresa es alto y las prácticas de negociación son altas, entonces el desempeño en la adquisición tecnológica es alto (en 5 casos).

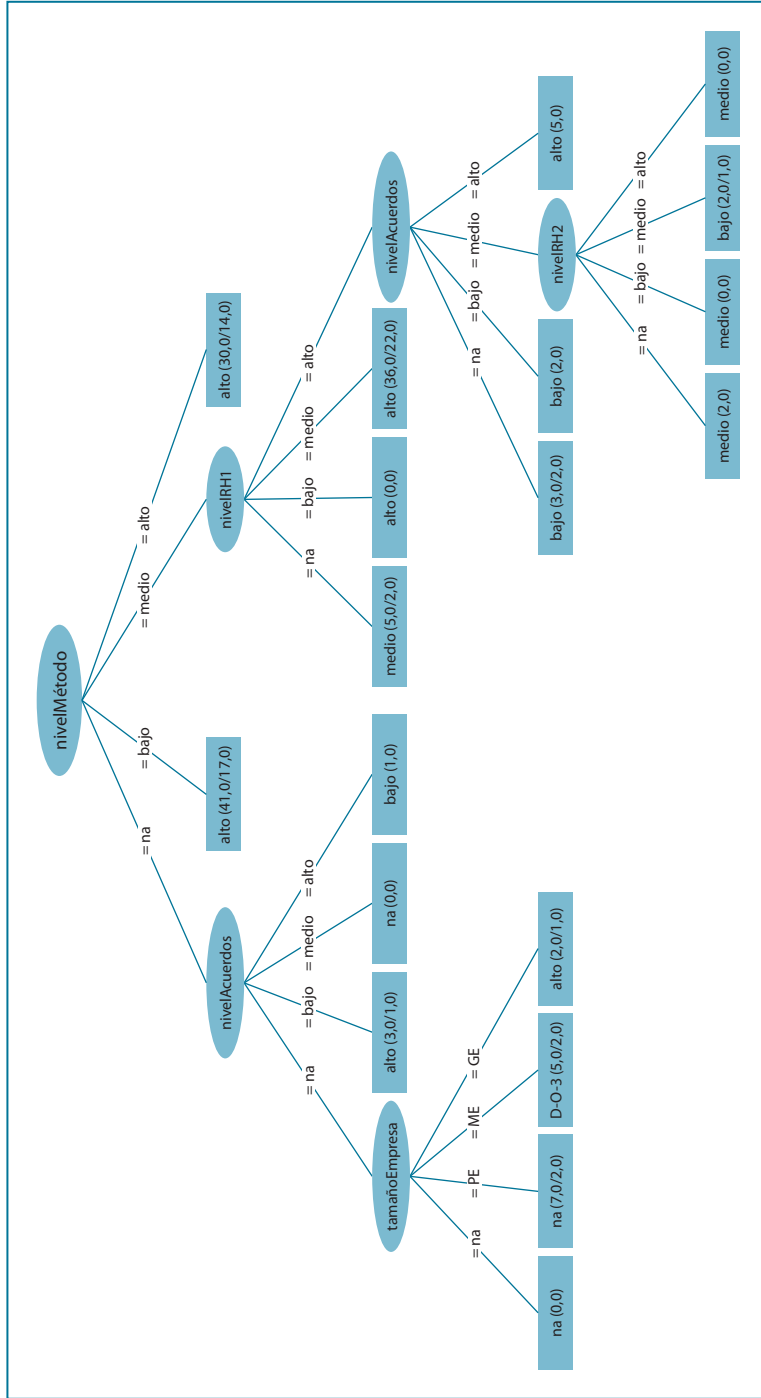
En este ejercicio se vuelven a evidenciar como relevantes las prácticas de gestión desplegadas para la adquisición de las tecnologías, en este caso en particular, los métodos para seleccionarlas y las prácticas para negociaciárlas, las cuales aparecen mediadas por el nivel de formación de los recursos humanos en toda la planta, antes que los concentrados en producción o actividades de I+D.

Un número importante de casos sigue otro patrón, claro, pero opuesto a los anteriores, el cual indica que si el nivel del método de selección de la tecnología es bajo, entonces el desempeño en la adquisición tecnológica es alto. En los restantes patrones el número de casos es muy bajo o la información es incompleta, razón por la cual no se ahonda en su análisis.

INTERPRETACIÓN DE LOS RESULTADOS

Los resultados arrojados por estos análisis efectuados en el ámbito de la empresa y apenas exploratorios indican que los hallazgos en otros contextos (países y empresas) en torno a que el capital humano, evaluado, por el nivel de formación de las personas ocupadas en las empresas está asociado de manera positiva y significativa con las capacidades de absorción del conocimiento, en este caso, no se cumplen (Lall, 2001; Vega *et al.*, 2008).

Gráfico 5
Reglas de asociación entre recursos humanos, prácticas y desempeño: árbol de decisión tipo J48



Fuente: elaboración propia partir de datos de la EIBYC (2005).

En ello pueden incidir varios factores presentes en las empresas industriales de ByC: el mínimo porcentaje de personas con formación de maestría y doctorado en las empresas, el hecho de que la inmensa mayoría de ellas laboren en el área administración, que las personas con menor nivel de formación de toda la empresa se encuentren en el área de producción, o que las personas que laboran en los departamentos de I+D o sus sucedáneos representen menos del 1% del total (Malaver y Vargas, 2006).

El hecho de que, en general, en las áreas técnicas el recurso humano sea el menos formado o que la participación del que labora en I+D sea marginal en las empresas puede traducirse en que incidan muy poco en la cultura y las estrategias empresariales, en las cuales la tecnología y la innovación no tienen un papel central, fenómeno en el cual coinciden distintos trabajos efectuados en el país (Vargas *et al.*, 2003; Calderón *et al.*, 2009). En ese sentido, es posible que el foco y las mayores capacidades empresariales se encuentren en las áreas comerciales o de producción, como señala para otros contextos Rangone (1999), o como sugieren a título de hipótesis Malaver y Vargas (2006), para la industria de ByC.

En esos resultados también pueden incidir razones de índole metodológica. Por ejemplo, no haber usado porcentajes de trabajadores según niveles de formación, que permitieran mejorar la comparación entre las diferentes perspectivas desde las cuales se abordaron las empresas (por tamaños y grados de innovación), o no mirarlas de acuerdo con tipologías relacionadas con la intensidad en el uso de las tecnologías.

En todo caso, los resultados del análisis precedente indican que aquello que en otros contextos es una buena premisa de partida, en la industria de ByC es un fenómeno que debe estudiarse en profundidad, si se quieren formular políticas realistas. De otro modo, estas serán, cuando menos, ineficaces.

Ahora bien, los resultados sugieren, aunque no de manera concluyente, que las capacidades de absorción acumuladas y desplegadas en las prácticas para adquirir las tecnologías externas tienen una incidencia positiva e importante en el grado del cambio tecnológico incorporado y en los aprendizajes suscitados por el tipo y forma de adquisición de la maquinaria y equipo.

Esos resultados sugieren finalmente que, en este caso, la relación directa entre *inputs* y *outputs* no es automática, que en los procesos se ponen en juego capacidades acumuladas y aprendizajes que inciden en grado significativo en los resultados (desempeños alcanzados). Esto resulta favorable a los planteamientos evolutivos, incluso cuando se usa información proveniente de instrumentos como las encuestas de innovación que, por su naturaleza, resultan bastante limitadas para captar fenómenos tan complejos y dinámicos como los asociados a las capacidades de absorción.

CONCLUSIONES

En los ejercicios efectuados no se encuentra una relación clara entre el nivel de formación de los recursos humanos y los desempeños alcanzados en términos de la absorción de conocimiento externo, por la vía de la adquisición de la tecnología. En consecuencia, antes que asumirse como una premisa de partida la relación positiva entre el capital humano y el cambio técnico, en el caso estudiado esta relación debe considerarse como problemática y, por lo tanto, debe estudiarse con mayor profundidad, a fin de diseñar políticas e instrumentos de estímulo más acordes con esta realidad.

Los resultados muestran, por el contrario, que la asociación entre las capacidades para la absorción de ese conocimiento, expresadas en las prácticas de incorporación de las tecnologías exógenas a las empresas y el desempeño alcanzado en su adquisición es positiva. Por lo tanto, en el diseño de las políticas debería estimularse de manera explícita este factor, y en las encuestas, ser objeto de indagación.

Estos resultados, dado su carácter exploratorio, no son concluyentes. Sin embargo, dejan abiertos promisorios interrogantes en torno a la medición de las capacidades de absorción y sugieren la necesidad de efectuar estudios y utilizar herramientas metodológicas que permitan ahondar en la caracterización de este fenómeno, vital para el desarrollo tecnológico de países y regiones en vía de desarrollo.

REFERENCIAS

- Arundel, A. (2007). Innovation survey indicators: What impact on innovation policy? En *Science, technology and innovation: Indicators in a changing world, responding to policy needs*. Paris: OCDE.
- Bell, M. y Pavitt, K. (1993). Technological accumulating and industrial growth: Contrasts between developer and developing countries. *Industrial and Corporate Change*, 2 (2), 157-202.
- Calderón, G.; Álvarez, C. y Naranjo, J. (2009). Orientación estratégica y recursos competitivos: un estudio en grandes empresas industriales de Colombia. *Cuadernos de Administración*, 22 (38), 49-72.
- Cohen, W. M. y Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128-152.
- Dahlman, C. y Nelson, R. (1993, junio). *Social absorption capability, national innovation systems and economic development*, documento presentado en la UNU/INTECH Research Conference, Maastricht, Países Bajos.
- Durán, X.; Ibáñez, R.; Salazar, M. y Vargas, M. (1998). *La innovación tecnológica en Colombia: características por tamaño y tipo de empresa*. Bogotá: DNP.
- Grant, R. (2005). *Contemporary strategy analysis*. London: Blackwell.
- Hitt, M.; Ireland, D. y Hoskinsson, R. (1999). *Administración estratégica: competitividad y conceptos de globalización*. México: Thomson.
- Hurtienne, T. y Messner, D. (1999). Nuevos conceptos de competitividad internacional en países industrializados y en países en desarrollo. En K. Esser (ed.), *Competencia global y libertad de acción nacional* (pp. 41-68). Caracas: Nueva Sociedad.
- Jaramillo, H.; Lugones, G. y Salazar, M. (2000). *Manual para la normalización de indicadores de innovación tecnológica en América Latina y el Caribe: manual de Bogotá*. Bogotá: OEA/RICYT/Tres Culturas.
- Jiawei, H. y Kamber, M. (2006). *Data mining concepts and techniques* (2a ed.). San Francisco: Elsevier.
- Kim, L. y Nelson, R. (2000). Introduction. En L. Kim y R. Nelson (eds.), *Technology, learning and innovation: Experiences of newly industrializing economies* (pp. 1-12). Cambridge: Cambridge University Press.
- Lall, S. (1994). Las capacidades tecnológicas. En J. J. Salomon, F. Sagasti y C. Sachs (comps.), *Una búsqueda incierta-ciencia, tecnología y desarrollo*. México: Editorial de la Universidad de Naciones Unidas, Centro de Investigación y Docencia Económicas y Fondo de Cultura Económica.
- . *Competitiveness, technology and skills*. London: Edward Elgar.

- Lee, W. (2000). The role of science and technology policy in Korea's industrial development. En L. Kim y R. Nelson (eds.), *Technology, learning and innovation: experiences of newly industrializing economies* (pp. 269-290). Cambridge: Cambridge University Press.
- Malaver, F. y Vargas, M. (2006). *Capacidades tecnológicas, innovación y competitividad de la industria de Bogotá y Cundinamarca: resultados de una encuesta de innovación*. Bogotá: Cámara de Comercio de Bogotá, Observatorio Colombiano de Ciencia y Tecnología, Agenda Regional de Ciencia y Tecnología y Consejo Regional de Competitividad.
- . (2009). *Los indicadores de innovación en Colombia: diagnóstico y propuestas para su desarrollo*. Bogotá: Observatorio Colombiano de Ciencia y Tecnología y Colciencias.
- Morgan, X. y Kaufmann, X. (2005). *Data mining practical machine learning tools and techniques* (2a ed.). San Francisco: Elsevier.
- Nelson, R. y Sampat, B. (2001). Las instituciones como factor que regula el desempeño económico. *Economía Institucional*, 3 (5), 18-51.
- Nonaka, I. y Takeuchi, H. (1994). *La organización creadora de conocimiento*. México: Oxford.
- Pérez, C. (2001). Cambio tecnológico y oportunidades de desarrollo como blanco móvil. *Revista de la CEPAL*, (75), 115-136.
- Polanco, X. (2009, 5 de marzo). *Análisis de la información: un programa*. Documento procedente de la conferencia dictada en la Pontificia Universidad Javeriana, Bogotá, Colombia.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1 (1), 81-106.
- Rangone, A. (1999). A resource-based approach to strategy analysis in small-medium sized enterprises. *Small Business Economics*, 12, 233-248.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Vargas, M.; Malaver, F. y Zerda, A. (eds.), (2003). *La innovación tecnológica en la industria colombiana*. Bogotá: OCyT y Colciencias.
- Vega, J.; Gutiérrez, A. y Fernández de Lucio, I. (2008). *An analytical model of absorptive capacity* (working paper 2008/2). Valencia: Universidad Politécnica de Valencia.
- Yao, Y. Y. (2003). Information-theoretic measures for knowledge discovery and data mining. En *Entropy measures, maximum entropy principle and emerging applications* (pp. 115-136). Berlin: Springer.
- Zahra, S. y George, G. (2002). Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review*, 27 (2), 185-203.

ANEXO 1

EJEMPLO BÚSQUEDA DE REGLAS DE ASOCIACIÓN

Un ejemplo para ilustrar la búsqueda de reglas de asociación es el problema del clima en la decisión de juego de un partido de tenis (Quinlan, 1986). En este caso existen cuatro atributos de entrada: pronóstico, temperatura, humedad y viento, y uno de salida: si la persona juega o no al tenis. Estos atributos están etiquetados en diferentes categorías como lo muestra el Cuadro 1.

Cuadro 1

Patrones del ejemplo del clima

Pronóstico	Temperatura	Humedad	Viento	Juega
Soleado	Caliente	Alta	FALSO	No
Soleado	Caliente	Alta	VERDADERO	No
Nublado	Caliente	Alta	FALSO	Sí
Lluvioso	Templado	Alta	FALSO	Sí
Lluvioso	Frío	Normal	FALSO	Sí
Lluvioso	Frío	Normal	VERDADERO	No
Nublado	Frío	Normal	VERDADERO	Sí
Soleado	Templado	Alta	FALSO	No
Soleado	Frío	Normal	FALSO	Sí
Lluvioso	Templado	Normal	FALSO	Sí
Soleado	Templado	Normal	VERDADERO	Sí
Nublado	Templado	Alta	VERDADERO	Sí
Nublado	Caliente	Normal	FALSO	Sí
Lluvioso	Templado	Alta	VERDADERO	No

Fuente: Quinlan (1986).

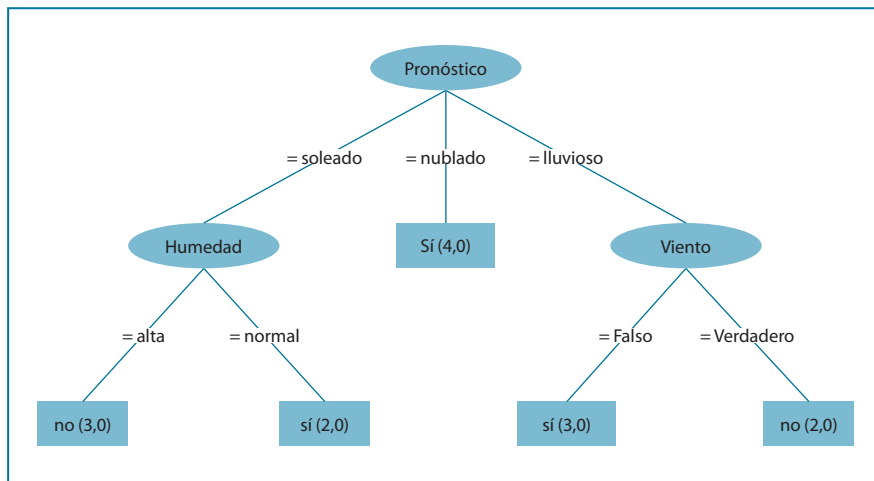
En este caso, el pronóstico y la temperatura tienen tres categorías; la humedad y el viento, dos, y la variable de decisión, dos. Esto genera 36 posibles combinaciones, de las cuales 14 están presentes en la tabla de patrones. Por ejemplo, un patrón se expresa así:

Si el pronóstico=soleado y la temperatura=caliente y la humedad=alta y viento=falso, entonces no juega tenis.

En el peor de los casos es necesario indagar por cada una de las variables en secuencia para conocer la salida. Por lo tanto, la pregunta que surge es: ¿existe una secuencia de inferencia más corta para llegar al mismo resultado?

Para ello son útiles los árboles de decisión. Por ejemplo, de acuerdo con el árbol de decisión que se presenta en el Gráfico 1, se encuentra la regla: “si el pronóstico es igual a nublado, entonces sí juega tenis”, donde tan sólo es necesario conocer el valor que toma la variable pronóstico para saber la decisión de juego.

Gráfico 1
Árbol de decisión del problema del clima



Fuente: Quinlan (1986).

ANEXO 2

LECTURA DE LOS RESULTADOS DE LOS ÁRBOLES DE DECISIÓN

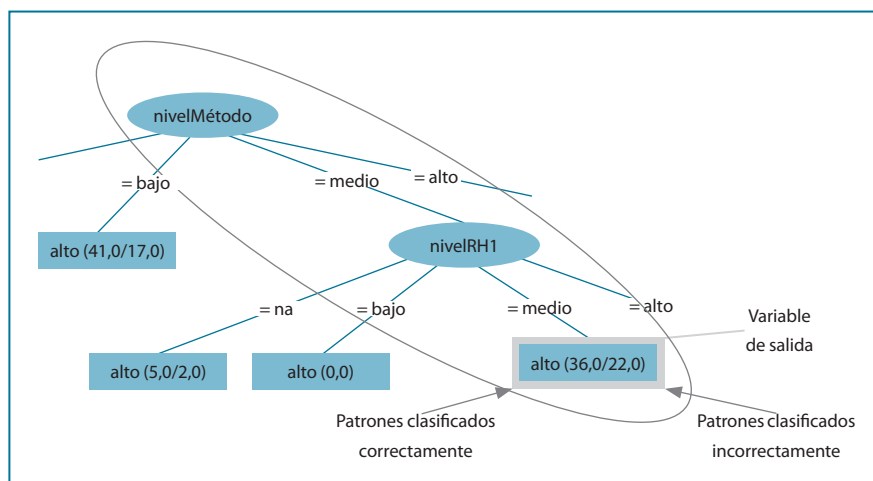
En los árboles de decisión aplicados a la información de la EIByC (2005), las reglas (o patrones) se leen de la siguiente forma:

Si *nivel del método de selección=medio (2)* y *nivel de RH1=medio (2)* entonces *el desempeño en la adquisición=alta (3)*.

En el Gráfico 1 se presenta un ejemplo de la forma como se puede leer la información mostrada por el árbol:

Gráfico 1

Lectura información de un árbol de decisión



Fuente: elaboración propia a partir de los datos de la EIByC (2005).

ÁRBOL 1: CAPACIDAD DE ABSORCIÓN

Patrones clasificados correctamente: 80 (55,6 %).

Patrones clasificados incorrectamente: 64 (44,4%).

Total de patrones: 144.

Capítulo 7

Internacionalización del financiamiento de las actividades de desarrollo tecnológico de las empresas establecidas en Colombia

Fredy Alexander Gómez Jiménez
Jorge Robledo Velásquez

La colaboración internacional de tipo empresarial en actividades de ciencia y tecnología (CyT) ha sido una actividad que tradicionalmente se ha llevado a cabo con mayor frecuencia entre organizaciones de países desarrollados, e incluso hoy en día esta actividad continúa siendo muy fuerte. Sin embargo, un importante crecimiento de este tipo de relacionamiento ha tomado forma en aquellos países llamados en desarrollo. De acuerdo con la OCDE (2008) y particularizando en las actividades de investigación y desarrollo experimental (I+D), las empresas extranjeras ubicadas en países en desarrollo incrementaron sus gastos de I+D entre la mitad y tres veces la cantidad invertida en el país de origen, y aunque los gastos aún son modestos, su crecimiento es muy relevante. En otro estudio, en este caso de la Comisión Europea (European Commission, 2007), se encontró que los flujos de gastos de I+D, los cuales solían estar dirigidos hacia países desarrollados, parecen estar virando a favor de las economías emergentes.

Varias de las razones que explican este fenómeno de desplazamiento de capitales están asociadas a la intensificación de la competencia global, los rápidos cambios tecnológicos, el acceso al personal de I+D en países en desarrollo y el aprovechamiento de la diferencia de costos entre los países en desarrollo y desarrollados (Reddy, 1997). Richards y Yang (2007) argumentan que las corporaciones multinacionales (CMN), las cuales solían concentrar sus actividades de I+D sólo en países desarrollados (Vonortas, 1997; Katz *et al.*, 1996), están ahora comenzando a invertir en países en desarrollo, dado que la infraestructura para realizar actividades de investigación ha mejorado, los mercados locales están creciendo y los costos de I+D son considerablemente más bajos.

De todas maneras, los efectos generales actuales de la globalización, como la estandarización de los productos de forma global y las facilidades de comunicación alrededor del mundo, han facilitado enormemente la globalización de actividades de I+D y de capacitación tecnológica. Adicionalmente, los cambios del mercado laboral en países en desarrollo ha incrementado considerablemente el atractivo de establecer filiales en estos países.

En Colombia, la internacionalización de las actividades de CyT ha sido un componente tradicional de la política pública. En la actualidad, la política vigente tiene como objetivo fundamental consolidar la proyección internacional del país en esta dimensión, que les facilita a los centros de I+D y grupos relacionados el acceso a los recursos financieros e intelectuales internacionales, a fin de favorecer la inserción de la ciencia, la tecnología y la innovación colombianas en el contexto internacional (Colombia-Colciencias, 2008). Para poder alcanzar este objetivo se contemplan como líneas de acción: la búsqueda y utilización de fuentes internacionales de cooperación, el apoyo a la movilidad de investigadores e innovadores, el aprovechamiento de la diáspora en CyT y el desarrollo de la cooperación horizontal con otros países caribeños y de Latinoamérica.

El interés del presente capítulo se centra en el acceso a recursos financieros internacionales por parte de las empresas establecidas en el país, como una medida del grado de internacionalización de las actividades empresariales de innovación y desarrollo tecnológico y como un aporte al

diagnóstico de la realidad colombiana en esta materia de interés político público. En este sentido, el estudio busca contribuir al conocimiento de un fenómeno que se ha estudiado muy poco en Latinoamérica (Grosse, s. f.). En particular, el capítulo estudia las características de la financiación internacional de las actividades de desarrollo tecnológico (ADT) realizadas por las empresas establecidas en el país, según se desprende de los resultados de la Segunda Encuesta Colombiana de Innovación.

El capítulo se estructura en cuatro secciones: en la primera se presentan las definiciones y clasificaciones que utiliza la Segunda Encuesta de Innovación para el tratamiento del fenómeno bajo estudio; en la segunda se reseña la metodología utilizada, basada, por una parte, en técnicas descriptivas para realizar un primer examen exploratorio de los datos y, por la otra, en técnicas de análisis multivariante (TAM), para reducir variables e identificar y caracterizar grupos de empresas similares; en la tercera se exponen y analizan los resultados, y, finalmente, se presentan las conclusiones del estudio.

DEFINICIONES, CLASIFICACIONES Y VARIABLES DE ANÁLISIS

Para la Segunda Encuesta de Innovación, las ADT son “todas aquellas acciones llevadas a cabo por la empresa, tendientes a poner en práctica conceptos, ideas y métodos necesarios para la adquisición, asimilación e incorporación de nuevos conocimientos” (Colombia-DANE, DNP y Colciencias, 2005: 23). Según esta misma fuente, estas actividades cubren las siguientes tecnologías:

Tecnologías incorporadas al capital: Incorporación a la empresa de conceptos, ideas y métodos, a través de la compra de maquinaria y equipo con desempeño tecnológico mejorado (incluso software integrado) vinculado con las innovaciones implementadas por la empresa. Conforman lo que se conoce como cambio técnico “incorporado”. Esto constituye nuevos conocimientos adquiridos a través

del análisis y uso de nuevos procesos mecánicos, materiales de partes y piezas y en general de nuevos conceptos e ideas incorporadas en la maquinaria. (p. 24).

Tecnologías de gestión: La inversión en tecnologías de gestión comprende la adquisición de conocimientos y el procesamiento de información orientados a ordenar, disponer, organizar, graduar o dosificar el uso de los recursos productivos para obtener mayor productividad o competitividad. (p. 25)

Tecnologías transversales: Estas inversiones corresponden a la incorporación de conceptos, ideas y métodos como resultado de una actividad de investigación llevada a cabo, ya sea en una forma rutinaria o no, por fuera de la empresa o a pedido de ésta. Estas tecnologías incluyen inversión en actividades de transferencia o adquisición de tecnología, inversión en actividades de transferencia o adquisición de tecnología, inversión en tecnologías de la información y la comunicación (TICs), inversión en actividades de biotecnología y programas de diseño industrial. (p. 33)

Proyectos de I+D: Comprenden el trabajo creativo emprendido sistemáticamente para incrementar el acervo de conocimientos, y el uso de este conocimiento para concebir nuevas aplicaciones. Pueden incluir el desarrollo de prototipos y plantas piloto. Un proyecto de I+D puede ser de investigación básica, aplicada o de desarrollo experimental. (p. 35)

Capacitación tecnológica: Comprende la capacitación en temas estrechamente relacionados con las tecnologías centrales en el proceso productivo de la empresa. Estas tecnologías pueden ser “blandas” (gestión y administración) o “duras” (tecnología de procesos productivos), que involucran un grado de complejidad significativo (no evidente) que requiere de un personal capacitador altamente especializado. (p. 36)

En relación con las entidades internacionales de financiación de ADT, las cuales serán objeto de análisis en el apartado “Reconocimiento de insti-

tuciones financieras para el financiamiento de ADT”, la Segunda Encuesta de Innovación, a través del *Manual de diligenciamiento* (Colombia-DANE, DNP y Colciencias, 2005), las clasifica y define como se muestra a continuación:

Banca de inversión: Se consideran establecimientos de crédito las instituciones financieras internacionales cuya función principal consista en captar en moneda legal recursos del público en depósitos, a la vista o a término, para colocarlos nuevamente a través de préstamos, descuentos, anticipos u otras operaciones activas de crédito. (p. 78)

Banca comercial internacional: Son instituciones financieras con el objeto de realizar operaciones activas de crédito, entre países. (p. 78)

Aportes casa matriz: Se refiere a los recursos propios de la empresa provenientes del país de origen de la misma. (p. 78)

Organismos internacionales —OEA, ONU, UE—: Se refiere a los recursos provenientes de organismos de cooperación técnica internacional como la Organización de Estados Americanos, Organización de las Naciones Unidas, Unión Europea, etc. (p. 78)

Cooperación internacional: Se refiere a los recursos provenientes de países que tienen programas de cooperación técnica, los cuales se brindan en su mayoría a través de proyectos en diferentes áreas. (p. 78)

Programa CARANA: La Agencia Internacional para el Desarrollo del gobierno de los Estados Unidos (USAID) ha asignado a la empresa CARANA la dirección de un proyecto de desarrollo empresarial, que busca incrementar rápidamente los niveles de ventas, empleo y rentabilidad de las pymes colombianas. (p. 78)

Finalmente, las variables “tamaño de empresa” y “tipo de empresa” (según el origen extranjero o nacional del capital) que se utilizan para caracterizar los *clusters* o agrupaciones encontradas en el estudio se definen como sigue: por el tamaño, una empresa es micro si su número de empleados es menor a 10; pequeña, si está entre 10 y 50; mediana, si está entre

51 y 200, y grande si es mayor a 200. En cuanto al tipo de empresa, una empresa se considera extranjera si su capital social de origen extranjero es mayor al 50%; en caso contrario, se considera nacional.

METODOLOGÍA

ORIGEN DE LOS DATOS

El principal insumo para realizar el trabajo está conformado por los datos producidos por la Segunda Encuesta Colombiana de Innovación —la Segunda Encuesta de Innovación y Desarrollo Tecnológico en el Establecimiento Industrial Colombiano (EDIT II)—, aplicada en 2005 para los años 2003 y 2004. Los datos se obtuvieron por entrevista directa a 6.172 establecimientos industriales y tuvo cobertura nacional representativa de agrupaciones industriales según el Código Industrial Internacional Uniforme (CIIU). La unidad de selección y observación es el establecimiento industrial, definido como la unidad económica que, bajo una forma jurídica única o un solo propietario y en una sola ubicación física, se dedica a la producción del grupo más homogéneo posible de bienes manufacturados.

La información se recolectó a través de las direcciones territoriales y subsedes del Departamento Administrativo Nacional de Estadística (DANE), en un período de cuatro meses (marzo-junio de 2005). Esta encuesta incluyó los establecimientos industriales que ocupan 10 o más personas o que obtengan una producción anual igual o superior a COL\$65 millones en el año de referencia.

HERRAMIENTAS PARA EL ANÁLISIS DE DATOS

El estudio tiene un primer enfoque descriptivo para analizar las entidades internacionales de financiamiento de ADT y un enfoque no supervisado de técnicas estadísticas exploratorias de análisis multivariante para analizar

la colaboración internacional en actividades de I+D y capacitación tecnológica (CT). Se realizó un preprocesamiento de los datos, que consistió en la identificación y tratamiento de *outliers* o datos atípicos para las variables seleccionadas, y se construyeron histogramas y *BoxPlots*.

Las TAM se emplearon para la reducción o clasificación de variables (análisis de factor y análisis de *cluster*), y fueron útiles en el hallazgo de perfiles y características de grupos de empresas similares.

RESULTADOS

Los resultados se presentan en dos apartados. El primero es, en esencia, un estudio descriptivo acerca de las entidades internacionales de financiamiento de ADT. El segundo apartado presenta los resultados del estudio de reducción y clasificación de variables mediante las TAM.

RECONOCIMIENTO DE INSTITUCIONES FINANCIERAS PARA EL FINANCIAMIENTO DE ACTIVIDADES DE DESARROLLO TECNOLÓGICO

Con base en la pregunta de la encuesta: “¿Reconoce o ha escuchado la siguiente entidad internacional de financiamiento?” y la clasificación de las fuentes de financiación presentada en la sección “Definiciones, clasificaciones y variables de análisis”, se construye el Cuadro 1, el cual muestra que la visibilidad de las entidades de financiamiento oscila entre el 22,3% y el 7,2%, valores muy por debajo de lo esperado, lo que revela un profundo desconocimiento de las posibilidades de financiamiento internacional de las ADT que ofrecen oportunidades para las empresas. Una causa de este desconocimiento podría ser la falta de difusión o publicidad comercial de parte de las entidades financieras para dar a conocer sus productos; sin embargo, en el Capítulo 9 de este libro, Otálora, Hurtado y Quimbay exploran una hipótesis alternativa, en la cual favorecen una explicación fundamentada en la falta de interés o motivación de los empresas por conocer las fuentes financieras de las ADT.

Cuadro 1

Reconocimiento (visibilidad) de las entidades de financiamiento

¿Reconoce la entidad de financiamiento?	Banca de inversión (%)	Banca internacional (%)	Casa matrix (%)	OEA, ONU, UE (%)	Cooperación internacional (%)	CARANA (%)
Sí	22,2	22,3	7,5	17,7	8,9	7,2
No	77,8	77,7	92,5	82,3	91,1	92,8
Total	100,0	100,0	100,0	100,0	100,0	100,0

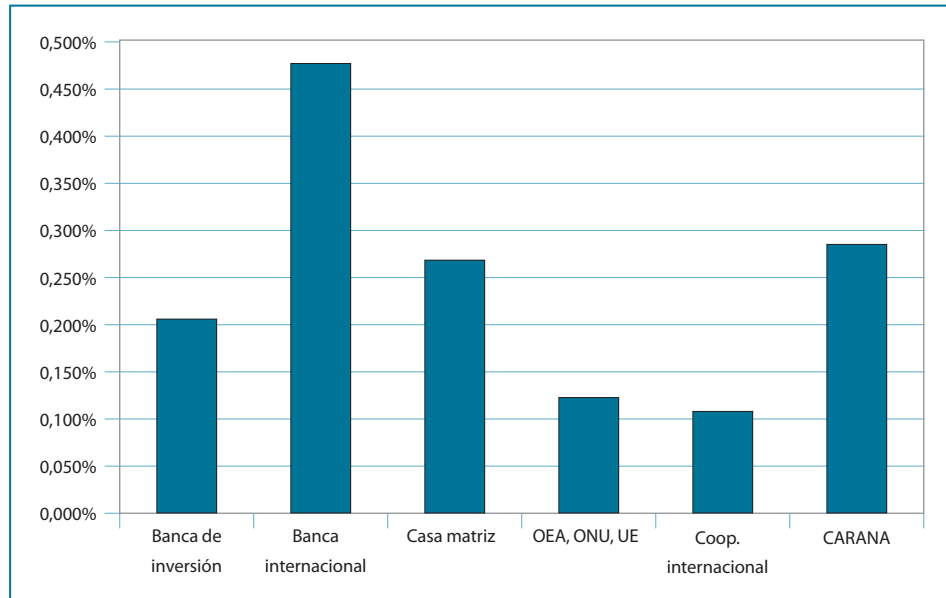
Fuente: elaboración propia a partir de datos de la EDIT II.

Luego de señalar lo anterior, del Cuadro 1 cabe también destacar la visibilidad que posee el sistema bancario (banca de inversión y banca internacional) entre las entidades financiadoras de las ADT, con porcentajes de respuestas positivas cercanas al 22% en cada caso. De hecho, del total de respuestas afirmativas, el sistema bancario posee el 52% de ellas. En contraste, la cooperación internacional es prácticamente invisible como fuente de financiación de las ADT; al tiempo que un programa específico, CARANA, es casi tan visible como toda la cooperación internacional junta.

Más allá de la visibilidad de las entidades de financiamiento, el interés podría recaer en la efectiva utilización de estas entidades como fuente de recursos para financiar las ADT de la empresa. A este respecto, y según los datos de la encuesta, sólo 14 de las 6.212 firmas usaron una fuente de financiamiento internacional. En otras palabras, únicamente el 1% de las firmas estudiadas emplearon en algún momento una o varias de las fuentes de financiación mencionadas para las ADT. Como se muestra en el Gráfico 1, la banca internacional se convierte en la vía más utilizada por las empresas para la consecución de recursos (0,47% del total de las firmas estudiadas).

Gráfico 1

Porcentaje de firmas que utilizaron recursos para las ADT provenientes de las fuentes financieras internacionales



Fuente: elaboración propia a partir de datos de la EDIT II.

FINANCIACIÓN INTERNACIONAL DE ACTIVIDADES DE I+D Y DE CAPACITACIÓN TECNOLÓGICA

Este apartado se enfoca en la financiación de las actividades de I+D y de CT; al mismo tiempo identifica el grado de participación del capital extranjero. El Cuadro 2 muestra el porcentaje de recursos financieros que las firmas establecidas en Colombia invirtieron en actividades de I+D y en actividades de CT durante 2003 y 2004. Estas últimas actividades son, sin lugar a dudas, el enfoque privilegiado de desarrollo tecnológico adoptado por la gran mayoría de las compañías, puesto que casi la mitad de los

recursos reportados se destinan a esta actividad. Como era de esperarse, la investigación básica posee la menor participación con un porcentaje del 3%.

Cuadro 2

Distribución de los gastos en I+D y capacitación tecnológica

Actividad	Total invertido (US\$)*	Porcentaje
Investigación básica	2.948	3
Investigación aplicada	20.675	21
Desarrollo experimental	30.195	30
Capacitación tecnológica	46.692	46
Total	100.510	100

* Valores en miles de dólares. La conversión se hizo de acuerdo con el promedio de la tasa de cambio en los años 2003 y 2004, sin ajuste inflacionario.

Fuente: elaboración propia a partir de datos de la EDIT II.

Para el análisis se optó por aplicar un enfoque pasivo o no supervisado y realizar pruebas de reducción de variables y asociación sobre las variables seleccionadas, de modo que se pudiera observar qué información y patrones de interés emergían. Se seleccionaron las variables relacionadas con la inversión en actividades de investigación aplicada, desarrollo experimental y actividades de capacitación tecnológica durante el período 2003-2004. La variable relacionada con la inversión en investigación básica no fue tomada en cuenta, por ser un valor nulo en la casi totalidad de establecimientos colombianos (Cuadro 3).

Cuadro 3

Variables seleccionadas para el estudio

Variable	Código
Inversión en actividades de investigación aplicada en el 2003-2004	INVAPLIC
Inversión en actividades de desarrollo experimental en el 2003-2004	DESEXP
Inversión en actividades de capacitación tecnológica en el 2003-2004	CAPTECN

Fuente: elaboración propia.

El propósito es identificar agrupaciones de empresas o *clusters* con una característica o perfil similar que permita analizarlos y caracterizarlos. Primero se realizó un *análisis de factor* para reducir las variables y, posteriormente, un *análisis de cluster* para identificar las agrupaciones. A fin de verificar la adecuación entre datos y técnica, se realizó una prueba de Kaiser-Meyer-Olkin (KMO o medida de adecuación) y un test de esfericidad de Barlett con su significancia sobre las tres variables presentadas en el Cuadro 3.

Los valores de 0,582 y 0,000 se obtuvieron para la prueba de KMO y para la significancia del test de esfericidad respectivamente, lo que permite concluir que los datos son adecuados para un análisis de factor (KMO>0,5 y valor de $p<0,05$). Al realizar la extracción de los factores por medio del *software* SPSS, se obtuvieron dos factores que explican el 76% de la varianza, por medio de la metodología de los componentes principales. El Cuadro 4 presenta la relación encontrada entre los dos factores y las tres variables seleccionadas y referidas a las actividades de I+D y CT.

Cuadro 4

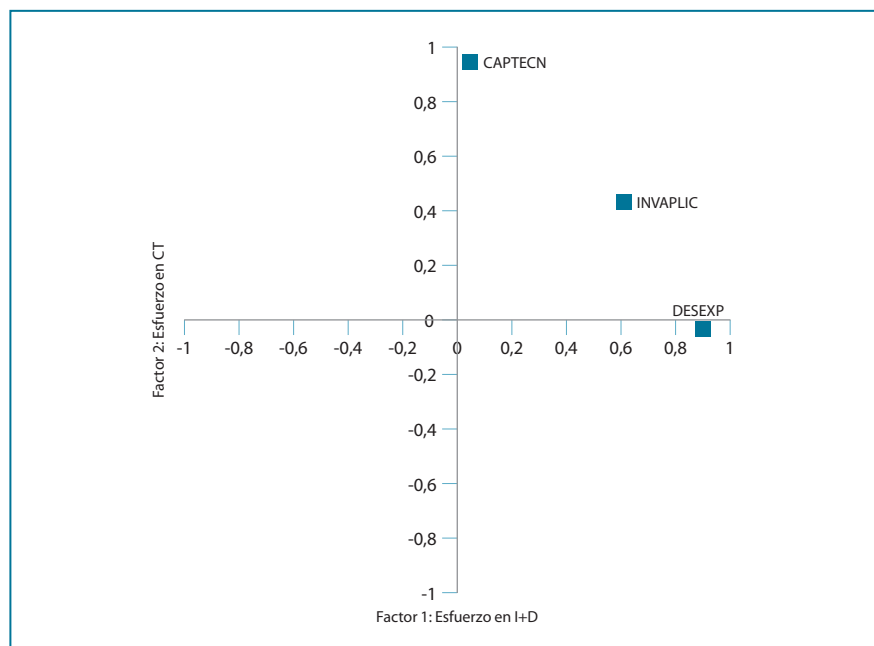
Factores encontrados mediante el análisis de componentes principales

Código	Factor 1	Factor 2
INVAPLIC	0,615	0,433
DESEXP	0,903	-0,030
CAPTECN	0,050	0,946

Método de rotación: varimax con normalización de Kaiser (converge en tres iteraciones).
Fuente: elaboración propia.

Como se puede apreciar en el Gráfico 2, el factor 1 tiene afinidad con los esfuerzos en I+D (investigación aplicada y desarrollo experimental) a través de los recursos invertidos, mientras que el factor 2 representa la variable relacionada con la CT. El factor 1 se etiqueta como esfuerzos en I+D; el factor 2, como esfuerzos en CT.

Gráfico 2
Componentes del análisis de factor



Fuente: elaboración propia.

Posteriormente se realizó un análisis de *cluster* sobre las puntuaciones de los dos factores encontrados anteriormente. Al aplicar el análisis por medio de la metodología de las K-medias, el *software* SPSS reveló los dos *clusters* que se presentan en el Cuadro 5 y que son etiquetados como los establecimientos investigadores y los establecimientos capacitadores.

Cuadro 5
Nombre y número de establecimientos en cada *cluster*

Nombre del cluster	Número de establecimientos
Investigadores	20
Capacitadores	6.196
Total	6.216

Fuente: elaboración propia.

Características del *cluster* 1: los establecimientos investigadores

Estas firmas se caracterizan por que realizan las mayores inversiones en actividades de I+D y de CT respecto al total de la muestra. Sólo corresponden al 0,32% del total, lo que evidencia la poca iniciativa que tiene la empresa promedio para realizar actividades investigativas. Estas empresas poseen en promedio 1.275 empleados, por lo que es un *cluster* conformado principalmente por grandes empresas. El 40% son empresas cuyo capital extranjero es mayor o igual al 50% de la composición accionaria, y el 60% son empresas de capital mayoritariamente nacional, lo que muestra un amplio predominio relativo de las empresas con capital extranjero frente a las de capital nacional, teniendo en cuenta que el número absoluto de estas es muy superior.

Los recursos obtenidos para financiar actividades de I+D provienen principalmente de fuentes nacionales, a excepción de dos empresas de las 20, de las cuales una utilizó recursos provenientes de Alemania para realizar investigación aplicada, y la otra, recursos provenientes de Estados Unidos para investigación aplicada y desarrollo experimental. En otras palabras, los establecimientos investigadores se caracterizan por no utilizar recursos provenientes del exterior, sino de fuentes nacionales, para financiar sus actividades investigativas.

Características del *cluster* 2: los establecimientos capacitadores

Casi todas las empresas de la muestra pertenecen a este *cluster* (99,7%), por lo que sus características son representativas del establecimiento industrial colombiano. Este *cluster*, a diferencia de los establecimientos investigadores, no dedica recursos significativos en actividades de I+D ni en CT; sin embargo, la mayoría de las firmas que pertenecen a este *cluster* tienen alguna inversión en actividades de CT, obviamente sin llegar alcanzar los niveles encontrados en los establecimientos investigadores. Es decir, este *cluster* muestra que la firma promedio en Colombia no invierte significativamente en I+D y solamente de forma tímida en actividades de CT, de donde proviene el nombre de establecimientos capacitadores.

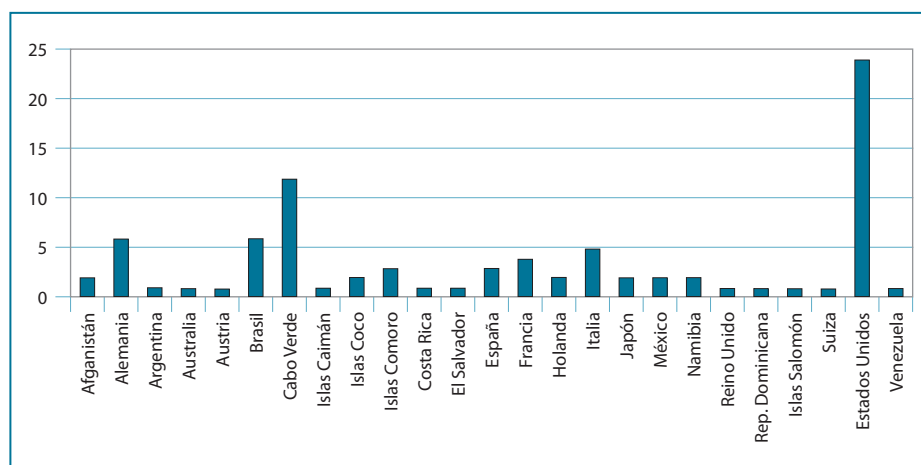
En promedio, las empresas de este *cluster* están conformadas por 81 empleados, lo que evidencia son representativas de la pyme colombiana. Sólo el 5% de las firmas poseen un capital accionario extranjero mayor o igual al 50%, mientras que el 95% son consideradas empresas nacionales.

A semejanza de los establecimientos investigadores, los cuales mostraron una participación muy pobre en el financiamiento extranjero de sus actividades, el *cluster* de los establecimientos capacitadores financia sus actividades con recursos de fuentes nacionales (alrededor del 93% de las empresas). Cerca del 97% de los recursos invertidos provienen de tales fuentes y solamente el 7% de las empresas del *cluster* utilizaron recursos de fuentes extranjeras para financiar sus actividades.

El Gráfico 3 presenta un desglose del 3% de los recursos financieros provenientes de fuentes extranjeras, según su país de procedencia y el número de empresas financiadas. Se observa el predominio estadounidense (24 empresas financiadas), seguido de Cabo Verde y, más atrás, de Alemania y Brasil. Las fuentes financieras para las demás empresas están dispersas en un número alto de países, cada uno de los cuales financia menos de cinco empresas.

Gráfico 3

Número de establecimientos financiados por país

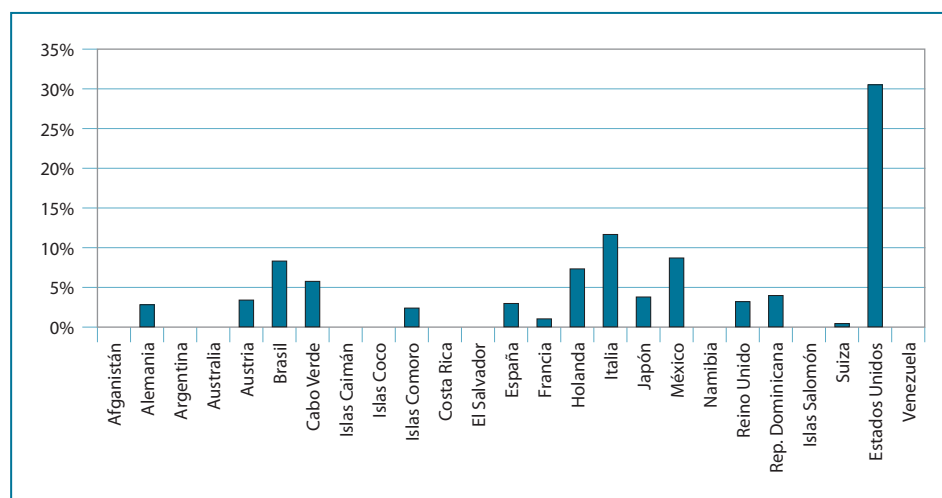


Fuente: elaboración propia.

Estados Unidos no sólo es el país que más número de empresas financia del *cluster* de establecimientos capacitadores, sino que también es el que más recursos aporta respecto a los demás países. El Gráfico 4 ilustra esta situación, en la que se aprecia que esa nación aporta alrededor del 30% del total de los recursos provenientes del exterior.

Gráfico 4

Porcentaje de la financiación total extranjera por país



Fuente: elaboración propia.

En el Gráfico 4, de nuevo, se observa la amplia dispersión de los recursos según su fuente, si se descuenta el aporte estadounidense. Sin embargo, si se consideran los países europeos en conjunto (Alemania, Austria, España, Francia, Holanda, Italia, Suiza e Inglaterra), la participación sería muy relevante y sobrepasaría incluso la de Estados Unidos en el aporte extranjero para la financiación de actividades de I+D y CT.

CONCLUSIONES

Respecto a la financiación internacional de las ADT de las empresas estudiadas y considerando las fuentes previamente identificadas, el trabajo

evidencia un acceso muy limitado a los recursos provenientes de tales fuentes, con sólo el 1% del total de las empresas que reportaron usar, por lo menos, una de las fuentes. Este resultado descriptivo muestra un escaso flujo de recursos financieros del exterior hacia Colombia para este tipo de actividades.

Esta situación se corresponde con la poca visibilidad de las entidades financieras internacionales como fuentes de recursos para ADT. De esta visibilidad, la mayor parte la acapara el sistema bancario (banca de inversión y banca internacional), aun cuando también es notable la cooperación internacional; pero por el hecho de que es prácticamente invisible como fuente de recursos empresariales para ADT.

En resumen, las ADT de las empresas colombianas están escasamente internacionalizadas en lo que respecta al flujo de recursos financieros externos. Este hecho, junto con la pobre visibilidad de las fuentes internacionales de recursos y la baja intensidad de las actividades empresariales de I+D y CT, configura un escenario que debe ser motivo de examen y tratamiento en la doble perspectiva de la política pública y la gestión empresarial.

Por otra parte, el estudio muestra que los recursos de financiación de actividades de I+D y de CT provienen principalmente de fuentes nacionales, tanto para el *cluster* de los establecimientos investigadores como para el de los capacitadores. Sin embargo, las actividades de CT en este último *cluster* generan una financiación internacional mucho más fuerte y diversa en sus fuentes que la encontrada para el *cluster* de los investigadores.

Las fuentes de financiamiento internacional de actividades de CT para el *cluster* de los establecimientos capacitadores se sitúan principalmente en Estados Unidos y Europa (como suma total de los países que la componen); no obstante, como país individual, Estados Unidos posee un liderazgo en número de firmas financiadas y en porcentaje sobre el total financiado frente a los demás países, los cuales muestran una amplia dispersión en número y una escasa participación en empresas financiadas y recursos aportados.

De los dos *clusters* encontrados, los establecimientos investigadores se caracterizan por su esfuerzo en I+D, el tamaño grande de sus empresas y la alta proporción de capital accionario extranjero. Estas, sin embargo, no

son características representativas de la muestra de empresas del estudio. Tal representatividad recae en el *cluster* de establecimientos capacitadores, que invierten muy poco en I+D y tímidamente en CT. El tamaño de sus empresas es predominantemente pequeño y mediano, y su composición accionaria está dominada por capital nacional.

REFERENCIAS

- Beers, C. van; Berghäll, E. y Poot, T. (2008). R&D internationalization, R&D collaboration and public knowledge institutions in small economies: Evidence from Finland and the Netherlands. *Research Policy*, 37, 294-308.
- Behrman, I. N. y Fischer, W. A. (1980). Overseas R&D. En R. Nelson, G. Silverberg y L. Soete (eds.), *Activities of transnational companies* (pp. 496-527). London: Oelgeschlager Pinter Publishers.
- Colombia. DANE, DNP y Colciencias (2005). *Manual de diligenciamiento de la segunda encuesta*. Bogotá: autores.
- Colombia. Colciencias (2008). *Colombia construye y siembra futuro: política nacional de fomento a la investigación y la innovación*. Bogotá: autores.
- Dunning, J. H. (1992). Multinational enterprises and the globalization of innovatory capacity. En O. Granstrand, L. Hakanson y S. Sjölander (eds.), *Technology management and international business: internationalization of R&D and technology* (pp. 19-51). Sussex: John Wiley & Sons.
- Europe Commission (2007). *Europe in the global research landscape*. Brussels: autor. Recuperado el 21 de octubre de 2009, de ftp://ftp.cordis.europa.eu/pub/foresight/docs/prospective-report2007_en_12_web.pdf.
- Freeman, C (1993). *The economics of industrial innovation* (3rd ed.). Londres: Frances Pinter.
- Grosse, R. (s. f.). *R&D by TNCs in Latin America*. En prensa.
- Hidalgo, A.; Fernández, C. y López, J. (2002). *Evaluación de los 10 de años de IBEROEKA*. Madrid: Fundación General de la Universidad Politécnica de Madrid.
- Katz, R.; Rebutisch, E. S. y Allen, T. J. (1996). A study of technology transfer in multinational cooperative joint venture. *IEEE Transactions on Engineering Management*, 43 (1), 97-105.
- Organization for Economic Co-operation and Development (OCDE), (2005). *Oslo manual: Guidelines for collecting and interpreting innovation data* (3rd ed.). Paris: European Commission, Organization for Economic Co-operation and Development, Statistical Office of the European Communities.

- Organization for Economic Co-operation and Development (OCDE), (2008). *Recent trends in the internationalization of R&D in the enterprise sector: Special session on globalization*. Paris: autor.
- Reddy, P. (1997). New trends in globalization of corporate R&D and implications for innovation capability in host countries: A survey from India. *World Development*, 25 (11), 1821-1837.
- Richards, M. y Yang, Y. (2007). Determinants of foreign ownership in international R&D joint ventures: Transaction costs and national culture. *Journal of International Management*, 13, 110-130.
- Schumpeter, J. (1939). *Business cycle: A theoretical and statistical analysis of the capitalist process*. New York: McGraw Hill.
- Vonortas, N. S. (1997). Research joint ventures in the US. *Research Policy*, 26, 577-595.

Capítulo 8

La financiación de la innovación: un análisis a partir de la Encuesta de Innovación de Bogotá y Cundinamarca¹

Jaime Humberto Sierra González
Florentino Malaver Rodríguez
Marisela Vargas Pérez

Este trabajo indaga por la relación entre la financiación y la innovación en la industria colombiana y profundiza en los aspectos micro en el caso particular de Bogotá y Cundinamarca (ByC). Para ello se concentra en las fuentes y los mecanismos de financiación a los que acuden las empresas cuando innovan, en la existencia de restricciones financieras y en la efectividad de los instrumentos de política pública de apoyo a la financiación de la innovación.

¹ Una versión previa de este trabajo fue presentada en el I Congreso de Gestión Tecnológica e Innovación, realizado en la Universidad Nacional de Colombia, Bogotá, Colombia, en agosto de 2008. Para la realización de los ejercicios de análisis de correspondencia se contó con el apoyo de Édgar Bueno, estadístico y miembro del Observatorio Colombiano de Ciencia y Tecnología (OCyT). Entre tanto, en la realización de los análisis econométricos se contó con el apoyo de David Londoño, PhD en Economía y profesor de la Universidad Tecnológica de Bolívar.

El marco que guía el análisis es producto de la articulación de elementos analíticos provenientes de las perspectivas de la asimetría de la información: la evolutiva y la institucionalista. Para el análisis micro, centrado en el caso de ByC, se utiliza la Encuesta de Innovación y Desarrollo Tecnológico en Bogotá y Cundinamarca (EIByC), realizada en 2005, y cuyos resultados se publicaron en Malaver y Vargas (2006). Los resultados se analizan acudiendo a la estadística descriptiva, complementada con análisis de correspondencias y análisis econométrico.

El texto se desarrolla en cinco secciones. En la primera se presenta el marco conceptual, centrado en la relación entre la innovación y las finanzas; en la segunda se expone la metodología utilizada; en la tercera se esbozan los rasgos más protuberantes de la financiación de la innovación en el caso colombiano; en la cuarta se examina la relación finanzas-innovación, a partir de los datos micro recogidos en la EIByC; por último, se presentan las conclusiones del estudio y se esbozan algunas líneas de profundización, que parten de reconocer que este es un trabajo apenas exploratorio.

ENFOQUE ANALÍTICO

Hoy cabe poca duda en torno a la importancia fundamental de la innovación. Esta, sin embargo, surge de procesos costosos, llenos de incertidumbre y riesgo, asociados a fallas de mercado que inciden en la dinámica y dirección del desarrollo tecnológico. Profundizar en esas fallas es prioritario, sobre todo en el ámbito micro —poco estudiado en Colombia—, con el fin de contribuir a fortalecer las políticas orientadas a potenciar los impactos de la innovación en la economía y el desarrollo social.

INNOVACIÓN Y DESARROLLO

En la teoría económica existe un amplio consenso en torno al papel del progreso técnico como motor del crecimiento, la competitividad y el

desarrollo;² también en torno a que el desarrollo tecnológico endógeno, basado en la innovación, es obstaculizado por fallas de mercado y factores del contexto que inducen la intervención pública para estimular el aumento de la inversión en actividades de innovación en el nivel requerido para lograr crecimientos que generen desarrollo económico.³

En general, las fallas de mercado están asociadas con factores como información imperfecta y problemas de apropiación de los productos de la innovación que inciden en niveles de inversión subóptima y excluibilidad (por problemas de disponibilidad, accesibilidad, costo y capacidad para absorber el conocimiento).⁴ Sin embargo, existen diferencias importantes en lo concerniente a la intervención del Estado para remediar esas fallas, debido a distintas concepciones de los determinantes del cambio técnico y de la innovación.

El análisis se concentra aquí en las fallas, que se traducen en restricciones de carácter financiero que enfrentan las empresas cuando innovan; pero se presentan los matices surgidos de las diferencias relacionadas, por ejemplo, con la introducción de factores asociados con el aprendizaje y los contextos organizacional, cultural e institucional que restringen la financiación de la innovación. De esa manera se busca superar la relación predominantemente cuantitativa y mecánica que existe entre finanzas e innovación (O'Sullivan, 2005).

INNOVACIÓN Y FINANZAS

Las características propias de los proyectos de innovación (incertidumbre, riesgo, intangibilidad y alto costo) hacen que el análisis de los criterios de

² Al respecto puede consultarse Schumpeter (1934), Solow (1956), Swan (1956), Arrow (1962), Ahmad (1966), Elster (1982), Ruttan (1997) y Barro y Sala-i-Martin (1997).

³ Véanse Solow (1956); Swan (1956); Arrow (1962); Winter (1964); Vernon (1966); Dasgupta y Stiglitz (1980); Nelson y Winter (1982); Dosi (1988a, 1988b, 1997); Romer (1986 y 1990); Mankiw, Romer y Weil (1992); Grossman y Helpman (1991); Aghion y Howitt (1992); Bell y Pavitt (1993); Freeman (1994); Nelson (2001), y Cantwell (2005).

⁴ Para mayor detalle, especialmente en lo relacionado con la financiación de las actividades de I+D e innovación se puede ver Becchetti y Sierra (2002).

inversión sea más arduo de lo convencional, pues no existe un “mercado perfecto de tecnologías”. Esto hace necesario analizar los incentivos que mueven a las empresas a embarcarse en proyectos de innovación. En un mundo de competencia imperfecta, la expectativa de obtener significativas rentas monopolísticas futuras constituye la principal motivación empresarial; sin embargo, esas expectativas son morigeradas por la posibilidad de ulteriores innovaciones e imitaciones de otras empresas (debido a los *spillovers* inmanentes a la actividad de investigación) y por la incertidumbre de los resultados de los procesos innovadores. A ello deben sumarse las restricciones internas y externas debidas a la asimetría informativa entre financiadores (internos y externos) y propietarios de los proyectos innovadores.

Estas restricciones están presentes en los sistemas financieros de los países, cuya capacidad para apoyar la inversión en actividades de investigación, desarrollo e innovación (I+D+i) depende de: (i) el monto; (ii) el precio de los fondos provistos; (iii) la capacidad relativa de evaluar a los empresarios y seleccionar buenos proyectos; (iv) la capacidad relativa de hacer seguimiento a las empresas para evitar el desperdicio de los recursos, y (v) la eficiencia relativa para diseminar información sobre la empresa y sus actividades de investigación, a fin de crear los *spillovers* tecnológicos (Allen y Gale, 1995).

La compleja interacción de todos estos factores da lugar a la hipótesis del *pecking order* o jerarquía de preferencias financieras, basada esencialmente en los costos de los fondos y en consideraciones relativas a la propiedad y el control de los proyectos. Según esta, las empresas que innovan prefieren financiar los proyectos, en primer lugar, con fondos propios; en segundo lugar, con fondos de deuda bancaria, y, por último, con capital nuevo (participación de nuevos accionistas). Esta jerarquía de las fuentes de fondos estaría reforzada por la incertidumbre sobre los flujos de caja derivados de las actividades de I+D+i, la acentuación de la asimetría informativa, los mayores costos de quiebra y liquidación (debidos a la alta especificidad de los activos) y los efectos de los riesgos de fuga de información sobre la propiedad y el control (Becchetti y Sierra, 2002).

En ese contexto se sitúa el tema de las fuentes y los mecanismos de financiación de los proyectos de ciencia, tecnología e innovación (CTI) pues, debido a las complejidades asociadas con este tipo de proyectos, las empresas no pueden tomar decisiones expeditas y sobre bases seguras en relación con los procesos de innovación. De allí se desprende la necesidad de la intervención del Estado y de unas políticas de CTI que estimulen las inversiones del sistema empresarial (O'Sullivan, 2005; Léger y Swaminathan, 2007).⁵ Dicha intervención se hace a través de los incentivos para elevar la inversión privada que proviene de los fondos propios de los empresarios, los intermediarios financieros, el mercado de capitales, o de la financiación pública efectuada mediante transferencias directas e indirectas.

En el escenario bosquejado es relevante profundizar en el análisis micro de las finanzas de la innovación basándose en la asimetría informativa y la incertidumbre paramétrica, y dar cabida al estudio de la incertidumbre fundamental que caracteriza la innovación, pues:

... si la incertidumbre es fundamental, como ocurre en la innovación, los agentes económicos no sólo no pueden tener certeza siquiera sobre el posible estado del mundo que se obtendrá al final, sino que tampoco la pueden tener sobre los estados del mundo posibles. Por ende, al invertir en innovación, no existen pautas objetivas que permitan tomar decisiones o resolver desacuerdos. (O'Sullivan, 2005: 257-258)

Aquí, tanto los juicios subjetivos basados en las percepciones como los sistemas de creencias tienen un papel central y se resalta, por ende, el carácter experiencial e interpretativo de la toma de decisiones en materia de asignación de recursos a una actividad esencialmente incierta como la

⁵ La relación entre los realizadores de los proyectos de CTI y los potenciales financiadores está caracterizada por: (a) el alto peso de la innovación subsecuente en el comportamiento de las empresas con CTI; (b) el bajo *matching* entre financiadores y realizadores de proyectos de CTI, debido a problemas de asimetría informativa, y (c) el impacto sobre el potencial de éxito de estos proyectos de las complejidades creadas por la interacción de distintos componentes y actores (Becchetti y Sierra, 2002).

innovación. Esta es, precisamente, una de las mayores limitaciones del análisis microfinanciero de la innovación, y nace de una concepción limitada y estática de la racionalidad, que excluye las dimensiones subjetiva y experiencial de la toma de decisiones (O'Sullivan, 2005).

La perspectiva aquí asumida es oportuna, habida cuenta de la poca literatura existente sobre las relaciones entre la innovación y el acceso a recursos de financiación por parte de las empresas en países en desarrollo. En efecto, la literatura teórica y empírica disponible se centra en los países más avanzados (Becchetti y Sierra, 2002; Ayyagari, Demirguc-Kunt y Maksimovic, 2007; Mohnen *et al.*, 2008), mientras que los pocos estudios sobre los países en desarrollo se basan en datos genéricos que dificultan profundizar en los análisis (Ayyagari, Demirguc-Kunt y Maksimovic, 2007; Léger y Swaminathan, 2007). Los incipientes hallazgos disponibles indican que el papel de los mecanismos de asignación de recursos financieros a la innovación incide directamente en el crecimiento económico de los países en desarrollo (Ayyagari, Demirguc-Kunt y Maksimovic, 2007).

METODOLOGÍA

Este trabajo tiene un carácter exploratorio tanto en los objetivos que persigue (en cuanto a la relación finanzas-innovación) como en las técnicas de análisis utilizadas. La información proviene de los resultados arrojados por la EIByC, encuesta que se aplicó a 400 empresas manufactureras de ByC por la Cámara de Comercio de Bogotá y el Observatorio Colombiano de Ciencia y Tecnología (OCyT), en el 2005.⁶

⁶ El universo del estudio está conformado por 2.404 empresas manufactureras de ByC con más de 10 empleados en el 2002 (según los registros mercantiles de la Cámara de Comercio de Bogotá y de los directorios industriales del Departamento Administrativo Nacional de Estadística), a partir del cual se realizó un muestreo aleatorio estratificado. El tamaño de muestra fue de 400 empresas, que es representativa para el Código Industrial Internacional Uniforme (CIIU), revisión dos, a tres dígitos, con un error estándar relativo de 3,73%. La estimación de los indicadores para el universo estudiado se garantizó con el cálculo de factores de expansión para cada uno de los grupos establecidos del CIIU.

En concordancia con el carácter exploratorio del trabajo, cada una de las secciones del texto que presentan los resultados obtenidos inicia con un análisis descriptivo, orientado a identificar las características más relevantes de la relación finanzas-innovación en ByC. Los resultados de este análisis se complementan y contrastan con:

- El uso de una de las técnicas propias de la minería de datos, el análisis de correspondencias, para facilitar la visualización de las asociaciones entre las variables.
- Análisis econométricos basados en modelos *logit* y *probit* que, dada la naturaleza discreta de la variable dependiente, permiten establecer valores confiables de la probabilidad de que una u otra variable utilizada como explicativa influya en el resultado binario propuesto. De este modo, se establecen relaciones entre un factor y un resultado con fuerza de significatividad estadística en los intervalos de confianza asumidos como válidos para hacer afirmaciones. Una particularidad en el uso práctico de estos modelos es que al hacer la estimación se obtiene un resultado indicativo, en términos de su signo y significatividad estadística, pero que no indica nada en términos de la magnitud del coeficiente. Para poder obtener el valor de la probabilidad asociada a un factor o variable explicativa es necesario estimar, en segunda instancia, un modelo de “efectos marginales” que arroja resultados—donde la magnitud se interpreta en términos probabilísticos— equiparables a los de los coeficientes de los modelos de regresión lineal.

CONTEXTO DE LA FINANCIACIÓN Y LA INNOVACIÓN EN COLOMBIA

Para cumplir los objetivos del trabajo, centrados en la relación finanzas-innovación, es preciso hacer una breve caracterización del contexto donde dicha relación se inscribe, esto es, el sistema financiero y las políticas de estímulo a la inversión en actividades de innovación en el país. Tal con-

texto contribuye a darles sentido a los avances en los esfuerzos y logros obtenidos en materia de innovación.

DESARROLLO DEL SISTEMA FINANCIERO COLOMBIANO Y DE LA INVERSIÓN EN INNOVACIÓN

El desarrollo del sistema financiero colombiano no corresponde a las necesidades del sector empresarial del país (Sarmiento y Salazar, 2005; Tenjo, López y Zamudio, 2006). El mercado accionario es muy pequeño (0,4% del PIB), tiene pocas empresas registradas (96 contra 366 de Brasil, 331 de México, 246 de Chile y 216 de Perú) y su grado de concentración (65%) es alto (Consejo Privado de Competitividad [CPC], 2007).

El sistema bancario está rezagado en el contexto latinoamericano: la proporción de crédito bancario al sector privado respecto al PIB es de apenas el 13,6%, al tiempo que el grado de ineficiencia reflejado en los costos como porcentaje de los activos totales alcanza el 5,1% (en Corea e India son inferiores al 2%). La cobertura, en términos de empresas con cuenta de ahorro, cuenta corriente y acceso al microcrédito, es baja (Asobancaria, 2009; Avendaño, 2006).⁷ Sin embargo, el crédito bancario es la principal fuente de financiación empresarial: los créditos comerciales representan el 17% del PIB, mientras los bonos corporativos no financieros de largo plazo apenas llegan al 3,8%, por debajo de Chile, Argentina y Brasil (CPC, 2007).

Las alternativas de inversión enfrentan dos fenómenos limitantes del acceso y el uso efectivo de las opciones existentes por parte de los potenciales inversionistas: (i) la baja capacidad de ahorro, que constriñe la

⁷ Claessens y Perotti proponen un marco de referencia para interpretar la variada literatura que explora los nexos entre desarrollo financiero y desigualdad y hace interesantes anotaciones sobre el papel de la regulación y la política financiera en contextos de subdesarrollo y asimetría de poder político. Los autores afirman que hay evidencia suficiente que demuestra que "... tanto la inequidad económica como el subdesarrollo (financiero) están determinados conjuntamente por factores institucionales que provocan un acceso desigual a los derechos políticos y contractuales [...]. La participación política sesgada permite que los intereses establecidos protejan sus rentas mediante la limitación del acceso financiero y, por ende, la supresión de la competencia y la entrada." (2007: 756).

disponibilidad de recursos de inversión de las familias colombianas, y (ii) la baja “cultura financiera y de inversión” promedio de los colombianos (Sierra y Londoño, 2008). Estos factores interactúan con otros que afectan la oferta de proyectos de inversión:

- El bajo número de emisores de valores negociables en el mercado público de capitales (MPC).
- El poco interés de las empresas por registrarse en el MPC (por exigencias de información, costo-beneficio) (Uribe, 2007).
- El bajo número de fondos de inversión.
- La virtual ausencia de capital de riesgo/capital semilla en Colombia (para fondar proyectos nuevos e innovadores).

De ese modo, el escaso desarrollo del sistema financiero, expresado en problemas de cobertura, ineficiencia y concentración, así como por las dificultades de índole cultural de los individuos y las empresas, se traduce en escasez, elevados costos y dificultad de acceso a recursos para financiar las actividades de inversión de las empresas. Así, a las fallas de mercado típicas se les adicionan los obstáculos surgidos del bajo desarrollo del sector y los problemas institucionales que, tal como menciona la literatura, se acrecientan en el caso de los proyectos relacionados con la innovación. Esto eleva los desafíos para la política pública.

POLÍTICA PÚBLICA DE ESTÍMULO A LA INVERSIÓN EN ACTIVIDADES DE I+D

En Colombia, desde el momento mismo de la “creación” del Sistema Nacional de Innovación (SNI), se planteó como un objetivo específico: “incrementar la inversión y la actividad privada en programas de investigación y desarrollo tecnológico, por medio de incentivos y mecanismos de tipo financiero” (Consejo Nacional de Ciencia y Tecnología [CONCyT], 1995a, s. p.).⁸

⁸ Los objetivos específicos del sistema de financiamiento contemplado por el SNI son: (a) incrementar sustancialmente los recursos públicos y privados para la innovación y el desarrollo tecnológico; (b) ampliar la cobertura y democratizar el acceso a las fuentes de financiamiento;

En concordancia con este objetivo, en la política de CTI para el período 1994-1998 se propuso una estrategia de modernización del sistema de financiación de la innovación y el desarrollo tecnológico basada en la creación de cinco modalidades de financiamiento para atender los requerimientos de los proyectos de innovación empresarial: (i) cofinanciación, (ii) riesgo compartido, (iii) financiamiento directo, (iv) capital semilla y (v) garantías para la innovación (CONCyT, 1995b); además, se puso como meta alcanzar en 1998 una inversión del 1% del PIB para las actividades de CTI.

Más de una década después, en la Política Nacional de Fomento a la Investigación y la Innovación (Colciencias, 2008), los instrumentos de fomento se agruparon en cuatro categorías:

- Los dirigidos a apoyar el financiamiento de proyectos de I+D+i (recuperación contingente, cofinanciación, crédito, riesgo tecnológico compartido y garantías).
- Los incentivos tributarios para la investigación y la innovación (deducciones y exenciones).
- Aquellos para financiar la formación de recursos humanos y otros apoyos (por ejemplo, vinculación de investigadores a empresas o financiación de registros de propiedad intelectual).
- Acciones para mejorar su efectividad: incrementar su divulgación, agilizar el trámite asociado a los incentivos fiscales, mejorar la atractividad y el acceso al incentivo relacionado con la línea de crédito, redefinir las condiciones de condonación del instrumento de riesgo compartido, potenciar la cofinanciación de los proyectos de I+D+i, promover los fondos de capital de riesgo y proponer nuevos instrumentos de fomento.

Como se verá, en este frente los avances han sido insuficientes.

(c) mejorar la eficiencia y disminuir los costos para el usuario final en el manejo de los recursos; (d) ofrecer capital de riesgo para inversiones en empresas de base tecnológica; (e) establecer mecanismos de garantías tecnológicas para respaldar proyectos de innovación de las pequeñas y medianas empresas, (f) aprovechar los beneficios de los incentivos fiscales para incrementar las inversiones en innovación.

FINANCIACIÓN DE LA INNOVACIÓN EN COLOMBIA

Pese a las presiones competitivas y a las políticas de estímulo, son muy pocos los esfuerzos de inversión realizados en Colombia. Según las cifras del OCyT (2008), la inversión en actividades de I+D apenas alcanza el 0,16% del PIB registrado en el 2007. Una comparación internacional muestra que los esfuerzos realizados son inferiores al promedio de América Latina (Cuadro 1). La desproporción es amplia cuando se observa que, por ejemplo, el PIB de Estados Unidos es cinco veces el de América Latina y los recursos invertidos en I+D son 27 veces más altos. Frente a Colombia esa diferencia es mayor.

Cuadro 1

Esfuerzos tecnológicos relativos en Latinoamérica y Colombia
Una visión comparada (2007)

Países	Inversión (%)	Colombia frente a cada país (veces)	
	I+D/PIB	PIB	I+D
Estados Unidos	2,60	101,1	1.462,0
Canadá	1,98	9,2	101,5
América Latina y el Caribe	0,54	20,9	61,2
Brasil	0,82	7,2	32,9
Colombia	0,16	1,0	1,0

Fuente: elaboración propia a partir de la RICyT (2007) y OCyT (2008).

A lo anterior deben agregarse dos hechos: (i) mientras en países desarrollados como Estados Unidos las empresas contribuyen con el 64,9% de la inversión nacional, en Colombia llega al 42,6% (Red Iberoamericana de Indicadores de Ciencia y Tecnología [RICyT], 2007; OCyT, 2008; CPC, 2007); (ii) en el 2008, la política se vuelve a plantear el desafío de elevar en el 2010 la inversión en CTI al 1% del PIB (Colciencias, 2008).

Sin ahondar en el balance de los resultados de más de una década de políticas explícitas de fomento a la inversión en CTI, que desborda los objetivos de este trabajo, una inversión en I+D que no supera el 0,16% del PIB y el hecho de que después de una década de la primera formula-

ción (1995) en 2008 se vuelva a plantear como meta una inversión del 1% para el 2010, cuando la dinámica del cambio técnico y el escenario competitivo han elevado sustancialmente las necesidades de inversión en CTI, indican que los esfuerzos de la política no han sido eficaces en el propósito de elevar la inversión pública y privada en estas actividades. En este frente el desafío para la política es enorme.

En ese contexto, es apenas lógico que los avances registrados entre 1995 y 2005 por parte de las encuestas de innovación sean inferiores a los retos competitivos del país. De hecho, como lo muestra el Cuadro 2, entre el 42% y el 51% de las empresas industriales del país y de ByC no innovan y apenas entre el 5,8% y el 9,1% generan innovaciones que otorgan competitividad internacional (Malaver y Vargas, 2005 y 2006; CPC, 2007; Colciencias, 2008).

Cuadro 2

Evolución de la innovación, 1995-2005

Tipo de empresa	Colombia		Bogotá y C/marca	
	Participación (%)		Participación (%)	
	EDT 1	EDIT II	EDT 1	EIByC
EIE	8,7	5,8	5,8	9,1
EIA	40,5	51,5	36,4	39,8
Empresas innovadoras	49,2	57,4	42,3	48,9
EPI	28,1	22,1	23,2	20,2
ENI	22,7	20,5	34,6	30,8
Empresas no innovadoras	50,8	42,6	57,7	51,1
Total	100,0	100,0	100,0	100,0

EIE: empresas innovadoras en sentido estricto; EIA: empresas innovadoras en sentido amplio; EPI: empresas potencialmente innovadoras; ENI: empresas no innovadoras.

Fuente: elaboración propia a partir de EDT 1 (1996), EDIT II (2005) y EIByC (2005).

En concordancia con los objetivos del artículo, en lo que sigue se ahondará en la relación entre las finanzas y la innovación desde una perspectiva micro, con el fin de contribuir al entendimiento de las restricciones de índole financiera para la innovación.

RELACIÓN FINANZAS-INNOVACIÓN EN BYC: UNA CARACTERIZACIÓN A PARTIR DE LA EIBYC

Con base en el marco planteado en la “Metodología”, enseguida se indaga por la existencia de restricciones financieras y el *pecking order* en la industria de ByC. Los resultados arrojados por la EIByC se analizan primero de manera descriptiva, luego mediante técnicas de visualización (por análisis de correspondencias) y, finalmente, se contrastan mediante ejercicios econométricos.

RESTRICCIÓN FINANCIERA: UNA MIRADA GLOBAL

En las empresas industriales de ByC, tal como lo muestra el Cuadro 3, se constata que en la región existen restricciones financieras para las empresas. En otras palabras, se registra el fenómeno que fue planteado por Weigand (1999) y verificado empíricamente en distintos países (Stiglitz y Weiss, 1981; Myers y Majluf, 1984; Stockdale, 2002; Canepa y Stone-man, 2002). El análisis descriptivo muestra que, en efecto, después de las barreras originadas en las condiciones del mercado —señaladas por cerca del 40% de las empresas que no innovaron— la carencia de recursos financieros constituye la segunda gran restricción y afecta al 27,6% de ellas. Estos resultados, por lo demás, concuerdan con los arrojados por las encuestas de innovación realizadas en el Reino Unido (Stockdale, 2002).

La carencia de recursos financieros es una importante restricción para el conjunto de las empresas que no innovaron en el período 2003-2005, pero hay diferencias notables entre ellas. Las restricciones de carácter financiero, junto con las del mercado, tienen mayor importancia entre las empresas que realizaron actividades de innovación, pero no obtuvieron innovaciones en el período, es decir, las empresas potencialmente innovadoras (EPI). En el caso de las empresas que no efectuaron actividades de innovación (ENI), ambas restricciones son menores frente a las de índole cultural, como la consideración de que innovar no es rentable o es innecesario (Malaver y Vargas, 2006; Sierra *et al.*, 2009).

Cuadro 3

Razón principal por la que no se innovó*

Razones	Empresas (%)
Innovar no es rentable	6,2
No considera necesario innovar	13,1
Carencia de recursos financieros para innovar	27,6
Carencia de personal capacitado para innovar	6,0
El mercado no requiere o valora los nuevos productos	31,8
Tamaño inadecuado del mercado	8,2
Otros	7,1
Total empresas (%)	100,0
Total empresas que no obtuvieron desarrollos novedosos y contestaron la pregunta (N°)	733,0

* Se calcula sobre las empresas que no obtuvieron desarrollos novedosos y contestaron esta pregunta.

Fuente: EIByC (2005).

Esto último se corrobora con el análisis de correspondencias, en el que se evalúan las correlaciones entre las variables señaladas en el Cuadro 3, el tamaño de las empresas⁹ y el subsector industrial al que pertenecen.¹⁰ En este análisis, en el que sólo se incluyen las empresas no innovadoras,¹¹ el Gráfico 1 confirma que en el caso de las empresas más pequeñas, que no innovan, las principales razones aducidas para no hacerlo obedecen a que consideran que la innovación no es rentable o es innecesaria. También, aunque con menos fuerza, aducen que no lo hacen por falta de recursos

⁹ A efectos del análisis, las empresas se dividieron entre: las pequeñas de menor tamaño (hasta 20 personas ocupadas), las pequeñas de mayor tamaño (hasta 50), las medianas (hasta 200) y las grandes (más de 200).

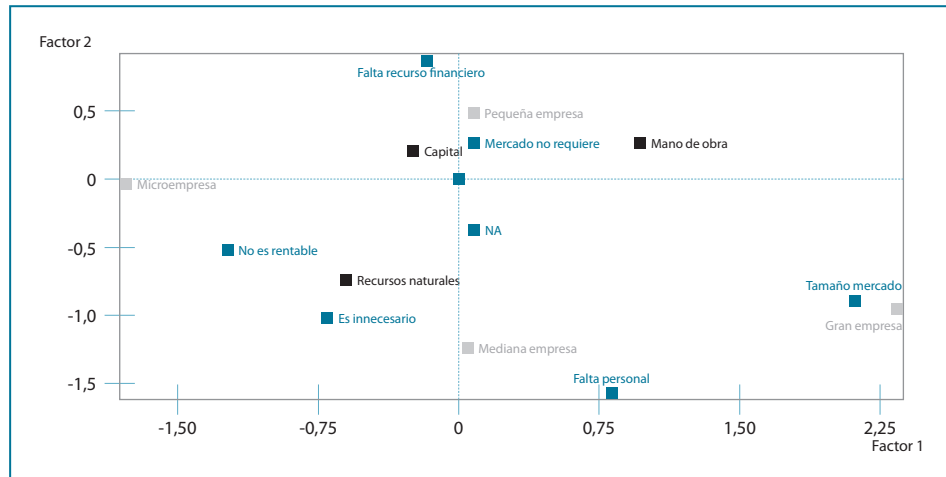
¹⁰ Las empresas, de acuerdo con la intensidad en el uso de los factores productivos, se agruparon en tres categorías: intensivas en capital, intensivas en recursos naturales e intensivas en mano de obra.

¹¹ Debe advertirse que en razón de que el histograma de valores propios obtenido no mostró saltos, la correlación entre las variables analizadas es baja; entonces, en el Gráfico 1 sólo se considera el primer plano factorial.

financieros. En abierto contraste, las grandes empresas tienen como principal razón asociada el tamaño inadecuado del mercado.

Gráfico 1

La relación entre el tamaño, el sector y las razones para no innovar en las empresas no innovadoras de ByC



Fuente: elaboración propia a partir de la EIByC (2005).

El contraste mediante el análisis econométrico para establecer la relación entre el carácter no innovador y el tamaño¹² muestra que sólo dos razones son significativas estadísticamente y presentan probabilidades de incidencia altas: la carencia de personal y la carencia de recursos financieros (Cuadro 4). Así mismo, la variable de control (tamaño) pone en evidencia que las empresas más pequeñas son las menos innovadoras.

¹² No se incluyó la variable *sector*, porque el limitado número de observaciones resultante de subdividir la muestra según este criterio obstaculizó, en un primer intento, el procedimiento al punto de que el modelo no corrió.

Cuadro 4

Razones por las que no innovó

Variable dependiente: no innovó	Modelo <i>probit</i> coeficiente (Error est.)	Efectos marginales dy/dx	Modelo <i>probit</i> coeficiente (Error est.)	Efectos marginales dy/dx
Carencia de personal	2,33*** (0,26)	0,64*** (0,02)	2,27*** (0,26)	0,63*** (0,02)
Carencia de recursos financieros para innovar	0,58** (0,32)	0,22*** (0,12)	0,65** (0,33)	0,25** (0,12)
Tamaño			-0,12*** (0,02)	-0,04*** (0,00)
Constante	-0,47*** (0,03)		-0,12 (0,07)	
Número de observaciones	1.903	1.903	1.903	...
LR $\chi^2(2)$	346,45		371,63	
Prob > χ^2	0,00		0,00	
Log likelihood	-1.090,91		-1.078,32	

* 90% de significancia; ** 95% de significancia; *** 99% de significancia.

Fuente: elaboración propia con base en EIByC (2005).

La restricción financiera obstaculiza el surgimiento de innovaciones; también frena a las empresas que realizan esfuerzos para innovar (EPI) e incluso a las que han obtenido innovaciones, bien sean empresas innovadoras en sentido amplio (EIA) o en sentido estricto (EIE). La carencia de recursos financieros y las dificultades para acceder a ellos constituyen la segunda limitación a las actividades de innovación y afectan al 29% de las empresas (Cuadro 5), casi al mismo nivel de las barreras originadas por la escasez de personal idóneo que afectan al 31%. Estos resultados son similares a los arrojados por las encuestas de innovación en Alemania (Canepa y Stoneman, 2002).

Cuadro 5

Principal obstáculo para innovar*

Obstáculos	Empresas (%)
Limitadas infraestructuras de soporte	8,5
Carencia de personal capacitado	14,0
Dificultades para capacitar el personal	7,6
Dificultades para conseguir personas idóneas	9,4
Carencia de recursos financieros	22,1
Dificultades para acceder a recursos financieros	6,9
Dificultades para acceder a información tecnológica	14,9
Otros	16,6
Total empresas (%)	100,0
Total empresas que tuvieron obstáculos (Nº)	1.337,0

* Se calcula sobre las empresas que tuvieron obstáculos a la hora de obtener desarrollos novedosos.

Fuente: EIByC (2005).

En este caso, también existen diferencias entre los distintos tipos de empresas. Para las empresas que obtuvieron las innovaciones con más alto grado de novedad (las EIE), la restricción financiera constituye el mayor obstáculo para innovar, pero es superada por las dificultades asociadas al personal en el caso de las empresas menos innovadoras, especialmente las que aún no han obtenido innovaciones (las EPI) (Sierra *et al.*, 2008).

De ese modo, mientras la falta de los recursos humanos requeridos para innovar afecta a las empresas que tienen las menores capacidades tecnológicas, en el caso de las más innovadoras —y que, por lo tanto, acreditan capacidades endógenas para innovar— la restricción financiera es el factor que más inhibe la innovación. En otros términos, la restricción financiera crece a medida que aumenta la capacidad de innovación de las empresas.

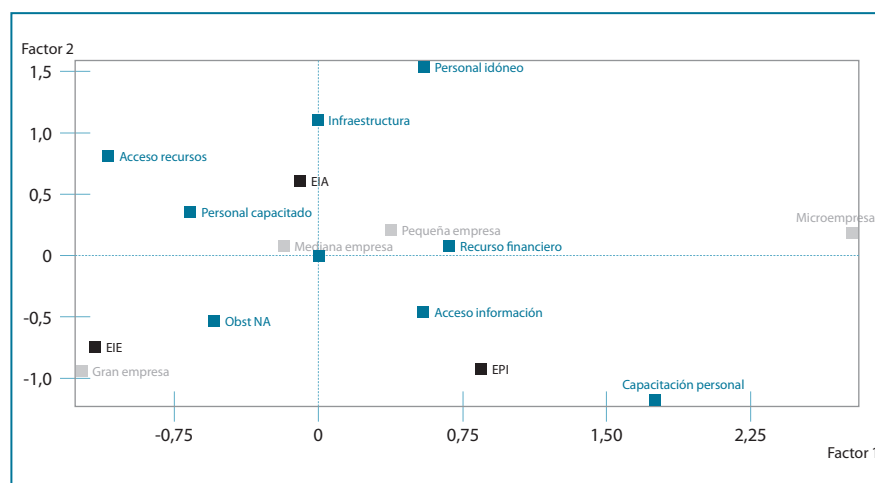
El análisis de correspondencias realizado para visualizar la correlación existente entre el grado de innovación de las empresas,¹³ el tamaño y los

¹³ En este caso, las empresas se agrupan entre aquellas que desarrollaron actividades de innovación en el período de observación, pero no obtuvieron innovaciones, esto es, las EPI; y las

obstáculos para innovar reportados por las empresas muestra que las variables incluidas en el análisis no están muy correlacionadas. Por ello, el análisis sólo se efectúa sobre el primer plano factorial obtenido, que se puede interpretar como asociado al tamaño, puesto que en este se localizan, de izquierda a derecha, las empresas grandes, las medianas, las pequeñas medianas y las más pequeñas (Gráfico 2). El gráfico muestra con nitidez una alta correlación entre el tamaño de las empresas y grado de innovación: las grandes son las más innovadoras (EIE), las medianas tienden a serlo en sentido amplio (EIA), las pequeñas medianas¹⁴ son las más potencialmente innovadoras (EPI) y las más pequeñas son las menos innovadoras.

Gráfico 2

Relación entre grado de innovación, tamaño y obstáculos para innovar en las empresas que adelantaron actividades de innovación en ByC



Fuente: elaboración propia a partir de la EIByC (2005).

que obtuvieron innovaciones para el mercado internacional, como producto de actividades de I+D, es decir, las EIE, y las otras empresas innovadoras, esto es, las EIA. No se incluyen las empresas no innovadoras (ENI), pues se analizan los obstáculos que enfrentan las empresas cuando innovan.

¹⁴ A las empresas pequeñas de mayor tamaño, según el número de empleados, por simplicidad se les denomina pequeñas medianas.

El Gráfico 2 también ilustra que las grandes empresas, que a la vez son las más innovadoras (EIE), presentan como principales obstáculos para innovar el acceso a los recursos financieros y a recursos humanos capacitados. Las empresas medianas señalan como principales obstáculos para innovar las dificultades para acceder a infraestructura y para conseguir personas idóneas. Las empresas pequeñas medianas, que son las más potencialmente innovadoras (EPI), señalan como principales obstáculos el acceso a información tecnológica, la carencia de recursos financieros, las dificultades para capacitar al personal y, en alguna medida, la consecución de personas idóneas.

El contraste econométrico de los resultados anteriores, efectuado a través de un modelo *probit*, indica que la restricción financiera (carencia de recursos y dificultad de acceso a ellos) es significativa. Los resultados del ejercicio señalan, además, que el nivel de la restricción financiera (55% de probabilidad de incidencia de esta variable en el caso de las empresas que innovaron) supera los problemas relacionados con el personal (32%) y de la inadecuada infraestructura de soporte (32%). Esto se confirma, en segunda instancia, controlando por el tamaño de las empresas,¹⁵ lo que pone en evidencia que las pequeñas empresas enfrentan barreras significativas para realizar proyectos innovadores (Cuadro 6).

Cuadro 6
Obstáculos para la innovación

Variable dependiente: innovó	Modelo <i>probit</i> coeficiente (error est.)	Efectos marginales dy/dx	Modelo <i>probit</i> coeficiente (error est.)	Efectos marginales dy/dx
Carencia de personal capacitado	0,32*** (0,08)	0,12*** (0,03)	0,20** (0,08)	0,08** (0,03)
Dificultades para capacitar personal	0,03 (0,09)	0,01 (0,03)	-0,03 (0,10)	-0,01 (0,03)

Continúa

¹⁵ Es decir, incluyendo una variable adicional en el modelo (tamaño) para verificar si sus resultados son estables y no se modifican (signo, significatividad y, eventualmente, magnitud de los coeficientes).

Variable dependiente: innovó	Modelo <i>probit</i> coeficiente (error est.)	Efectos marginales dy/dx	Modelo <i>probit</i> coeficiente (error est.)	Efectos marginales dy/dx
Dificultades para conseguir personas idóneas	0,55*** (0,09)	0,21*** (0,03)	0,63*** (0,09)	0,24*** (0,03)
Carencia de recursos financieros	0,65*** (0,07)	0,25*** (0,02)	0,73*** (0,07)	0,28*** (0,02)
Dificultad para acceder a recursos financieros	0,75*** (0,09)	0,28*** (0,03)	0,72*** (0,09)	0,27*** (0,03)
Dificultad para acceder a información	0,16** (0,07)	0,06** (0,03)	0,19** (0,08)	0,07** (0,03)
Limitada infraestructura y soporte	0,85*** (0,08)	0,31*** (0,02)	0,86*** (0,08)	0,32*** (0,02)
Tamaño			0,22*** (0,02)	0,09*** (0,00)
Constante	-0,49*** (0,02)		-1,11*** (0,07)	
Número de observaciones	2.405	2.405	2.405	2.405
LR $\chi^2(2)$	482,61		584,14	
Prob > χ^2	0,00		0,00	
Log likelihood	-1.425,42		-1.374,66	

* 90% de significancia, ** 95% de significancia, *** 99% de significancia.

Fuente: elaboración propia con base en ElByC (2005).

FUENTES DE FINANCIACIÓN DE LA INNOVACIÓN EMPRESARIAL Y *PECKING ORDER*

Para complementar la indagación sobre las restricciones financieras existentes se identificaron las fuentes de financiación de la innovación de las empresas industriales de la región ByC, con el propósito de poner a prueba la hipótesis del *pecking order*. Al respecto, el Cuadro 7 revela que en la estructura de financiación de las empresas manufactureras de la región predomina la financiación de las actividades de innovación con recursos internos. Revela, además, que este comportamiento es más acentuado en las empresas más grandes y en las de menor tamaño. Esto sugiere, por lo menos para el caso de las grandes, una clara preferencia por la financiación

interna frente a la externa, en la medida en que estas empresas disponen de mayor capacidad de generación de recursos para financiar sus innovaciones (Cohen y Levin, 1989).¹⁶

Cuadro 7

Fuentes de financiamiento según tamaño: porcentaje de montos

Financiamiento	Tamaño					Total
	1	2	3	4	5	
Externo	25,5	48,8	58,6	45,5	29,1	39,6
Interno	74,5	51,2	41,4	54,5	70,9	60,4
Total	100,0	100,0	100,0	100,0	100,0	100,0

1: empresas con menos de 20 empleados; 2: entre 20 y 49; 3: entre 50 y 99; 4: entre 100 y 199; 5: con 200 o más.

Fuente: EByC (2005).

En términos de las capacidades, las empresas más innovadoras (EIE) financian sus actividades de innovación con recursos internos en un porcentaje mucho mayor que las empresas menos innovadoras; de hecho, sobrepasa el 80% de los recursos totales invertidos¹⁷ (Sierra *et al.*, 2008). También es destacable el hecho de que las EPI y las EIA financien sus innovaciones con una proporción mayor de recursos externos que internos.

Al desglosar tanto las fuentes internas como las externas a las que acuden las empresas, varios fenómenos se destacan. Las empresas más grandes acuden en mayor grado a sus recursos internos y a otras empresas del grupo al que pertenecen, es decir, para ellas aplica el *pecking order*. También acuden en mayor porcentaje a la banca comercial nacional e internacional e incluso a la casa matriz (Cuadro 8).¹⁸ Esto indica que las empresas grandes

¹⁶ Otros factores, asociados a las características de las agrupaciones industriales, los patrones de desarrollo tecnológico a los que pertenecen, los índices de concentración industrial, etc., pueden incidir de manera diferenciada en los niveles de innovación, pero escapan a los objetivos y alcances de este trabajo

¹⁷ Esto, a título de hipótesis por verificar en trabajos posteriores, podría estar asociado con una abundante liquidez interna y con el menor costo de esos fondos respecto a los de deuda, en el caso de las más empresas más innovadoras.

¹⁸ También a manera de hipótesis, esto podría relacionarse con el carácter de los proyectos. Entre

tienen acceso a una mayor diversidad de fuentes de financiación y acuden a fuentes institucionales como Bancoldex y Colciencias.

Cuadro 8

Fuentes de financiamiento según tamaño: porcentaje de empresas

Fuentes	Tamaño					Total
	1	2	3	4	5	
Fuentes externas públicas						
Bancoldex	4,4	6,2	6,5	7,3	4,0	5,8
PNPC	0,0	0,0	0,0	0,0	0,0	0,0
Fomipyme	0,0	0,0	0,0	2,1	0,0	0,3
Proexport	0,0	0,9	2,7	2,1	0,0	1,2
Expopyme	0,0	0,0	0,0	0,0	0,0	0,0
Sena	0,0	1,4	0,0	6,8	0,0	1,5
Colciencias	0,0	1,9	0,0	1,7	1,0	1,1
Fuentes externas privadas						
Banca comercial nacional	11,9	19,4	21,9	17,2	29,5	19,4
Fuentes externas internacionales						
Banca comercial internacional	1,0	0,7	0,0	2,1	5,7	1,4
Casa matriz	0,9	0,0	1,3	0,0	1,6	0,6
Cooperación internacional	0,0	0,0	0,0	0,0	0,0	0,0
Organismos internacionales	0,0	0,0	0,0	0,0	0,0	0,0
Fuentes internas						
Recursos propios	39,9	50,7	46,4	51,3	54,5	48,4
Otras empresas del grupo	0,0	0,0	1,7	0,0	6,3	1,0
Total	100,0	100,0	100,0	100,0	100,0	100,0

PNPC: Programa Nacional de Productividad y Competitividad.

1: empresas con menos de 20 empleados; 2: entre 20 y 49; 3: entre 50 y 99; 4: entre 100 y 199; 5: con 200 o más.

Fuente: ElBYC (2005).

más complejo el proyecto, su evaluación por parte de los prestatarios (bancos u otras entidades) es más difícil y genera mayores costos que se cargan al proyecto y, de paso, este se expone al riesgo de selección adversa.

Entre las pymes, por el contrario, las empresas de tamaño medio son las que acuden en mayor grado a y usan un mayor número de instrumentos financieros institucionales. Las empresas más pequeñas acuden en menor porcentaje tanto a las fuentes internas como a las externas privadas, sobre todo a las institucionales. Se destaca, igualmente, en ese nivel la existencia del *pecking order*.

Por lo demás, en todos los casos, el mecanismo público de apoyo financiero más utilizado es Bancoldex. También sobresale el pequeño porcentaje de empresas que acude a Proexport y Fomipyme, el hecho de que ninguna empresa haya acudido al Programa Nacional de Productividad y Competitividad (PNPC) y a Expopyme, y que esta situación se presente entre las pymes, que son las destinatarias de estos mecanismos de apoyo para la financiación.

Las fuentes de financiación usadas por los distintos grupos de empresas, de acuerdo con los diferentes grados de innovación, muestran con nitidez que las EIE se comportan de manera bastante similar a las empresas de mayor tamaño. En efecto, es muy superior el porcentaje de empresas que utiliza sus recursos y los de otras empresas del grupo para financiar su innovación (Cuadro 9). Un mayor porcentaje de ellas usa las fuentes externas privadas nacionales e internacionales y, desde la perspectiva institucional, son las que acuden en mayor proporción a Colciencias, con lo cual se confirma que un mayor porcentaje de estas empresas (EIE) realiza actividades de innovación que cumplen con los criterios técnicos, financieros y de relaciones con actores del SNI para obtener este apoyo institucional. Sin embargo, estas empresas acuden exclusivamente a Colciencias y a Bancoldex.

En el caso de las EIA y de las EPI, si bien tienen en sus recursos internos la mayor fuente de financiación, lo más destacable es que acuden a una mayor diversidad de fuentes institucionales (sobre todo las EPI), aunque sólo un porcentaje muy pequeño de ellas lo hace, excepto en lo que se refiere a Bancoldex que es, de lejos, la institución oficial de financiación externa más utilizada.

Cuadro 9

Fuentes de financiamiento por grado de innovación: porcentaje de empresas

Fuentes	Grado de innovación			Total
	EIE	EIA	EPI	
Fuentes externas públicas				
Bancoldex	8,4	7,6	10,1	5,8
PNPC	0,0	0,0	0,0	0,0
Fomipyme	0,0	0,0	1,5	0,3
Proexport	0,0	2,2	1,5	1,2
Expopyme	0,0	0,0	0,0	0,0
Sena	0,0	2,2	1,5	1,5
Colciencias	5,1	0,6	1,7	1,1
Fuentes externas privadas				
Banca comercial nacional	30,1	26,6	27,5	19,4
Fuentes externas internacionales				
Banca comercial internacional	2,9	0,9	3,6	1,4
Casa matriz	2,8	0,9	0,0	0,6
Cooperación internacional	0,0	0,0	0,0	0,0
Organismos internacionales	0,0	0,0	0,0	0,0
Fuentes internas				
Recursos propios	86,8	61,5	62,0	48,4
Otras empresas del grupo	11,5	0,0	0,0	1,0
Total	100,0	100,0	100,0	100,0

EIE: empresas innovadoras en sentido estricto; EIA: innovadoras en sentido amplio; EPI: potencialmente innovadoras.

PNPC: Programa Nacional de Productividad y Competitividad.

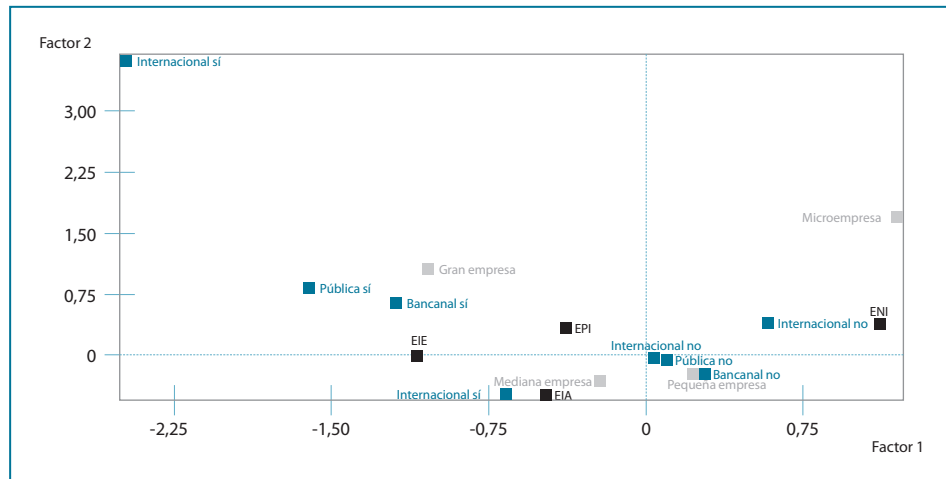
Fuente: ElByC (2005).

Ahora bien, el análisis de correspondencia entre las variables grado de innovación, tamaño y fuentes de financiamiento sugiere una fuerte correlación en ellas. Esta se visualiza en el Gráfico 3, que presenta el primer plano factorial. En este se observa que las grandes empresas y las IE, tienen la mayor correlación con las fuentes de financiamiento públicas, la banca comercial del orden nacional y la internacional. La financiación

interna, como es de esperarse, es más empleada por empresas medianas o grandes. En el polo opuesto, las empresas de menor tamaño, que son las menos innovadoras, no usan fuentes de financiación.

Gráfico 3

Relación entre grado de innovación, tamaño y fuentes de financiación de la innovación en las empresas que adelantaron actividades de innovación en ByC



Fuente: elaboración propia a partir de la EIByC (2005).

El análisis econométrico, mediante un modelo *probit*, intenta explicar cuál es la probabilidad de que una empresa innovadora acuda a ciertas fuentes de financiación. El ejercicio muestra que, en materia de utilización e importancia de las distintas fuentes, hay esencialmente preferencia por dos fuentes oficiales (Bancoldex y Proexport) y por la banca privada nacional, pero siempre de manera subsidiaria a los recursos propios.¹⁹

Se confirma así la importancia primaria de los recursos propios en la financiación de proyectos innovadores y, sucesivamente, la de la banca nacional y de los fondos oficiales, además del papel nulo (relación inversa según el signo negativo del coeficiente) de la banca internacional. Estos

¹⁹ También aquí la única variable de control empleada es el tamaño de la empresa (variable de cinco valores categóricos).

resultados se confirman cuando se incluye la variable de control del tamaño de las empresas y se ratifica que las pequeñas empresas tienen una relación sumamente problemática con el tema de financiación de la innovación.

Cuadro 10
Fuentes de financiación

Variable dependiente: innovó	Modelo <i>probit</i> Coeficiente (error est.)	Efectos marginales dy/dx	Modelo <i>probit</i> Coeficiente (error est.)	Efectos marginales dy/dx
Bancoldex	0,28** (0,12)	0,11** (0,04)	0,35*** (0,12)	0,14*** (0,04)
Proexport	0,57** (0,28)	0,22** (0,10)	0,59** (0,29)	0,22** (0,10)
Sena	-0,37* (0,21)	-0,14* (0,07)	-0,45** (0,21)	-0,17** (0,07)
Colciencias	0,11 (0,27)	0,04 (0,11)	0,11 (0,28)	0,04 (0,11)
Banca nacional	0,59*** (0,07)	0,23*** (0,02)	0,55*** (0,07)	0,21*** (0,02)
Banca internacional	-0,85*** (0,25)	-0,29*** (0,03)	-1,14*** (0,27)	-0,36*** (0,05)
Recursos propios	0,88*** (0,05)	0,34*** (0,01)	0,89*** (0,05)	0,34*** (0,02)
Tamaño			0,20*** (0,02)	0,08*** (0,00)
Constante	-0,59*** (0,04)		-1,13*** (0,07)	
Número de observaciones	2.359	2.359	2.359	2.359
LR $\chi^2(2)$	381,85		462,72	
Prob > χ^2	0,00		0,00	
Log likelihood	-1.442,52		-1.402,09	

* 90% de significancia, ** 95% de significancia, *** 99% de significancia.

Fuente: elaboración propia con base en la EIByC (2005).

El análisis de los efectos marginales indica que la fuente de financiación más utilizada es la de recursos propios (34%), seguida por la banca nacional (24%), Proexport (20%) y Bancoldex (14%), estas dos últimas fuentes estatales. Además, se confirma que recurrir a la banca internacional no es una opción (coeficiente negativo y significativo) y que las empresas que innovan no recurren al Sena (Cuadro 10).

También aquí los efectos marginales ratifican la significancia de los fondos financieros propios y de la relación con los bancos comerciales del país, el papel secundario de los fondos oficiales para la promoción de la innovación y la nula relación con los bancos internacionales. Todo esto parece confirmar la conclusión relativa a la existencia del *pecking order* en las empresas de ByC que desean realizar proyectos de innovación.

OPERACIÓN DE LA RESTRICCIÓN FINANCIERA Y POLÍTICA

Los ejercicios efectuados han mostrado, por una parte, que los esfuerzos tecnológicos realizados en Colombia son muy pocos, que esto se traduce en una inversión tan baja en las actividades de I+D, que en esas condiciones cabe esperar una ampliación de la brecha tecnológica con los países desarrollados e, incluso, con los de América Latina. Esto revela tanto un déficit como enormes desafíos para la política. Por otra parte, mostraron que existe la restricción financiera y el *pecking order*, que es necesaria una política pública que contribuya a superar esas fallas del mercado y que esta necesidad se revela mayor ante el mínimo uso de los instrumentos de fomento, sobre todo, por parte de las empresas de menor tamaño.

Con el propósito de brindar elementos útiles para la formulación de la política pública, enseguida se presentan los resultados de la indagación por cómo opera la restricción financiera, en otros términos, por los factores que obstruyen el acceso a la financiación y, cuando esta se ha dado, por los problemas que han encontrado las empresas.

Las empresas aducen que la principal razón por la cual no usaron financiación externa es la existencia de recursos internos suficientes (Cuadro 11), según lo señala cerca del 60% de ellas. En segundo lugar, lejos de la

anterior, aparece el desconocimiento de las fuentes de financiación, señalado por el 10,1% de las empresas. Este problema de información atañe tanto a la incapacidad de las empresas para apalancar sus recursos con los existentes en el entorno financiero e institucional²⁰ como a deficiencias en las políticas de promoción del uso de los instrumentos públicos.²¹ Después aparecen los problemas relacionados con el sistema financiero, como los trámites exigidos, con la poca pertinencia de las líneas de crédito y las garantías exigidas, que constituyen restricciones para el 14% de las empresas.

Cuadro 11

Principal razón por la cual no se usó financiación*

Razones	%
Solicitudes de crédito no aprobadas	3,2
Desconocimiento de las fuentes	10,1
Trámites exigidos	7,2
Líneas de crédito poco pertinentes	4,8
Garantías exigidas	2,0
Experiencias pasadas negativas	1,4
No se innovó	3,2
Recursos internos suficientes	59,9
Otros	8,3
Total empresas (%)	100,0
Total empresas sin financiación externa (Nº)	1.076,0

* Se calcula sobre empresas sin financiación externa.

Fuente: ElByC (2005).

²⁰ Esa incapacidad expresa un débil desarrollo de las capacidades de vinculación de las empresas industriales de la región ByC (Malaver y Vargas, 2006).

²¹ Colciencias (2008) señala, por ejemplo, que el 39% de las empresas manufactureras colombianas no usa los instrumentos financieros de apoyo a la innovación por desconocimiento de ellos.

Los resultados del análisis econométrico señalan como las principales razones para no utilizar financiación externa, en orden de importancia: el desconocimiento de las fuentes, la suficiencia de los recursos internos y los trámites exigidos (Cuadro 12).

Cuadro 12

Razones por las que no se utilizó la financiación

Variable dependiente: innovó	Modelo <i>probit</i> Coeficiente (error est.)	Efectos marginales dy/dx	Modelo <i>probit</i> Coeficiente (error est.)	Efectos marginales dy/dx
Solicitud de crédito no aprobada	0,20 (0,17)	0,08 (0,06)	0,19 (0,17)	0,07 (0,07)
Desconocimiento de las fuentes	0,85*** (0,12)	0,31*** (0,03)	0,84*** (0,12)	0,31*** (0,03)
Trámites exigidos	0,39*** (0,13)	0,15*** (0,05)	0,41*** (0,13)	0,16*** (0,05)
Líneas de crédito poco pertinentes	-0,03 (0,14)	-0,01 (0,05)	0,03 (0,14)	0,01 (0,05)
Experiencias pasadas negativas	-0,29 (0,19)	-0,11 (0,07)	-0,14 (0,19)	-0,05 (0,07)
No realizó innovaciones	-1,51*** (0,25)	-0,43*** (0,03)	-1,48*** (0,25)	-0,43*** (0,03)
Garantías exigidas	0,34* (0,20)	0,13** (0,07)	0,40** (0,19)	0,15** (0,07)
Recursos internos suficientes	0,69*** (0,05)	0,26*** (0,02)	0,61*** (0,05)	0,24*** (0,02)
Tamaño			0,19*** (0,02)	0,07*** (0,00)
Constante	-0,28*** (0,03)		-0,78*** (0,06)	
Número de observaciones	2.359	2.359	2.405	2.405
LR $\chi^2(2)$	234,78		313,44	
Prob > χ^2	0,00		0,00	
Log likelihood	-1.549,33		-1.510,00	

* 90% de significancia, ** 95% de significancia, *** 99% de significancia.

Fuente: elaboración propia con base en EIByC (2005).

El análisis sobre cómo opera la restricción, según los grados de innovación de las empresas, señala un patrón similar al del total, excepto para el caso de las ENI, que en el 56,3% de los casos indican que no accedieron por no innovar, y en la cuarta parte de estas empresas por experiencias pasadas negativas. Las EIE señalan como segunda razón por la cual no accedieron, las garantías exigidas, y las EPI mencionan los trámites necesarios (Sierra *et al.*, 2008).

Ahora bien, entre las empresas que usaron financiación externa, el 75% manifiesta no haber tenido problemas (Sierra *et al.*, 2008). Por lo tanto, el problema parece ser más del acceso al financiamiento que de la relación misma, una vez esta se ha logrado establecer por las empresas.

Entre las empresas que accedieron a la financiación y tuvieron problemas, el 80,6% señala el alto costo del crédito (Cuadro 13); en segundo lugar, aparece el nivel de garantías y el respaldo exigido, percibido como un obstáculo por el 49,2% de las empresas; en tercer lugar, la poca claridad en los criterios de evaluación y selección, señalado por cerca del 20% de las empresas, y en cuarto lugar, la falta de información y claridad sobre las condiciones del crédito, así como las altas exigencias técnicas para especificar el proyecto.²²

Cuadro 13

Problemas con la financiación en la industria de la región*

Razones	%
Falta de información y/o claridad sobre las condiciones del crédito	13,0
Altos costos del crédito	80,6
Demasiadas exigencias técnicas para especificar el proyecto	10,6
Poca claridad en los criterios de evaluación y selección	19,2
Garantías y respaldo exigido	49,2
Total empresas con problemas en la financiación (N°)	149,0

* Se calcula sobre empresas con problemas en la financiación.

Fuente: EIByC (2005).

²² Los ejercicios econométricos mostraron como significativa solamente la variable de los altos costos.

Como hipótesis podría plantearse que los mayores costos de los fondos se deberían a la aplicación de mecanismos de selección, seguimiento y control (*screening, monitoring, and enforcement*), destinados a evitar los problemas de selección adversa y riesgo moral que surgen debido al desconocimiento de los detalles de los proyectos por parte de los banqueros (Weigand, 1999). Tal desconocimiento pareciera estar asociado al débil desarrollo del sistema financiero, en particular para evaluar proyectos innovadores.

Los ejercicios también aportan indicios sobre las restricciones de carácter más institucional. La garantía y respaldo exigidos, así como la poca claridad en los criterios de evaluación y selección de los proyectos o solicitudes de financiación de las innovaciones, son percibidos como problemáticos por un porcentaje mayor de las empresas más innovadoras. Mientras tanto, la falta de información o de claridad sobre las condiciones del crédito son resaltados por las empresas menos innovadoras (Sierra *et al.*, 2008). Todo esto sugiere fallas de mercado asociadas al bajo desarrollo del sector financiero tradicional para evaluar y financiar proyectos de innovación.

En síntesis, los resultados sugieren que la asimetría de la información y el alto riesgo e incertidumbre asociados a este tipo de proyectos elevan los costos de acceso al capital (Weigand, 1999; Mohnen *et al.*, 2008). También sugieren que el sistema financiero colombiano está poco desarrollado para evaluar proyectos de innovación por las garantías exigidas —similares a las pedidas para financiar proyectos comunes de inversión física—, y el poco desarrollo de mecanismos para valorar intangibles asociados con la innovación y su calidad. Eso explicaría por qué las empresas que más innovan son las que perciben con más intensidad ese tipo de problemas.

CONCLUSIONES

El trabajo muestra la existencia de restricciones financieras que inhiben el surgimiento y el desarrollo de proyectos y procesos de innovación. También muestra que los instrumentos de política pública han sido insuficientes para remediar esas fallas de mercado, como lo muestra el poco

uso de estos por parte de las empresas industriales de ByC, especialmente por parte de las empresas de menor tamaño, que también son las menos innovadoras.

Los resultados del trabajo muestran elementos útiles para entender cómo operan las restricciones financieras en las empresas estudiadas. A las limitaciones provenientes del escaso desarrollo del sistema financiero, en general, se agregan las específicas, esto es, las relacionadas con el relativo desconocimiento de las características de los procesos y proyectos de innovación, que se revela en el poco desarrollo de los mecanismos de evaluación de este tipo de proyectos, y que ensancha los requisitos, trámites y costos de los créditos obtenidos y reduce el acceso a los recursos para financiar las actividades de I+D. Estos factores de índole institucional se agregan a los costos de los créditos que operan como la principal restricción.

Todo esto indica que la política debe contemplar tanto lo referente a los incentivos y apoyos con recursos directos como el débil desarrollo del sector financiero y los factores de orden institucional que inciden en esta situación. También se deben considerar las deficiencias en su aplicación, que inciden en la poca utilización de los instrumentos institucionales de financiación de la innovación.

El ejercicio efectuado plantea interrogantes que requieren ahondar en la indagación. Por ejemplo: ¿qué tan cierta es la suficiencia de recursos internos de las empresas para financiar la totalidad de los proyectos y, en particular, si las pymes tienen la capacidad de generar suficiente liquidez para atender sus necesidades de corto y mediano plazos y financiar, simultáneamente, proyectos que encierran alto riesgo e incertidumbre y cuyo horizonte de resultados es de mediano y largo plazos? O ¿será más bien que, dadas las restricciones financieras externas (desconocimiento de las fuentes oficiales, inconveniencia o exageradas condiciones de las fuentes privadas, inexistencia de capital de riesgo), las empresas se resignan a hacer lo que puedan con los fondos propios disponibles?

Al respecto son relevantes los trabajos de Tenjo *et al.* (2006) y Sarmiento y Salazar (2005) en relación con los factores que determinan la estructura de capital de las empresas colombianas, pues concluyen que las decisiones de financiación de las empresas nacionales están marcadas por

las imperfecciones de los mercados de recursos (derivadas de problemas de información imperfecta, que se manifiestan en concentración del crédito, poca disponibilidad de crédito de largo plazo, incipiente desarrollo del mercado accionario) que generan restricciones financieras en los mercados de fondos de financiamiento. ¿Cuánto más cierto será esto en el caso de la financiación de proyectos de innovación?

REFERENCIAS

- Aghion, P. y Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60 (2), 323-351.
- Ahmad, S. (1966). On the theory of induced innovation. *Economic Journal*, 76, 344-357.
- Allen, F. y Gale, D. (1995). A welfare comparison of intermediaries and financial markets in Germany and the US. *European Economic Review*, 39 (2), 179-209.
- Arrow, K. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29 (3), 155-173.
- Asobancaria (2009). *Reporte de bancarización a septiembre de 2008*. Bogotá: Vicepresidencia Económica-Dirección de Estudios y Regulación Financiera CIFIN.
- Avendaño, H. (2006). ¿Es demasiado costoso el microcrédito en Colombia? *Carta Financiera de ANIF*, 133, pp. 1-13. Recuperado el 23 de octubre de 2009, de <http://www.mincomercio.gov.co/econtent/Documentos/EstudiosEconomicos/documentos/EsCostoso-MicrocreditoenColombia.pdf>.
- Ayyagari, M.; Demirguc-Kunt, A. y Maksimovic, V. (2007). *Firm innovation in emerging markets*. Washington: World Bank Policy Research Working Papers.
- Barro, R. y Sala-i-Martin, X. (1997). Technological diffusion, convergence and growth. *Journal of Economic Growth*, 2 (1), 1-26.
- Becchetti, L. y Sierra, J. H. (2002). Financing innovation: Trodden and unexplored paths. *Cuadernos de Administración*, 15 (24), 7-40.
- Bell, M. y Pavitt, K. (1993). Technological accumulation and industrial growth: Contrasts between developed and developing countries. *Industrial and Corporate Change*, 2 (2), 157-210.
- Canepa, A. y Stoneman, P. (2002). *Financial constraints on innovations: A European cross country study*. Maastricht: Kiel Institute of World Economics.
- Cantwell, J. (2005). Innovation and competitiveness. En J. Fagerberg, D. Mowery y R. Nelson (Eds.), *The Oxford handbook of innovation*. Oxford: Oxford University Press.

- Claessens, S. y Perotti, E. (2007). Finance and inequality: Channels and evidence. *Journal of Comparative Economics*, 35 (4), 748-773.
- Cohen, W. y Levin, R. (1989). Empirical studies of innovation and market structure. En S. Schmalense y R. Willing (Eds.). *Handbook of industrial organization*. North Holland: Elsevier.
- Colciencias (2008). *Colombia construye y siembra futuro: política nacional de fomento a la investigación y la innovación*. Bogotá: autor.
- Consejo Nacional de Ciencia y Tecnología (1995a). *Política Nacional de Innovación y Desarrollo Tecnológico*. Documento de trabajo. Bogotá: Colciencias.
- (1995b). *Ciencia y tecnología para un desarrollo sostenible y equitativo: implementación de la Política Nacional de Ciencia y Tecnología: 1994-1998*. Documento de trabajo. Bogotá: Colciencias.
- Consejo Privado de Competitividad (CPC), (2007). *Informe nacional de competitividad 2007*. Bogotá: autor.
- Dasgupta, P. y Stiglitz, J. (1980). Industrial structure and the nature of the innovative activity. *Economic Journal*, 90, 266-293.
- Dosi, G. (1988a). The nature of the innovative process. En G. Dosi *et al.* (Eds.), *Technical change and economic theory*. London: Francis Pinter.
- (1988b). Sources, procedures and microeconomic effects of innovation. *Journal of Economic Literature*, 26 (3), 1120-1171.
- (1997). Opportunities, incentives and the collective patterns of technological change. *Economic Journal*, 107 (444), 1530-1547.
- Elster, J. (1982). *Explaining technical change: A case study in the philosophy of science*. Cambridge: Cambridge University Press.
- Fagerberg, J.; Mowery, D. y Nelson, R. (Eds.), (2005). *The Oxford handbook of innovation*. Oxford: Oxford University Press.
- Freeman, C. (1994). The economics of technical change. *Cambridge Journal of Economics*, 18, 463-514.
- Grossman, G. y Helpman, E. (1991). Quality ladders in the theory of growth. *Review of Economic Studies*, 58, 43-61.
- Léger, A. y Swaminathan, S. (2007). *Innovation theories: relevance and implications for developing country innovation*. Berlin: German Institute for Economic Research.
- Malaver, F. y Vargas, M. (2005). Las políticas de ciencia, tecnología e innovación en Colombia 1990-2005: sus logros, fracasos y desafíos. *Cuadernos de Administración*, 18 (30), 39-78.
- (2006). *Capacidades tecnológicas, innovación y competitividad de la industria de Bogotá y Cundinamarca: resultados de una encuesta de innovación*. Bogotá: Cámara de Comercio de Bogotá, Observatorio Colombiano de Ciencia y Tecnología, Agenda Regional de Ciencia y Tecnología y Consejo Regional de Competitividad.

- Mankiw, G.; Romer, P. y Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107 (2), 407-37.
- Mohnen, P.; Palm, F. C.; Schim van der Loeff, S. y Tiwari, A. (2008). *Financial constraints and other obstacles: Are they a threat to innovation activity?* Tokio: United Nations University.
- Myers, S. C. y Majluf, N. S. (1984). Corporate financing decisions when firms have investment information that investment do not have. *Journal of Financial Economics*, 48, 56-57.
- Nelson, R. (2001). Evolutionary theories of economic change. En A. Nicita y U. Pagano (Eds.), *The evolution of economic diversity*. London: Routledge.
- y Winter. S. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press.
- O'Sullivan, M. (2005). Finance and innovation. En J. Fagerberg, D. Mowery y R. Nelson (Eds.), *The Oxford handbook of innovation*. Oxford: Oxford University Press.
- Observatorio Colombiano de Ciencia y Tecnología (OCyT), (2008). *Indicadores de ciencia y tecnología Colombia 2007*. Bogotá: OCyT y Colciencias.
- Red de Indicadores de Ciencia y Tecnología (RICyT), (2007). *El estado de la ciencia: principales indicadores de ciencia y tecnología iberoamericanos/interamericanos*. Buenos Aires: RICyT.
- Romer, P. (1986). Increasing returns and long run growth. *Journal of Political Economy*, 94 (5), 1002-1037.
- (1990). Endogenous technical change. *Journal of Political Economy*, 98 (5), 1002-1037.
- Ruttan, V. (1997). Induced innovation, evolutionary theory and path dependence: Sources of technical change. *Economic Journal*, 107, 1520-1529.
- Sarmiento, R. y Salazar, M. (2005). *La estructura de financiamiento de las empresas: una evidencia teórica y econométrica para Colombia 1997-2004*. Documentos de Economía (16). Bogotá: Facultad de Ciencias Económicas y Administrativas, Departamento de Economía, Universidad Javeriana.
- Schumpeter, J. (1934). *The theory of economic development*. Cambridge: Harvard University Press.
- Sierra, J. H.; Malaver, F. y Vargas, M. (2008, 14-15 de agosto). *Finanzas e innovación en Colombia: un análisis a partir de la EIByC*. Documento procedente del I Congreso Internacional de Gestión Tecnológica e Innovación, Bogotá, Colombia, Universidad Nacional de Colombia.
- Sierra, J. H. y Londoño, D. (2008). RSE y mipymes: alternativas para apoyar su desarrollo en Colombia. *Contaduría Universidad de Antioquia* (52), 13-44.
- Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70 (1), 65-94.
- Stiglitz, J. y Weiss, A. (1981). Credit rationing in markets with imperfect information en *American Economic Review*, 71 (3), 393-410.

- Stockdale, B. (2002). UK Innovation Survey 2001. *Economic Trends*, 580, 36-42.
- Swan, T. (1956). Economic growth and capital accumulation. *Economic Record*, 32 (2), 394-361.
- Tenjo, F.; López, E. y Zamudio, N. (2006). *Determinantes de la estructura de capital de las empresas colombianas (1996-2002)*. Borradores de Economía (380). Bogotá: Banco de la República.
- Tiwari, A. K.; Mohnen, P.; Palm, F. C. y Schim van der Loeff, S. (2008). Financial constraints and R&D investment: Evidence from CIS, forthcoming. En A. Kleinknecht, R. Ott, C. van Beers y R. Verburg (Eds.), *Determinants of innovative behaviour: A firm's internal practices and its external environments*. London: Palgrave Publishers.
- Uribe, J. M. (2007). *Caracterización del mercado accionario colombiano, 2001-2006: un análisis comparativo*. Borradores de Economía (456). Bogotá: Banco de la República.
- Vernon, R. (1966). International investment and international trade in the product cycle. *Quarterly Journal of Economics*, 80, 190-207.
- Weigand, J. (1999). Innovation, investment and corporate finance. En D. Mueller, A. Haid y J. Weigand (Eds.). *Competition, efficiency, and welfare: Essays in honour of Manfred Neumann*. Dordrecht: Kluwer Academic Publishers.
- Winter, S. (1964). Economic natural selection and the theory of the firm. *Yale Economic Essays*, 4, 225-272.

Capítulo 9

Interés de las empresas por la financiación de sus actividades de I+D+i: un análisis en el marco de las redes complejas para el sector industrial manufacturero colombiano

Diego Leonardo Otálora Montenegro
Rafael Germán Hurtado Heredia
Carlos Quimbay Herrera

En este capítulo presentamos un análisis, en el marco de los estudios de las redes complejas, sobre el interés que tienen los empresarios del sector industrial manufacturero colombiano en acceder a financiación para actividades de investigación, desarrollo experimental e innovación (I+D+i) y sobre la visibilidad de las fuentes de financiación. El análisis se sustenta en el desarrollo de un modelo con dos parámetros cuantificables, cuyo fin es establecer el interés de los actores de un grupo para interactuar con actores de otro grupo, según sea la visibilidad de los últimos. Los fundamentos teóricos del modelo son el análisis de redes sociales y la mecánica estadística. El modelo se caracteriza por tener dos parámetros fáciles de interpretar, asociados al interés y la visibilidad, y umbrales que permiten establecer tendencias (Otálora, Hurtado y Quimbay, 2009a y 2009b).

La aplicación realizada parte de considerar estudios empíricos según los cuales la estructura financiera de las empresas presenta características acordes al grado de inversión en I+D+i que estas despliegan. Para explicarlo se han propuesto varios enfoques teóricos (Pereiras y Huergo 2006; Rajan y Zingales, 1995; Aghion *et al.*, 2004), pero, en forma general, se sabe que este tipo de inversión tiene restricciones financieras debido a sus características. En nuestra propuesta de análisis, la estructura financiera de una empresa o de un sector empresarial se expresa en sus patrones de relacionamiento con las fuentes de financiación y, así, evidencia el grado de factibilidad financiera para la innovación en dicha empresa o sector.

De este modo, guiados por los conceptos del análisis de redes sociales, sustentamos la aplicación del modelo en dos parámetros fundamentales, con los cuales se capturan rasgos generales del sistema real y nos permiten *parametrizar* el establecimiento de relaciones de financiación entre empresas y fuentes de financiación: el interés de las empresas por conocer fuentes de financiación, que en adelante llamaremos *interés*, y la visibilidad de las fuentes de financiación ante las empresas, que denominaremos *visibilidad* (Wasserman y Faust, 1994).

El *interés* de las empresas por financiar actividades de I+D+i puede realizarse, más allá de los recursos propios, con conocimiento e información sobre las fuentes de financiación externas, en particular de su existencia y sus propósitos. La *visibilidad* depende del interés de las fuentes de financiación por hacerse visibles, de sus relaciones con otras fuentes de financiación y de las relaciones que tengan las empresas con que se relacionan. Así, el *interés* y la *visibilidad* son los polos de la relación de conocimiento necesaria para establecer la relación de financiación. Por lo anterior, asumimos que el alcance del conocimiento que una empresa posee sobre las componentes y las características del sistema financiero es una medida de su *interés* en la financiación de la innovación y expresa su estructura financiera.

Así es como estudiar la estructura relacional asociada al conocimiento que tienen las empresas acerca de la existencia de fuentes de financiación para actividades de I+D+i y extraer información sobre el *interés* y sobre la *visibilidad* evidencia los propósitos de estos actores ante la financiación de la innovación. Teniendo en cuenta lo anterior, nos centramos en la

relación, definida por el conocimiento que tienen las empresas sobre la existencia de otros actores (en nuestro caso, fuentes de financiación) que sirven a sus propósitos e intereses para desarrollar actividades específicas. En adelante nos referiremos a esta relación como *relación de conocimiento*.

Nuestro análisis, enmarcado en los estudios sobre redes complejas,¹ aprovecha la correspondencia entre la dinámica de los sistemas y las propiedades topológicas de las redes que los representan relacionamente, propiedades que se han abordado con la aplicación de herramientas conceptuales y analíticas de la mecánica estadística (Albert y Barabasi, 2002; Park y Newman, 2004). En particular, desarrollamos un modelo para redes bipartitas,² que parte de la definición de relaciones entre nodos de distinto tipo a escala microscópica para establecer propiedades de conectividad de la red como un todo a escala macroscópica (Albert y Barabasi 2002; Newman, Barabasi y Watts, 2006). Este modelo extiende a redes bipartitas los resultados obtenidos para redes de un modo³ y expuestos en la literatura (Garlaschelli y Loffredo, 2008).

Una ventaja de este modelo es que permite establecer el *interés* de los empresarios en la financiación de actividades de I+D+i, independientemente de las restricciones establecidas por las fuentes de financiación y que determinan la relación de financiación misma. Si bien el acceso al proceso de financiación es indirecto, esta aproximación tiene la ventaja de desacoplar el *interés* de los empresarios por acceder a financiación para sus actividades de la ocurrencia de la financiación. En cuanto a la *visibilidad* de las fuentes de financiación, esta hace parte del modelo e incorpora la

¹ *Red compleja*: representación relacional de un sistema real con interacciones fuertes o frecuentes entre sus elementos que determinan una dinámica no lineal. Estos sistemas están compuestos típicamente por un gran número de elementos.

² *Red bipartita*: red cuyos elementos (nodos) se pueden dividir en dos conjuntos disyuntos y las relaciones se trazan de un conjunto al otro exclusivamente.

³ *Red de un modo*: red donde todos sus elementos (nodos) son de un mismo tipo, siendo todos los nodos estadísticamente homogéneos en un principio. *Red de dos modos o bimodal*: red bipartita en la que cada conjunto disyunto se distingue por poseer características diferentes.

dinámica relacional de las fuentes de manera independiente al *interés* de las empresas.

Aplicamos el modelo desarrollado al sector industrial manufacturero colombiano utilizando información proveniente de la Segunda Encuesta Colombiana de Innovación —Segunda Encuesta de Desarrollo e Innovación Tecnológica en la Industria Manufacturera Colombiana 2003-2004 (EDIT II)—, realizada por el Departamento Administrativo Nacional de Estadística (DANE), el Departamento Nacional de Planeación (DNP) y el Departamento Nacional de Ciencia, Tecnología e Innovación (Colciencias),⁴ en 2005 (DANE, DNP y Colciencias, 2006).

Este capítulo está organizado de la siguiente manera: en la primera sección se presenta el modelo; en la segunda se reseñan los aspectos generales del sistema empresas-fuentes de financiación analizado y una caracterización de algunas de sus propiedades de conectividad; en la tercera se encuentran los resultados, y, finalmente, se presenta una discusión de estos y nuestras conclusiones. En el Anexo se encuentra el desarrollo formal del modelo.

PLANTEAMIENTO DEL MODELO

El modelo desarrollado para N_e empresas y N_f fuentes de financiación parte de la existencia de un vínculo entre una empresa n_i y una fuente de financiación m_j . Este vínculo, definido como una variable de Bernoulli, tiene asociado un parámetro p que determina la probabilidad de existencia del vínculo y es función del *interés* e_i de la empresa i , y de la *visibilidad* v_j de la fuente de financiación j . Esta función, que resulta de considerar la extracción óptima de información a partir de la red (Anexo), está dada por la ecuación (1).

⁴ En el 2005, cuando se realizó la EDIT II, el Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas.

$$p_{ij} = \frac{e_i v_j}{1 + e_i v_j} \quad (1)$$

Donde e_i y v_j son conocidas como variables ocultas de la tendencia a conectarse de los nodos.

Mediante una aplicación inversa de esta formulación es posible obtener los valores de las variables ocultas de *interés* y de *visibilidad*. Para ello calculamos el valor esperado del número de vínculos⁵ k de cada nodo n_i o m_j a partir de la expresión para el grado de cada nodo (Anexo):

$$k(n_i) = \sum_{j=1}^{N_f} \frac{e_i v_j}{1 + e_i v_j}, k(m_j) = \sum_{i=1}^{N_e} \frac{e_i v_j}{1 + e_i v_j} \quad (2)$$

Estas representan un sistema de $N = N_e + N_f$ ecuaciones no lineales acopladas, una para cada nodo de la red, con N incógnitas. Al resolver este sistema de ecuaciones obtenemos información característica independiente a partir del grado de los N nodos, cuya distribución (distribución de grado) es una característica relacional del sistema.

SISTEMA EMPRESAS-FUENTES DE FINANCIACIÓN

El sistema está compuesto por 6.212 empresas industriales manufactureras y 39 fuentes de financiación. Las empresas fueron entrevistadas en la EDIT II, de carácter censal. A cada empresa entrevistada se le dio la opción de identificar las fuentes de financiación de actividades de I+D+i que conoce, entre 39 fuentes inscritas en un cuestionario cerrado, con la pregunta “¿Conoce la fuente de financiamiento?”. A partir de esta pregunta se establece la adyacencia entre las empresas y las fuentes de financiación clasificadas en el cuestionario en seis grupos (cuadros 1 y 2).

⁵ El número de vínculos de un nodo n_i se define como el grado del nodo $k(n_i)$ y, en conjunto para todos los nodos, se expresa como una distribución de grado (Wasserman y Faust, 1994).

Cuadro 1

Clasificación de las fuentes de financiación según la EDIT II

Sector	Número fuentes de financiación	Convención
Sector público, cofinanciación	6	I
Sector público, acceso a crédito	14	II
Banca privada, acceso a crédito	5	III
Sector externo, acceso a crédito	6	IV
Otros recursos	5	V
Otras fuentes empresariales	3	VI

Fuente: EDIT II.

RESULTADOS Y DISCUSIÓN

Presentamos en dos partes los resultados de la aplicación realizada: en la primera exploramos las características básicas de conectividad de la relación de conocimiento, a partir de las propiedades de adyacencia que determinan la distribución de grado para los nodos. En la segunda parte aplicamos el modelo desarrollado para encontrar los parámetros de *interés* y *visibilidad* e interpretamos sus valores y distribución para la información contenida en la EDIT II.

ANÁLISIS DE LAS PROPIEDADES DE LA RED SEGÚN LA DISTRIBUCIÓN DE GRADO

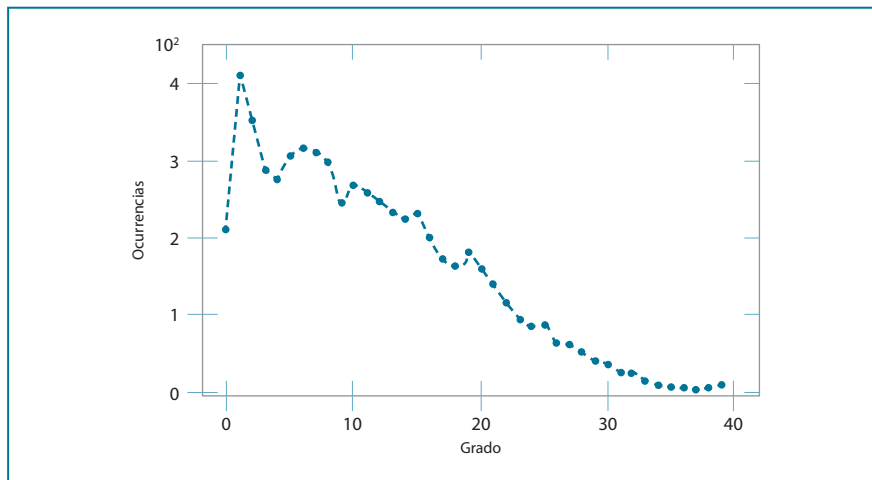
El conocimiento que manifiestan tener las empresas sobre la existencia de las fuentes de financiación define una red de afiliación representada por la matriz C con elementos $c_{ij} = 1$, si la empresa i dice conocer la fuente de financiación j , y $c_{ij} = 0$ de otra forma. Definimos el grado de una empresa $K(n_i)$ como el número de fuentes de financiación que la empresa dice conocer, es decir $K(n_i) = \sum_{j=1}^{N_f} c_{ij}$, y de forma análoga el grado de la fuente

de financiación $K(n_i)$, como el número de empresas que dicen conocerla $K(m_j) = \sum_{i=1}^{N_i} c_{ij}$.

El Gráfico 1a muestra la distribución de grado para las empresas en la red bimodal, definida por la relación de conocimiento entre empresas y fuentes de financiación. Observamos que la tendencia natural de las empresas es tener un número de vínculos en el rango inferior de la distribución de grado, mientras que pocas se encuentran en el rango superior. El Gráfico 1b muestra la misma distribución en escala logarítmica (Log-Log), donde observamos la existencia de un decaimiento tipo ley de potencias para el final de la distribución, al descontar los últimos dos datos, que corresponden a empresas que dicen conocer a 38 o 39 de las 39 fuentes de financiación y pueden atribuirse a una cola gruesa.

Gráfico 1a

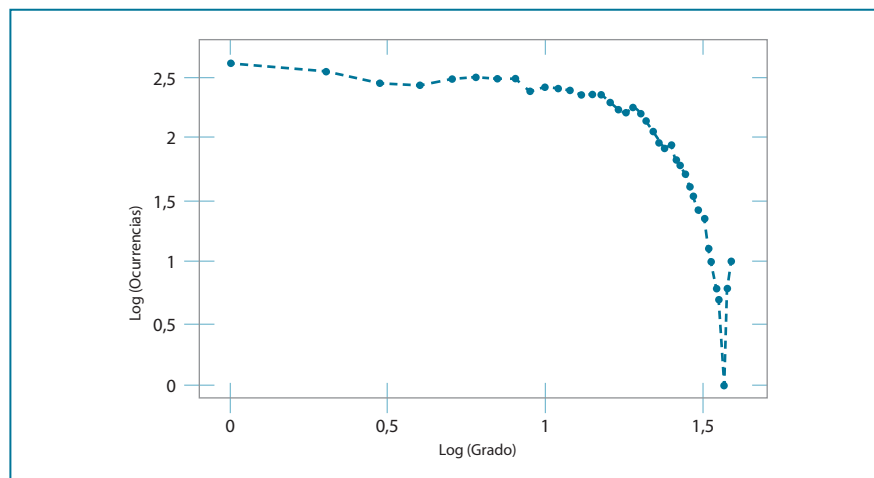
Distribución de grado para las empresas: muestra el número de empresas (ocurrencias) que dicen conocer un cierto número de fuentes de financiación (grado)



Fuente: EDIT II.

Gráfico 1b

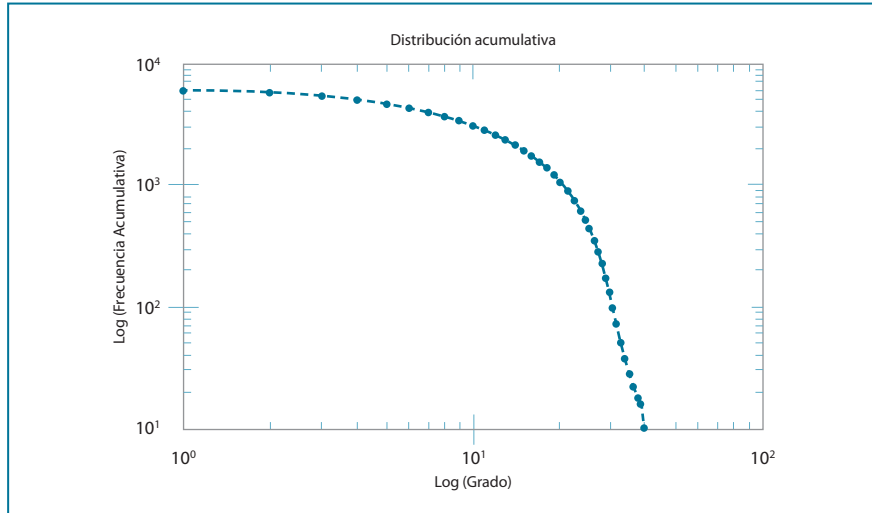
Distribución de grado para las empresas en escala logarítmica



Fuente: EDIT II.

Al graficar la distribución acumulativa para el grado de las empresas sobre ejes logarítmicos (Gráfico 2) podemos confirmar la aparición de un decaimiento de tipo ley de potencias para la cola de la distribución, el cual se relaciona con la regla de Pareto, según la cual este tipo de distribuciones aluden a que muchos tienen poco —en este caso, vínculos de conocimiento— y pocos tienen mucho —y que puede interpretarse desde la perspectiva de la distribución de riqueza— (Garlaschelli y Lofredo, 2004a y 2004b). Este comportamiento en la distribución de grado revela que la vinculación entre empresas y fuentes de financiación no es aleatoria, sino que resulta de un fenómeno evolutivo (Caldarelli *et al.*, 2002) o uno de un proceso de vinculación preferencial (Albert y Barabasi, 2002).

Gráfico 2

Distribución acumulativa del grado de las empresas en la red de conocimiento

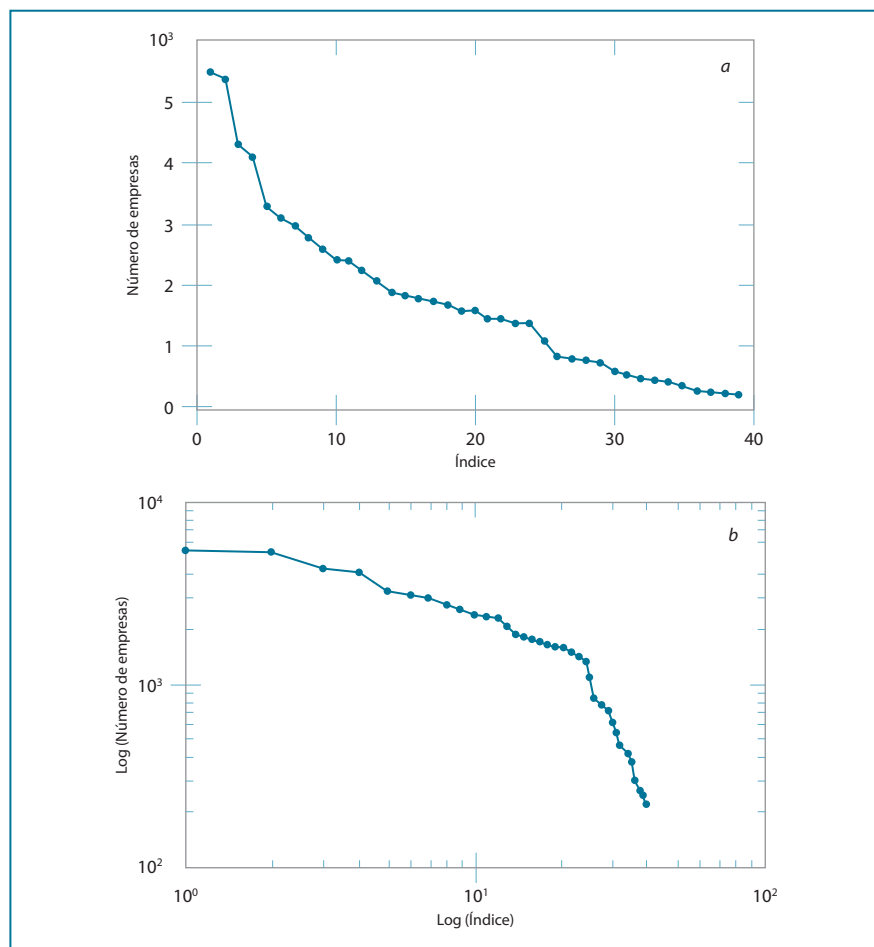
Nota: la grafica sobre ejes logarítmicos revela un comportamiento de ley de potencias.
Fuente: EDIT II.

Si fuera un fenómeno evolutivo, la distribución de grado reflejaría la acumulación de vínculos por parte de los nodos más antiguos en la red —en este caso, fuentes de financiación con mayor tiempo en el sistema—. En el caso de la vinculación preferencial, la acumulación reflejaría habilidades de las fuentes de financiación para atraer vínculos a ellas. Estas habilidades pueden ser endógenas o exógenas, en el primer caso hacen parte de los atributos de los nodos y en el segundo resultan de la posición de los nodos en la red.

En el Gráfico 3a se presenta el grado de cada fuente de financiación (ordenadas según su grado), donde se evidencia la existencia de un comportamiento de ley de potencia. Con el fin de corroborar esta apreciación, en el Gráfico 3b se presenta la misma información en escala logarítmica, y se aprecia la existencia de dos tendencias generales (A y B) que pueden ser asociadas independientemente a una ley de potencias (en cada caso con exponente distinto).

Gráfico 3

(a) Gráfica de Zipf para el número de empresas que conocen la fuente de financiación;
 (b) gráfica de Zipf-Mandelbröt para el número de empresas que conocen la fuente de financiación*



* En los dos gráficos el índice de cada fuente de financiación denota su posición respecto a los otros en orden descendente. Fuente: EDIT II.

Desde nuestro análisis podemos asegurar que los patrones relacionales de dos fuentes de financiación que pertenecen a distintas tendencias son diferentes. En la segunda columna del Cuadro 2 se muestra el grado de cada fuente de financiación de manera porcentual, con relación al número total de empresas, y en la cuarta columna, el rango A o B al que pertenece.

Cuadro 2

Porcentaje de conectividad y visibilidad para las fuentes de financiación

Fuente de financiación	Porcentaje	Visibilidad	Sector	Rango
Recursos propios	88,2	1,5E+02	VI	A1
Bancos comerciales	86,5	1,1E+02	III	A1
Corporaciones financieras	69,4	1,6E+01	III	A2
Compañías de financiamiento comercial	66,3	1,2E+01	III	A2
Fondos de inversión o fiducias	53,1	4,6E+00	III	A3
Proexport	50,0	3,7E+00	I	A3
Banca de inversion	48,1	3,2E+00	III	A3
Bancoldex: crédito	44,7	2,6E+00	II	A3
Bancoldex: inversión activos fijos	41,9	2,1E+00	II	A3
Cajas de compensación familiar	38,9	1,7E+00	V	A3
Sena (Ley 344 de 1996)	38,5	1,7E+00	I	A3
Fondo Nacional de Garantías	36,7	1,5E+00	II	A3
Bancoldex: Expopyme-Multipropósito	34,0	1,2E+00	II	A3
Finagro	30,4	9,6E-01	II	A3
Bancoldex-Colciencias: I+D	29,4	8,8E-01	II	A3
Bancoldex: gestión ambiental	28,3	8,1E-01	II	A3
Fomipyme	27,9	7,9E-01	I	A3
Colciencias: universidad-empresa	27,1	7,4E-01	I	A3
Financiamiento otra empresas	25,6	6,6E-01	VI	A3
Findeter	25,5	6,6E-01	II	A3
Fondos parafiscales	23,6	5,6E-01	VI	A3
Colciencias: otras líneas	23,4	5,6E-01	I	A3
Banca comercial internacional	22,3	5,1E-01	IV	A3
Banca de inversión	22,1	5,0E-01	IV	A3
Organismos internacionales: OEA, ONU, UE	17,7	3,3E-01	IV	A3
Financiamiento otras empresas del grupo	13,6	2,2E-01	VI	B
Centros de desarrollo tecnológico	13,2	2,1E-01	V	B
Recursos de universidades	12,6	1,9E-01	V	B

Continúa

Fuente de financiación	Porcentaje	Visibilidad	Sector	Rango
Fondo agropecuario de garantías	11,5	1,7E-01	IV	B
Fondos de capital de riesgo	9,8	1,3E-01	V	B
Cooperación internacional	8,9	1,1E-01	IV	B
Aportes casa matriz	7,5	9,0E-02	IV	B
CARANA	7,2	8,5E-02	V	B
Fondos departamentales o municipales	6,9	8,1E-02	II	B
Ecofondo	6,1	6,8E-02	II	B
Programa Transferencia Tecnología Agropecuaria	4,7	4,9E-02	I	B
Otros	4,2	4,1E-02	II	B
Programa Desarrollo Microempresa Rural	4,0	4,0E-02	II	B
Alianzas productivas para la paz	3,6	3,5E-02	II	B

Fuente: EDIT II.

Del Gráfico 3b y del Cuadro 2 argumentamos que las tendencias identificadas corresponden en el primer rango (A) a fuentes de financiación pertinentes para las empresas, y en el segundo (B), a fuentes menos pertinentes. Como argumento general, nótese que en la tendencia A se encuentran los actores tradicionales de la financiación para las empresas, asociados al sistema financiero nacional, mientras que en la tendencia B se encuentran actores que tienen relación con otros sectores —programa CARANA, Proyecto de Apoyo al Desarrollo de la Microempresa Rural (Pademer), Fondo agropecuario de garantías (FAG), el Programa Nacional de Transferencia de Tecnología Agropecuaria (Pronatta), Alianzas Productivas por la Paz (APP) y Ecofondo— o que tienen poca o ninguna actividad de financiación, como las universidades, los centros de desarrollo tecnológico (CDT), los fondos de capital de riesgo del país y los programas de cooperación internacional.

Nótese que entre más pendiente sea la recta definida por la ley de potencias, es mayor la variación en el número de empresas que las conocen, lo que interpretamos como un criterio de exclusividad. En el Gráfico 3a y en el Cuadro 2 podemos distinguir, con mayor detalle, la existencia de cuatro

rangos para el grado de las fuentes de financiación, los cuales corresponden a tres subdivisiones del rango A (A1, A2 y A3) y la cuarta idéntica a B.

En el primer rango de esta división (A1) se encuentran los recursos propios de las empresas (recursos propios) y la banca comercial, conocidos por alrededor del 87% de las empresas; en el segundo rango (A2) las corporaciones financieras y las compañías de financiamiento comercial, conocidas por alrededor del 67%; el tercer rango (A3) está conformado por 21 fuentes de financiación, conocidas por un intervalo del 20% al 50% de las empresas, y el cuarto rango (B), conformado por 14 fuentes de financiación conocidas por menos del 13% de las empresas.

En general, los recursos propios y el sector de la banca privada tienen los mayores valores de grado, mientras que las iniciativas del sector público se encuentran en los rangos tercero y cuarto. Entre los actores del sector público se distinguen aquellos con mayor grado, como los programas del Ministerio de Comercio, Industria y Turismo (como Bancoldex y Proexport), y aquellos con menor grado orientados a sectores o a fines ajenos a la industria manufacturera. El sector externo se ubica en bajos índices de popularidad, incluida la financiación proveniente de la casa matriz, de cooperación internacional y crédito externo.

De los gráficos 1a y 3a deducimos que, en general, pocas fuentes de financiación tienen un grado alto (muchas empresas que manifiestan conocerlas) y muchas fuentes de financiación tienen un grado bajo, lo cual, en nuestro caso, se puede interpretar como que el conocimiento que tienen las empresas de aquellas fuentes de financiación más comunes no implica el conocimiento de aquellas menos comunes y, en cambio, el conocimiento de las menos comunes implica el conocimiento de las más comunes.

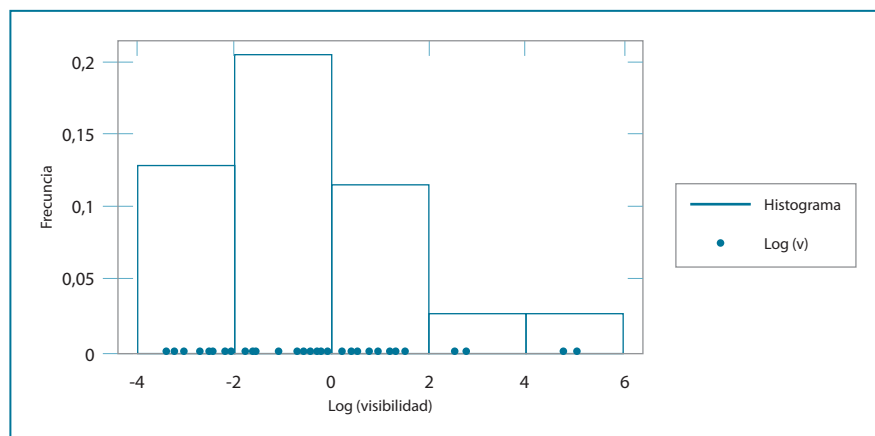
VISIBILIDAD DE LAS FUENTES DE FINANCIACIÓN E INTERÉS DE LAS EMPRESAS

En el Gráfico 4 se presenta la distribución de la *visibilidad* de las fuentes de financiación, obtenida según el modelo presentado en la sección “Planteamiento del modelo”. En el Gráfico 4a el histograma muestra la frecuencia que tiene cada *visibilidad* en términos del número de fuentes de financiación en un rango de *visibilidad*. El Gráfico 4b muestra en escala

logarítmica la *visibilidad* de las fuentes de financiación ordenadas según su *visibilidad*, de mayor a menor. Estos gráficos revelan una diferencia clara entre los diferentes tipos de fuentes de financiación, de acuerdo con los resultados de la sección anterior, Gráfico 3 y Cuadro 2.

Gráfico 4a

Histograma de frecuencias de la *visibilidad* de las fuentes de financiación*

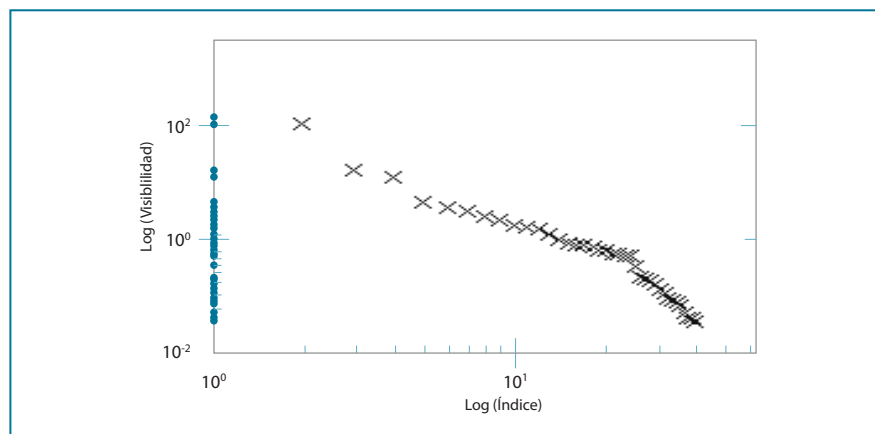


* Los puntos indican el valor particular de la *visibilidad*.

Fuente: EDIT II.

Gráfico 4b

De Zipf-Mandelbröt para la *visibilidad* de las fuentes de financiación*



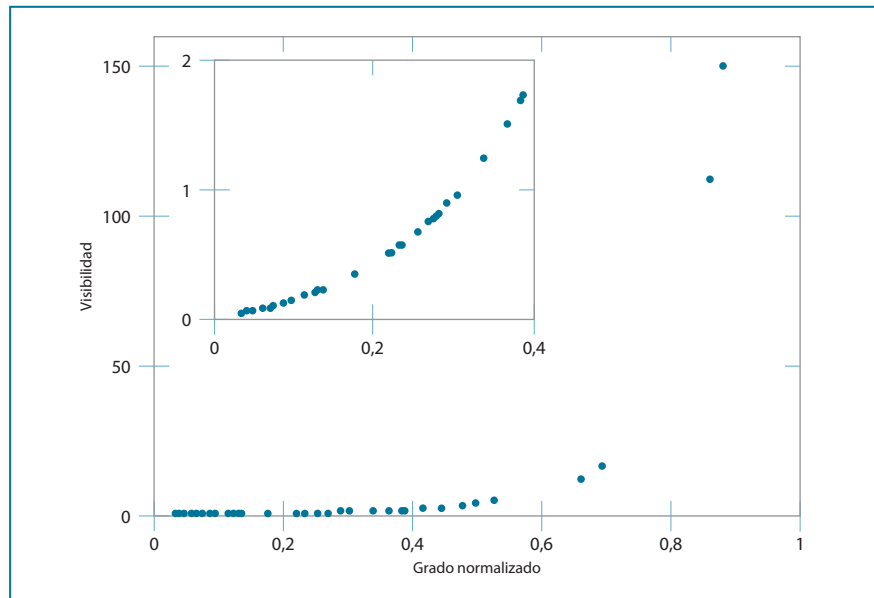
* El índice de cada fuente de financiación denota su posición respecto a las otras en orden descendente.

Fuente: EDIT II.

Esta diferencia es evidente en el Gráfico 4b, en la cual la *visibilidad* es mayor para las cuatro principales fuentes de financiación (con *visibilidades* entre 2 y 6 en el Gráfico 4a) y dista significativamente de la de las fuentes pertenecientes al grupo mayoritario (con *visibilidades* entre -4 y 2 en el Gráfico 4a). En el Cuadro 2 se muestran los valores de *visibilidad* para cada fuente de financiación, y en el Gráfico 5, la relación entre *visibilidad* y el grado, este normalizado a (1).

Gráfico 5

Visibilidad de la fuente de financiación contra grado normalizado a (1)



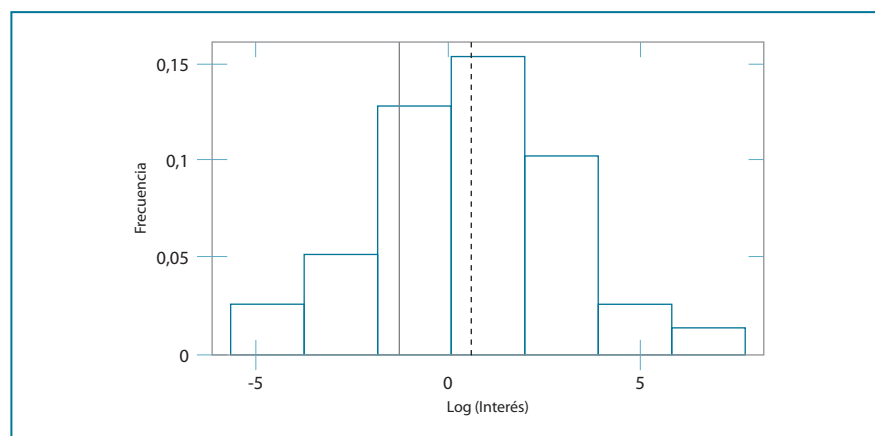
Nota: en el gráfico interior se muestra un acercamiento de la misma relación para el rango (0,4).
Fuente: EDIT II.

Los posibles valores de *interés* de las empresas siguen una distribución aproximadamente log-normal (con valor p de 0,998, en el test de Kolmogorov-Smirnov), así como se evidencia en el Gráfico 6a, el cual muestra que un grupo de empresas indiferentes ante un ambiente como el dado por las *visibilidades* de las fuentes de financiación tendría un comportamiento típico de distribución log-normal.

El valor esperado del *interés* calculado sobre la distribución en la cual las empresas son indiferentes ante las *visibilidades* de las fuentes de financiación (línea punteada en el Gráfico 6a) es mayor que el valor promedio calculado sobre el *interés* real de las empresas (línea continua en el Gráfico 6a). Lo anterior lo interpretamos como un signo de aversión de las empresas a construir relaciones con las fuentes de financiación para actividades de I+D+i, y por ende como un signo de aversión a las actividades de I+D+i.

Adicionalmente, en la distribución de empresas sobre los posibles valores de *interés* (Gráfico 6b) se muestra que la mayoría de las empresas se sitúan en los valores de *interés* por debajo del promedio esperado y que al aumentar el valor del *interés* el número de empresas con ese *interés* decae siguiendo una ley de potencias.

Gráfico 6a
Histograma de frecuencias del *interés* de las empresas

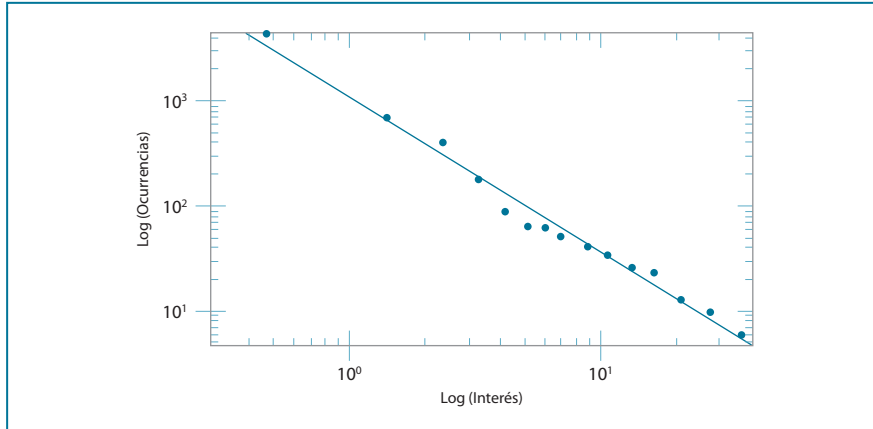


Nota: la línea punteada señala el *interés* esperado de la distribución; mientras que la línea continua, el *interés* promedio de las empresas encuestadas.

Fuente: EDIT II.

Gráfico 6b

Interés que muestra el número de empresas con *interés* en un rango dado sobre una escala logarítmica



Nota: el ajuste lineal indica que el número de empresas decae con el interés, siguiendo una ley de potencias.
Fuente: EDIT II.

Lo anterior se refleja parcialmente en las respuestas que dan los empresarios al cuestionamiento “La principal razón por la cual no utiliza las líneas de financiamiento que ofrece el sector público es”, al cual responden 5.487 empresarios. El 12,2% de ellos desconoce los mecanismos de financiación, el 26,6% manifiesta que les falta información, el 16,4% no cumple con los requisitos, el 9,2% tiene inconvenientes con el tiempo de los trámites y el 35,6% tiene otros motivos. Así, por lo menos, el 38,8% de las empresas refleja en su respuesta problemas de *interés*, teniendo en cuenta que ninguna de las fuentes de financiación permanece oculta.

CONCLUSIONES

En la aplicación realizada encontramos que, gracias a que el modelo desacopla mutuamente el interés de los empresarios y la visibilidad de las fuentes de financiación, podemos identificar el interés de los empresarios, en promedio, por conocer fuentes de financiación para actividades de I+D+i y si tienen poco interés o “aversión” o, por el contrario, tienen un alto in-

terés. En el caso colombiano, para reproducir las observaciones empíricas obtenidas a partir de la EDIT II, encontramos que los empresarios en promedio tienen poco interés o “aversión” por conocer fuentes de financiación, a pesar de que la red tiene una densidad del 0,285, es decir, que el 28,5% de los vínculos de conocimiento posibles existen. Este resultado parece ser sorprendente por la relativamente alta densidad de vínculos en la red y evidencia el alcance de la metodología desarrollada.

Por otra parte, la distribución de visibilidades está relacionada con la distribución de grado y marca las mismas tendencias encontradas en la primera parte del análisis presentado, el cual es de carácter descriptivo, y muestra un gran diferencia entre la percepción que tienen los empresarios de las fuentes de financiación tradicional (recurso propios, banca privada) y las otras formas de financiamiento.

REFERENCIAS

- Aghion, P.; Bond, S.; Klemm, A. y Marinescu, I. (2004). Technology and financial structure: Are innovative firms different? *Journal of the European Economic Association*, 2 (2-3), 277-288.
- Albert, R. y Barabasi, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74 (1), 47-97.
- Bianconi, G. (2008). Entropy of randomized network ensembles. *Europhys. Lett.*, 81 (2), 28005-28011.
- Burda, J.; Correia, D. y Krzywicki, A. (2001). Statistical ensemble of scale-free random graphs. *Phys. Rev.*, 64 (4), 46118-46121.
- Caldarelli, G.; Capocci, A.; de los Ríos, P. y Muñoz, M. A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.*, 89 (25), 258702-258706.
- Departamento Administrativo Nacional de Estadística (DANE), Departamento Nacional de Planeación (DNP), Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Francisco José de Caldas (Colciencias), (2006). *Innovación y desarrollo tecnológico en la industria manufacturera, Colombia 2003-2004*. Bogotá: autores.
- Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1 (3), 215-239.
- (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver: Empirical Press.

- Garlaschelli, D. y Loffredo, M. (2004a). Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.*, 93, 3868-3873.
- (2004b). Wealth dynamics on complex networks. *Phys. A.*, 338, 113-118.
- (2008). Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev.* 78, 15101-15105.
- Lundvall, B.-A. E. (1992). *National innovation systems: Towards a theory of innovation and interactive learning*. London: Printer.
- Mohenen, P.; Palm, F. C.; Schim van der Loeff y Tiwari, A. (2008). *Financial constraints and other obstacles: Are they a threat to innovation activity?* Tokio: United Nations University.
- Nelson, R. R. (Ed.), (1993). *National innovation systems: A comparative study*. New York: Oxford University Press.
- Newman, M.; Barabasi, A.-L. y Watts, D. J. (2006). *The structure and dynamics of networks*. New Jersey: Princeton University Press.
- Otálora, D.; Hurtado, R. y Quimbay, C. (2009a). *Un modelo simple para capturar la dinámica del sistema empresas-fuentes de financiación: el caso del conocimiento de la existencia de fuentes de financiación para innovación y desarrollo tecnológico en el sector industrial manufacturero colombiano*. Documento presentado para el XIII Seminario Latino-Ibero Americano de Gestión Tecnológica (Altec), 2009, Cartagena, Colombia.
- (2009b). Tratamiento mecánico-estadístico de redes bipartitas y aplicación a la relación de conocimiento entre empresas industriales manufactureras colombianas y fuentes de financiación para actividades de innovación y desarrollo tecnológico. Artículo sometido para publicación en la *Revista Colombiana de Física*.
- Park, J. y Newman, M. E. J. (2004). Statistical mechanics of networks. *Phys. Rev.*, 70 (6), 66117-66130.
- Pereiras, M. S. y Huergo, E. (2006). *La financiación de las actividades de investigación, desarrollo e innovación: una revisión de la evidencia sobre el impacto de las ayudas públicas*. Madrid: CDTI (España).
- Rajan, R. G. y Zingales, L. (1995). What do we know about capital structure?: Some evidence from international data. *Journal of Finance*, 50 (5), 1421-1460.
- Wasserman, S. y Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press
- Wasserman, S.; Scott, J. y Carrington P. J. (2005). *Models and methods in social network analysis*. Cambridge: Cambridge University Press.

ANEXO

En este anexo explicamos de forma concisa el trasfondo técnico desde el cual se obtiene la ecuación (1), que corresponde a la probabilidad de la ocurrencia del vínculo de conocimiento entre una empresa y una fuente de financiación como una función de las variables de *interés* y *visibilidad*.

Inspirados por los métodos de la física estadística, Park y Newman (2004) desarrollaron un método por el cual se logra extraer información a partir de configurar vínculos en una red. Su concepto básico de partida es el ensamble. Un ensamble en física estadística de redes G se obtiene al darle un peso estadístico $P(g \in G)$ a cada red g de un conjunto maximal dado. Para exponer las ideas básicas de esta metodología partiremos de la definición de entropía para el ensamble de redes:

$$S = - \sum_{g \in G} P(g) \ln(P(g)) \quad (1)$$

Imponemos que la distribución de probabilidad $P(g)$ está sujeta a la condición de reproducir los valores medidos de los observables x_i como valores esperados sobre el ensamble, es decir:

$$\langle x_i \rangle = \sum_{g \in G} P(g) x_i(g) \quad (2)$$

Y a la condición de normalización:

$$1 = \sum_{g \in G} P(g) \quad (3)$$

Al maximizar la expresión para la entropía (1) respetando las restricciones (2) y (3), obtenemos:

$$p(g) = \frac{e^{-H(g)}}{Z} \quad (4)$$

Donde $H(g)$ se define como el hamiltoniano del grafo:

$$H(g) = \sum_i \theta_i x_i(g) \quad (5)$$

Con $\Theta = \{\theta_j\}$ un campo de multiplicadores de Lagrange y Z es la función de partición dada por:

$$Z = \sum_{g \in \mathcal{G}} e^{-H(g)} \tag{6}$$

Conceptualmente, la imposición de las restricciones sobre los observables desempeña el papel de un baño térmico para el sistema, en cuanto la realización de un grafo con los valores promedio impuestos para los observables favorece algunas configuraciones de grafos y, por lo tanto, los grafos en el ensamble estadístico no tienen la misma probabilidad de ocurrencia.

En un grafo bipartito no direccionado se tienen dos conjuntos de nodos $\{n\}$ y $\{m\}$, con cardinales N y M , respectivamente, que en el caso de nuestra aplicación son el número de empresas y fuentes de financiación, y un conjunto de vínculos $\{v\}$, con cardinal L , que se trazan entre nodos de los conjuntos $\{n\}$ y $\{m\}$. A continuación extendemos el tratamiento desarrollado por Park y Newman (2004) para redes monomodales, $N \times N$, a redes bimodales, $N \times M$, que en general pueden ser representadas como grafos bipartitos.

Para el caso de las redes bipartitas no direccionadas tenemos la matriz de adyacencia con elementos:

$$\sigma_{ij} = \begin{cases} 1 & \text{si } i \text{ está conectado con } j \\ 0 & \text{de otra forma,} \end{cases} \tag{7}$$

Una propiedad específica de la red es su secuencia de grado $k_g(s)$, donde s representa cualquier nodo de la red $\{k_j\}_g$. En este caso, el hamiltoniano es:

$$H(g) = \sum_s \theta_s k_g(s) \tag{8}$$

Donde ahora se tiene un parámetro θ_s para cada vértice s , este parámetro corresponderá al valor de *interés* o *visibilidad*. El grado k del nodo s en el grafo g se extrae a partir de la matriz de adyacencia (9) de la siguiente manera:

$$k_g(s) = \sum_j \sigma_{js}, \text{ si } s \in \{n\} \tag{9a}$$

$$k_g(s) = \sum_i \sigma_{is}, \text{ si } s \in \{m\} \quad (9b)$$

Así, si el nodo s pertenece al grupo $\{n\}$ la sumatoria se desarrolla sobre el segundo índice (9a), y si pertenece al grupo $\{m\}$ la sumatoria se desarrolla sobre el primer índice (9b). De (8) y (9) el hamiltoniano es:

$$H(g) = \sum_{ij} (\theta_i + \theta_j) \sigma_{ij} \quad (10)$$

La función de partición ahora es:

$$\begin{aligned} Z &= \sum_{\{\sigma_{ij}\}} \exp\left(-\sum_{ij} (\theta_i + \theta_j) \sigma_{ij}\right) \\ &= \prod_{ij} \left(1 + e^{-(\theta_i + \theta_j)}\right) \end{aligned} \quad (11)$$

Entonces podemos calcular la probabilidad de ocurrencia de cada vínculo:

$$\begin{aligned} P_{ij} &= \langle \sigma_{ij} \rangle = \frac{\partial}{\partial} \left\{ -\frac{1}{Z} \frac{\partial Z}{\partial (\theta_i + \theta_j)} \right\} = \frac{e^{-(\theta_i + \theta_j)}}{1 + e^{-(\theta_i + \theta_j)}} \\ P_{ij} &= \frac{x_i y_j}{1 + x_i y_j} \end{aligned} \quad (12)$$

Donde x_i y x_j son variables ocultas que describen la tendencia de los nodos a conectarse (Bianconi, 2008; Garlaschelli y Loffredo, 2008; Caldarelli *et al.*, 2002).

Al observar las propiedades del ensamble que reproduce una distribución de grado encontramos que la probabilidad de que exista el vínculo entre dos nodos está dada por (12), donde cada nodo i de $\{n\}$ está caracterizado por un parámetro x_i y cada nodo j de $\{m\}$ lo está por un parámetro y_j . Estos parámetros desconocidos inicialmente, $\{x_i\}$ y $\{y_j\}$, identifican los nodos i con $k_g(i)$ y los j con $k_g(j)$. Para obtenerlos calculamos el valor esperado del grado para los nodos en $\{n\}$ y en $\{m\}$:

$$k_g(i) = \left\langle \sum_j \sigma_{ij} \right\rangle = \sum_j \frac{\partial F}{\partial (\theta_i + \theta_j)} = \sum_j \frac{x_i y_j}{1 + x_i y_j} \quad (13a)$$

$$k_g(j) = \left\langle \sum_i \sigma_{ij} \right\rangle = \sum_i \frac{\partial F}{\partial (\theta_i + \theta_j)} = \sum_i \frac{x_i y_j}{1 + x_i y_j} \quad (13b)$$

Las ecuaciones en (13) representan un sistema de $N+M$ ecuaciones no lineales acopladas, una para cada nodo de la red, y $N+M$ incógnitas x_i y y_j . Al resolver este sistema de ecuaciones habremos obtenido, a partir de los $N+M$ grados de los nodos, los $N+M$ elementos de información característica independiente en x_i y y_j .

Capítulo 10

Análisis comparativo entre árboles de regresión y clasificación (CART) y regresión logística: aplicación a la caracterización de empresas innovadoras colombianas

Sandra Carolina Serna Pineda

Juan Carlos Correa Morales

Jorge Robledo Velásquez

La definición, clasificación y caracterización de empresas que por su naturaleza, actividad o desempeño presentan un interés especial de tipo académico, político y gerencial, es un problema que los analistas han abordado de tiempo atrás —véase, por ejemplo, Kannebley Jr., Porto y Pazello (2005); Grinstein y Goldman (2006), y Koberg, Sarason y Rosse (1996)—. En nuestro caso, el problema se concreta en torno a la definición de empresas innovadoras y su clasificación y caracterización, con el objetivo básico de desarrollar herramientas analíticas que contribuyan a mejorar nuestro entendimiento de las dinámicas de la innovación y aplicarlo en la política pública y la gestión empresarial.

En estadística, la clasificación de individuos u objetos en determinados grupos o poblaciones ha sido ampliamente estudiada; para ello se han desarrollado varias técnicas, como el análisis discriminante lineal (Fisher,

1936), el análisis discriminante no métrico (Raveh, 1989), la regresión logística y los árboles de regresión y clasificación (Breiman *et al.*, 1984), entre otros.

En este capítulo se aplican dos de las técnicas mencionadas, los árboles de regresión y clasificación (CART, por su sigla en inglés) y la regresión logística a las empresas innovadoras colombianas, utilizando los datos de la Primera Encuesta Colombiana de Innovación. El propósito es doble: posibilitar el análisis comparativo de ambas técnicas y explorar su aplicación en la caracterización de las empresas innovadoras de producto, de proceso y de organización. La Primera Encuesta de Innovación se seleccionó porque la naturaleza y la estructura de sus datos era más propicia para lograr el propósito enunciado, sin que ello signifique que los resultados no sean útiles para abordar el análisis de otras encuestas de innovación y relacionadas.

La clasificación obtenida y las características representativas de los grupos de empresas permiten múltiples interpretaciones que enriquecen los análisis pertinentes desde diversas perspectivas de política pública y gestión empresarial. Además, con la aplicación paralela de dos técnicas de clasificación se pueden analizar comparativamente los resultados y se identifican las potencialidades y limitaciones de las técnicas respecto al objeto de estudio y el tipo de datos disponibles.

En la siguiente sección se presentan las definiciones y clasificaciones utilizadas para el análisis; luego, las dos técnicas de clasificación que se aplicarán a los datos de la Primera Encuesta de Innovación; después, los resultados obtenidos de la aplicación de las técnicas para, por último, concluir el trabajo.

DEFINICIONES Y CLASIFICACIONES

La Primera Encuesta Colombiana de Innovación —Encuesta sobre el Desarrollo Tecnológico del Establecimiento Industrial Colombiano (EDT)—, elaborada por el Departamento Nacional de Planeación (DNP) y el Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología

Francisco José de Caldas (Colciencias), en 1996, fue la primera encuesta de este tipo realizada en Colombia. Su aplicación permitió clasificar las empresas del sector manufacturero desde la perspectiva de la innovación.

El marco analítico de la Encuesta define tres categorías de empresas: empresas innovadoras, empresas potencialmente innovadoras y empresas no innovadoras. A su vez, las empresas innovadoras pueden ser innovadoras en sentido estricto (IE) —aquellas que realizan actividades de investigación y desarrollo (I+D) y otras actividades conducentes a la innovación y obtienen innovaciones de alcance internacional— o innovadoras en sentido amplio —aquellas que realizan cualquier actividad de innovación y logran innovaciones para el mercado nacional o la propia empresa— (Sistema Estratégico de Información [SEI], 1996; Salazar y Vargas, 1998).

Como empresa innovadora se define aquella que ha introducido una innovación tecnológica de producto o de proceso en el período bajo análisis; las potencialmente innovadoras, si bien no han llegado a la innovación, realizan alguna de las actividades de innovación que las conducirán, eventualmente, a introducir innovaciones. Para la Primera Encuesta se consideraron los siguientes tipos de innovación (SEI, 1996: 30):

1. *Innovación de productos*: adquisición, asimilación o imitación de nuevas tecnologías para mejorar tecnológicamente productos, para comenzar a producir productos que no existían en la empresa o para innovar con productos no existentes en el mercado.
2. *Innovación de procesos*: adquisición, asimilación o imitación de nuevas tecnologías para mejorar tecnológicamente procesos productivos existentes en la empresa, para comenzar a utilizar procesos que no existían en la empresa o para innovar en procesos inexistentes en el mercado. *Nota*: se entiende por mejora tecnológica de un producto o proceso el desarrollo de un producto o proceso existente con mejoras sustanciales en los beneficios generados o en su desempeño.
3. *Cambios en las formas de organización y administración*: esto incluye cambios tanto en la organización del proceso productivo como en la organización y gestión del establecimiento, en general, que

implican cambios radicales en las estrategias corporativas, basados fundamentalmente en las posibilidades abiertas por las nuevas tecnologías informáticas.

4. *Cambios en el empaque y embalaje*: se refiere a todo tipo de mejora que se introduzca en el embalaje o envoltorio del producto final de la firma, que no altera sustancialmente sus propiedades. Por ejemplo, pasar de una presentación de leche en caja sin troquel a otra con troquel.

El tercer tipo de innovación, cambios en las formas de organización y administración, se desagrega en dos formas específicas de innovación, a saber:

Cambios en las formas de organización y gestión del establecimiento: se refiere a los cambios en la gestión de la firma con el fin de realizar tanto los procesos productivos como los administrativos y de comercialización en una forma más eficiente. Es lo que se conoce como “tecnologías blandas”. Por ejemplo: re-ingeniería de procesos administrativos, planeación estratégica, calidad total y círculos de calidad, la sistematización de la firma.

Cambios en la organización y administración del proceso productivo: se trata de la introducción de cambios a la organización del proceso productivo tendientes a reducir tiempos muertos, desechos, tiempos de proceso y otros similares, todo ello con la línea de producción existente. Esto implica modificaciones en el “dibujo o diseño” de la línea de producción, pero con las mismas máquinas y equipos existentes. Por ejemplo: mejoras en la organización física (lay-out) de la planta, desverticalización del proceso productivo (outsourcing), métodos justo a tiempo (JIT), círculos de calidad. (SEI, 1996: 33)

ÁRBOLES DE REGRESIÓN Y CLASIFICACIÓN

Los CART (Breiman *et al.*, 1984) son una técnica estadística de clasificación no supervisada, considerada también una técnica de minería de datos. Los CART son un método no paramétrico de segmentación binaria, donde el árbol es construido dividiendo repetidamente los datos. En cada división, los datos son divididos en dos grupos mutuamente excluyentes. El nodo inicial es llamado *nodo raíz* o *grupo madre* y se divide en dos grupos hijos; luego, el procedimiento de partición se aplica a cada grupo por separado. Las divisiones se seleccionan de modo que la impureza de los hijos sea menor que la del grupo madre, y estas están definidas por un valor de una variable explicativa (Deconinck *et al.*, 2006). El objetivo es partir la respuesta en grupos homogéneos y, a la vez, mantener el árbol razonablemente pequeño. El análisis de los CART generalmente consiste en tres pasos (Timofeev, 2004):

1. Construcción del árbol máximo.
2. Poda del árbol, lo que resulta en una serie de árboles menos complejos derivados del árbol máximo.
3. Selección del árbol óptimo por validación cruzada (*cross-validation*). Existen varias medidas de impureza (criterios de división) que permiten analizar varios tipos de respuesta. Las medidas más comunes para árboles de clasificación, de acuerdo con Breiman *et al.* (1984), son:

- Índice de Gini: $g(t) = \sum_{j \neq i} p(j|t)p(i|t)$, donde i y j son categorías de la variable. El criterio de partición de s en el nodo t se define como: $\phi(s, t) = g(t) - p_L g(t_L) - p_R g(t_R)$, donde p_L es la proporción de casos enviados al nodo hijo del lado izquierdo y p_R es la proporción de casos enviados al lado derecho. Esta es la medida de impureza utilizada en el estudio pues, en general, se trabaja con variables categóricas.
- Índice de información o entropía: $\Delta_i = -\sum_{j=1}^k p_j(t) \ln p_j(t)$, donde $j=1, \dots, k$ es el número de clases de la variable categórica y

$p_j(t)$ es la probabilidad de clasificación correcta de la clase j en el nodo t .

Los dos índices anteriores se usan para una variable de respuesta categórica; además, está el índice de Twoing, muy similar al de Gini, que produce árboles más balanceados, pero es un algoritmo más lento, razón por la cual no se utiliza con frecuencia. Para árboles de regresión (cuando la variable respuesta es continua) se utilizan dos medidas de impureza: desviación por mínimos cuadrados, que es similar a los mínimos cuadrados en regresión, y suma de las desviaciones absolutas alrededor de la media.

El segundo paso consiste en podar el árbol, pues al no tener un criterio de parada para la división de los nodos, el modelo es sobreajustado y la información puede complicarse sin necesidad. Para ello se utiliza una función de costo-complejidad, como sigue:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}|$$

Donde $R(T)$ es el promedio de la suma de cuadrados entre los nodos, que puede ser la tasa de mala clasificación total o la suma de cuadrados de residuales total, dependiendo del tipo de árbol; $|\tilde{T}|$ es la complejidad del árbol, definida como el número total de nodos del sub-árbol, y α es el parámetro de complejidad. El parámetro $\alpha \geq 0$ es un número real. De esta manera se inicia con el árbol más grande y se termina con un árbol que contiene únicamente el nodo raíz.

Por último, se selecciona el árbol óptimo mediante un procedimiento de validación cruzada de partición en V . En general, $V=10$, así:

- Se divide la muestra en diez grupos mutuamente excluyentes y de aproximadamente igual tamaño.
- Se saca un conjunto por vez y se construye el árbol con los datos de los grupos restantes. El árbol se usa para predecir la respuesta del conjunto eliminado.
- Se calcula el error estimado para cada subconjunto.

- Se repiten los pasos 2 y 3 para cada tamaño de árbol.
- Se selecciona el árbol con la menor tasa de error estimado.

Al llegar a este punto se analiza el árbol obtenido.

REGRESIÓN LOGÍSTICA

Cuando se desea clasificar a un sujeto dentro de uno o más grupos previamente determinados a partir de un conjunto de características observadas del sujeto, es razonable pensar en usar una medida probabilística, en este caso el modelo de regresión logística puede considerarse una fórmula para calcular la probabilidad de pertenencia a uno de los grupos.

Además, la interpretación del resultado de la aplicación de esta metodología es sencilla, por tratarse en términos de probabilidad. El modelo de regresión logística se formula matemáticamente como se muestra en la ecuación siguiente (Press y Wilson, 1978):

$$P(x) = P(E | x) = \frac{1}{1 + e^{-\alpha - \beta^t x}}$$

Donde α y β son parámetros desconocidos que se estiman de los datos. Este modelo puede usarse para clasificar un objeto en una de dos poblaciones, cuando E sea el evento de que el objeto pertenezca a la primera población y x denote un vector de atributos del objeto que será clasificado. Una medida útil para verificar la calidad en las clasificaciones obtenidas por el modelo puede ser la tasa de mala clasificación, que es la proporción de observaciones mal clasificadas.

REGRESIÓN LOGÍSTICA MULTINOMIAL

Esta técnica se utiliza cuando se tiene más de una variable regresora y, por lo menos, una es de tipo cuantitativo. La técnica de regresión logística multinomial consiste en estimar la probabilidad de que una observación x pertenezca a uno de los grupos, dados valores de las p variables que la

conforman. El modelo compara $G-1$ categorías contra una categoría de referencia. Dadas n observaciones (y_i, x_i) , donde x_i es un vector con p variables y y_i es una variable aleatoria independiente multinomial con valores $1, 2, \dots, G$, la cual indica el grupo al cual pertenece cada observación. La probabilidad condicional de pertenencia de x_i a cada grupo está dada por:

$$P(y = j | x_i) = \frac{e^{\alpha_{ij} + \beta_{1j}^T x_i}}{1 + \sum_{k=2}^G e^{\alpha_{1k} + \beta_{1k}^T x_i}}$$

Donde $\alpha_{11} = \beta_{11} = 0$. Para clasificar la observación p -variada, en un grupo, se calcula la probabilidad de pertenencia a cada uno de los G grupos y se asigna la mayor probabilidad.

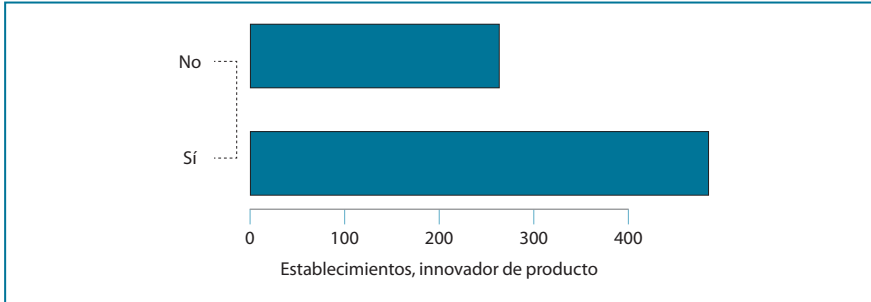
APLICACIÓN DE LAS TÉCNICAS

VARIABLES

Para construir los CART se utilizaron las siguientes variables de la Primera Encuesta de Innovación:

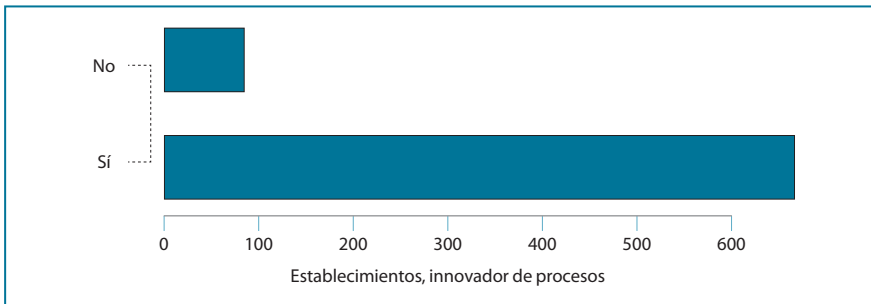
- Colocación de nuevos productos en el mercado: en el Gráfico 1 se observa que aproximadamente el 54,7% de los establecimientos encuestados puso nuevos productos en el mercado; aparecen 747 establecimientos, pues 138 (15,6%) no respondió o no conocía la respuesta a la pregunta.
- Mejora tecnológica de procesos: en el Gráfico 2 se observa que el 88,5% de los establecimientos realizó mejora tecnológica de proceso.
- Implementación de cambios en la gestión y administración del negocio: en el Gráfico 3 se observa que el 74% de los establecimientos encuestados realizó cambio organizacional.

Gráfico 1
Colocación de nuevos productos en el mercado



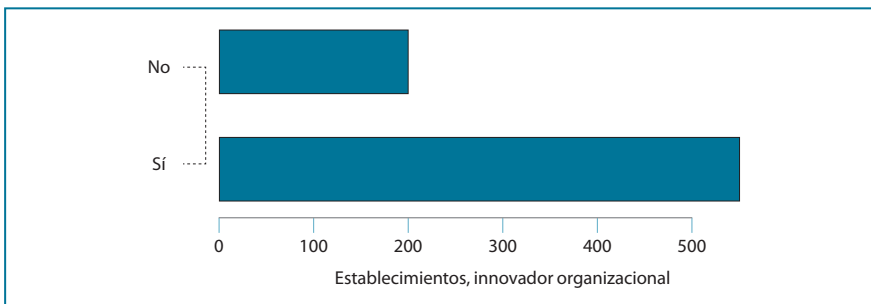
Fuente: elaboración propia a partir de datos de la EDT.

Gráfico 2
Mejora tecnológica de procesos



Fuente: elaboración propia a partir de datos de la EDT.

Gráfico 3
Implementación de cambio organizacional



Fuente: elaboración propia a partir de datos de la EDT.

Las variables seleccionadas para la caracterización son las siguientes:

1. Fuentes internas de la innovación. Cinco variables binarias donde Sí=1, No=2 (Cuadro 1).

Cuadro 1

Fuentes internas de la innovación

	Variable
1	Directivos del establecimiento
2	Personal de producción
3	Actividad permanente y organizada de I+D
4	Actividad no rutinaria de estudio técnico y de métodos en el establecimiento
5	Círculo de calidad

Fuente: elaboración propia.

2. Fuentes externas de la innovación. Diez variables binarias donde Sí=1, No=2 (Cuadro 2).

Cuadro 2

Fuentes externas de innovación

	Variable
1	Clientes
2	Proveedores
3	Universidades
4	Institutos de investigación o centros de desarrollo tecnológico
5	Consulta de bancos de datos, patentes, modelos y publicaciones técnicas
6	Ferias, exposiciones y conferencias
7	Contratación de consultores, expertos o investigadores independientes
8	Adquisición de patentes, licencias, <i>know-how</i>
9	Ingeniería reversa
10	Cursos de entrenamiento y capacitación del personal

Fuente: elaboración propia.

3. Código del tamaño del establecimiento.
4. Código Industrial Internacional Uniforme a dos/tres dígitos (CIIU3 [Rev. 2]).
5. Naturaleza jurídica del establecimiento.
6. Inversión bruta en maquinaria y equipo sobre las ventas del establecimiento en 1995.
7. Inversión bruta total sobre las ventas del establecimiento en 1995.
8. Valor de las exportaciones sobre las ventas del establecimiento en 1995.
9. Valor de las utilidades (o pérdidas) sobre las ventas del establecimiento en 1995.
10. Porcentaje de empleo calificado en producción (profesionales y posgraduados).
11. Encargados de la ejecución de las actividades innovadoras dentro de la empresa. Tres variables binarias donde Sí=1, No=2 (Cuadro 3).

Cuadro 3

Encargados de la ejecución de actividades innovadoras en el establecimiento

	Variable
1	Grupo normal de trabajo en el área de producción
2	Grupo de trabajo creado para la solución de un problema específico
3	Unidad o departamento de investigación y desarrollo

Fuente: elaboración propia.

12. Encargados de la ejecución de las actividades innovadoras en el exterior de la empresa. Siete variables binarias donde Sí=1, No=2 (Cuadro 4).

Cuadro 4

Encargados de la ejecución de actividades innovadoras en el exterior del establecimiento

	Variable
1	Otro establecimiento de la empresa o del grupo al cual pertenece el establecimiento
2	Investigador independiente

Continúa

	Variable
3	Universidades o centros de investigación
4	Centros de desarrollo tecnológico
5	Empresa privada o firma consultora
6	Conjuntamente con proveedores
7	Conjuntamente con clientes

Fuente: elaboración propia.

En total, se consideran 33 variables para la caracterización de las empresas innovadoras.

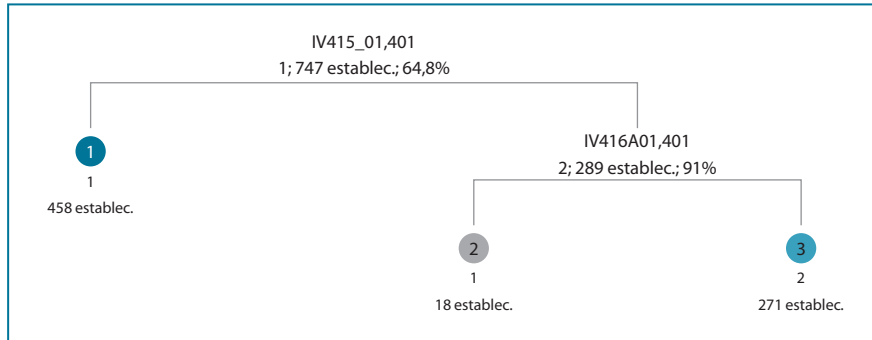
CART PARA EMPRESAS INNOVADORAS DE PRODUCTO

En el Gráfico 4 se muestra el árbol de clasificación obtenido a partir de las 33 variables mencionadas. El árbol clasifica los establecimientos en establecimientos innovadores de producto o no, utilizando el paquete estadístico R (*R Development Core Team*). Se puede notar que, según esta clasificación, un establecimiento es innovador de producto cuando:

- *La fuente de la innovación son los directivos del establecimiento.* Si la fuente interna de la innovación son los directivos del establecimiento, se puede afirmar que el establecimiento es innovador de producto; en este grupo se ubicaron 458 de los 747 establecimientos de la muestra, de los cuales ninguno quedó mal clasificado.
- *La fuente de la innovación son los clientes del establecimiento.* Para el caso en que la fuente de la innovación no son los directivos del establecimiento, es necesario verificar si la fuente externa de la innovación son los clientes, de modo que si la fuente de la innovación proviene de los clientes, entonces se tienen establecimientos innovadores de producto. Si la fuente de la innovación no son los directivos del establecimiento y tampoco son los clientes, se tienen establecimientos no innovadores.

El árbol del Gráfico 4 es bastante sencillo, por lo cual se buscará caracterizar estos grupos, es decir, buscar variables que interesa saber si son influyentes o no en los establecimientos innovadores de producto.

Gráfico 4

Árbol de clasificación para la innovación de producto

Fuente: elaboración propia.

Se realizaron pruebas adicionales sobre algunas variables que pueden ser determinantes en la clasificación de un establecimiento como innovador de producto o no. Para verificar este comportamiento se probó la independencia entre nodos del árbol e igualdad de proporciones entre las observaciones de los nodos del árbol, para cada variable de interés, así:

Primera variable. *La fuente interna de la innovación es la actividad permanente y organizada de I+D.* El Cuadro 5 presenta la proporción de establecimientos cuya fuente de innovación es I+D. Se desea verificar la hipótesis H_0 : las proporciones, probabilidad de éxito, en los grupos 1 y 2 es la misma (donde Grupo 1 corresponde a las observaciones del lado izquierdo del árbol, y Grupo 2, a las observaciones de los dos nodos del lado derecho). Para esta hipótesis tenemos un *valor* $p < 2,2 \cdot 10^{-16}$, luego se rechaza H_0 y se concluye que la proporción de respuestas afirmativas en el Grupo 1 (innovadores de producto) es mayor que la proporción de respuestas afirmativas en el Grupo 2, luego los establecimientos cuya fuente de innovación son los directivos también tienen como fuente de innovación la actividad continua y organizada de I+D.

Cuadro 5

Proporción de establecimientos cuya fuente de innovación es I+D

I+D Grupo 1 (%)	I+D Grupo 2 (%)
38	3,4

Fuente: elaboración propia.

Podemos utilizar también una prueba de independencia de los nodos del árbol usando el Cuadro 6.

Cuadro 6

Establecimientos cuya fuente de innovación es I+D

Grupo	Sí	No	Total
1	176	282	458
2	10	279	289

Fuente: elaboración propia.

Para la prueba *chi-cuadrado* (χ^2) se obtiene un *valor p* = 0, luego se rechaza H_0 ; hay independencia entre el grupo y la actividad permanente y organizada de I+D como fuente de innovación. Se concluye que hay asociación entre dichas variables. Luego, la actividad permanente y organizada de I+D como fuente interna de la innovación sí determina los grupos.

Ahora, según los resultados de las dos pruebas anteriores, se puede afirmar que los establecimientos innovadores de producto tienen como fuentes internas de innovación los directivos del establecimiento y la actividad permanente y organizada de I+D.

Segunda variable. *El encargado de ejecutar actividades innovadoras es el departamento de I+D*. Aquí se obtiene para la prueba de comparación de proporciones un *valor p* < 0,05, luego tenemos que hay una diferencia significativa entre la proporción de establecimientos donde las actividades innovadoras están a cargo del departamento de I+D, en el Grupo 1 y en el Grupo 2, y que esta proporción es mayor para el Grupo 1.

Además, según la prueba de independencia, hay asociación entre los grupos y la respuesta, luego la variable sí influye en la determinación de los grupos. Por lo tanto, se puede afirmar que los establecimientos innovadores de producto, además de tener como fuente interna de innovación

a los directivos del establecimiento, el encargado de las actividades innovadoras dentro del mismo es el departamento de I+D.

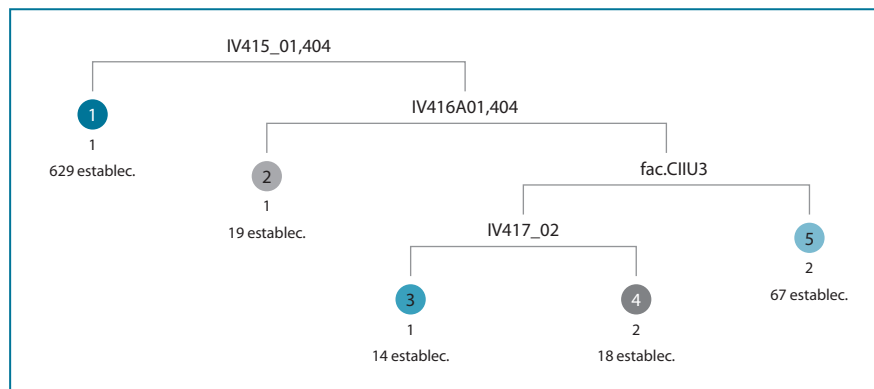
Tercera variable. *El establecimiento desarrolló proyectos de I+D entre 1989 y 1996.* Nuevamente aplicamos la prueba de igualdad de proporciones y la prueba de independencia a los datos. Los resultados indican que la proporción de establecimientos que desarrolló proyectos de I+D es estadísticamente mayor en el grupo de establecimientos innovadores de producto; además, hay asociación entre los grupos y la realización de proyectos de I+D. Luego podemos afirmar que los establecimientos innovadores de producto, aparte de tener los directivos como fuente interna de la innovación, también desarrollaron proyectos de I+D en los años previos a la encuesta.

CART PARA EMPRESAS INNOVADORAS DE PROCESO

En el Gráfico 5 se observa el árbol obtenido, donde se identifican cuatro variables que caracterizan a las empresas innovadoras de proceso:

Gráfico 5

Árbol de clasificación para la innovación de proceso



Fuente: elaboración propia.

- *La fuente de la innovación son los directivos o los clientes del establecimiento.* Si la fuente interna de la innovación son los directivos

del establecimiento, este último es innovador de proceso; de lo contrario, la fuente de la innovación son los clientes. Pero si la innovación no proviene de los clientes o los directivos entonces,

- *El encargado de la ejecución de actividades innovadoras es un grupo de trabajo creado para solucionar un problema específico.* Si el establecimiento pertenece al sector bebidas, textiles, papel o derivados, imprentas y editoriales, productos químicos, metales no ferrosos, maquinaria y aparatos eléctricos o equipo profesional y científico, y además la innovación es llevada a cabo por un grupo específico de trabajo, entonces hay innovación de proceso; de lo contrario, el establecimiento no es innovador.

Las variables complementarias son: el encargado de la ejecución de las actividades innovadoras dentro del establecimiento, esto es, la unidad o departamento de I+D; el establecimiento ha desarrollado proyectos de I+D, y la fuente interna de la innovación es la actividad permanente y organizada de I+D. Para las tres variables mencionadas, utilizando una prueba de comparación de proporciones, se rechaza la hipótesis de no asociación entre la innovación de proceso y la variable; luego se puede afirmar que hay asociación entre la innovación de proceso y las tres variables complementarias.

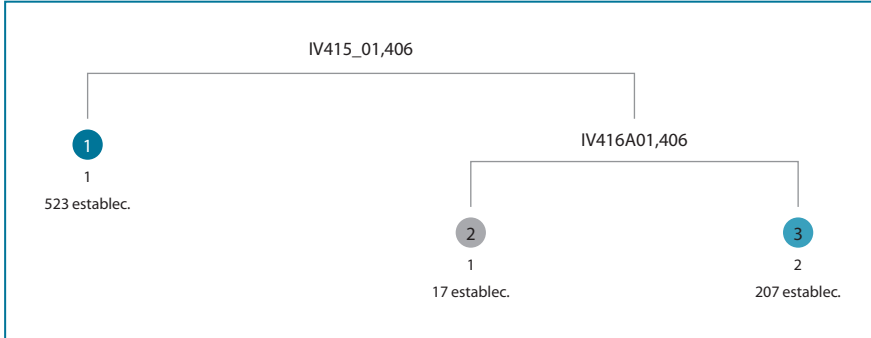
CART PARA EMPRESAS INNOVADORAS ORGANIZACIONALES

Por último, se muestra en el Gráfico 6 la clasificación para los establecimientos que han implementado cambios en la gestión y administración del negocio. Nuevamente, la innovación proviene de los directivos del establecimiento; de lo contrario, de los clientes.

Las variables complementarias son las mismas mencionadas en la sección “Variables complementarias: I+D”. Nuevamente se realiza una prueba de comparación de proporciones y una prueba de independencia de las cuales se obtiene que la innovación organizacional está relacionada con las variables complementarias.

Gráfico 6

Árbol de clasificación para la innovación organizacional



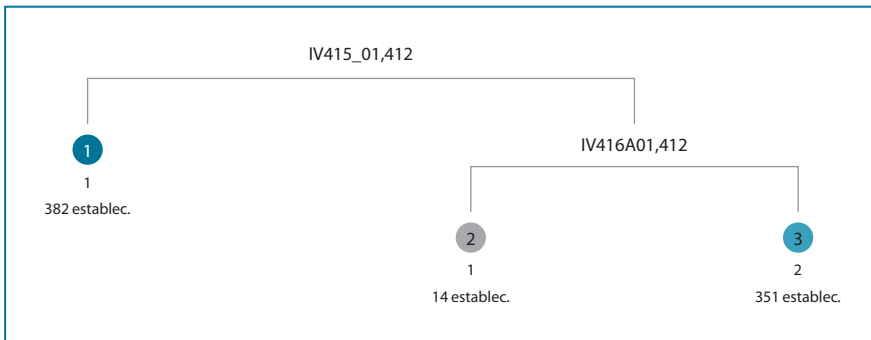
Fuente: elaboración propia.

CART PARA EMPRESAS QUE REALIZAN CAMBIOS EN EMPAQUE Y EMBALAJE

En este punto se obtiene el mismo árbol de clasificación que para el caso anterior y, nuevamente, al verificar para las variables complementarias, se cumple lo mismo. En el Gráfico 7 se observa la misma estructura de árbol; luego se concluye: para que cualquier establecimiento sea innovador, la fuente de la innovación ha de provenir de los directivos o los clientes del establecimiento.

Gráfico 7

Árbol de clasificación para cambios en empaque y embalaje



Fuente: elaboración propia.

REGRESIÓN LOGÍSTICA

El modelo de regresión logística es uno de los más aplicados, pero uno de sus mayores problemas es el de la separación, cuya consecuencia es la no existencia de los estimadores de máxima verosimilitud, pues el proceso iterativo para obtenerlos no converge y, por lo tanto, no se pueden realizar inferencias.

Al aplicar la regresión logística para el caso de las variables seleccionadas en la EDT se presenta este problema, conocido como *separación completa* o *separación cuasi completa* (Prieto Castellanos, 2005), y el modelo de regresión logística no converge; luego no se pueden obtener resultados para ninguno de los casos presentados anteriormente.

CLASIFICACIÓN VÍA COMPONENTES PRINCIPALES

Se analizaron los componentes principales con las variables relacionadas con la innovación, a fin de obtener una clasificación con una variable de innovación menos desagregada, un indicador de innovación. Se tomaron las variables: innovación de producto (IV401), innovación de proceso (IV404), innovación de gestión y administración organizacional (IV406), innovación de administración del proceso productivo (IV408) y, por último, innovación en empaque y embalaje (IV412). En el Cuadro 7 se observa que para explicar el 73% de la variabilidad de los datos se requieren tres componentes.

Cuadro 7

Componentes principales, indicador innovación

Valores propios					
	1,88627	0,9130619	0,8610286	0,743679	0,5959621
Vectores propios					
	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
IV401	-0,4238148	0,58483764	-0,01299091	0,6570438	-0,21557025

Continúa

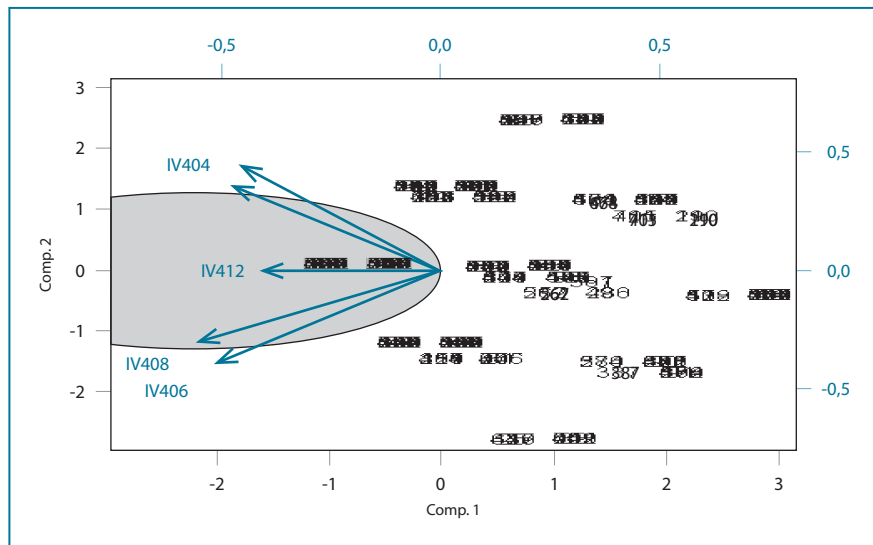
IV404	-0,4386165	0,47160784	-0,34313883	-0,6049861	0,31851084
IV406	-0,4751968	-0,52217255	-0,17655343	0,3485637	0,59064041
IV408	-0,5103776	-0,40358924	-0,19492784	-0,2050321	-0,70469523
IV412	-0,3764296	-0,00159446	0,90162021	-0,1968512	0,08141772
Importancia de los componentes					
	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Desviación estándar	1	0,9555427	0,9279163	0,8623682	0,7719858
Proporción de varianza	0,3772537	0,1826124	0,1722057	0,1487358	0,1191924
Proporción acumulada	0,3772537	0,5598661	0,7320718	0,8808076	1

Fuente: elaboración propia.

El primer componente presenta coeficientes altos en todos los tipos de innovación (valores del vector propio), luego es una medida de la *innovación global*, donde los tipos de innovación con mayor peso para los establecimientos son en administración del proceso productivo (IV408) y gestión y administración (IV406); el tipo de innovación con menor impacto sobre los establecimientos es en empaque y embalaje (IV412).

El segundo componente presenta dos valores positivos altos en la innovación de producto y proceso y dos valores negativos altos en innovación de la gestión y administración y administración del proceso productivo. Este componente parece medir el predominio de la innovación tecnológica de productos y procesos sobre la innovación organizacional. En el Gráfico 8 se observa un contraste en el tipo de innovación; la región gris contiene los establecimientos donde la innovación es ideal, pues se encuentra en equilibrio, con valores altos en ambos componentes (por la construcción de los componentes, entre más negativo, más innovador).

Gráfico 8
Biplot, primera y segunda componentes

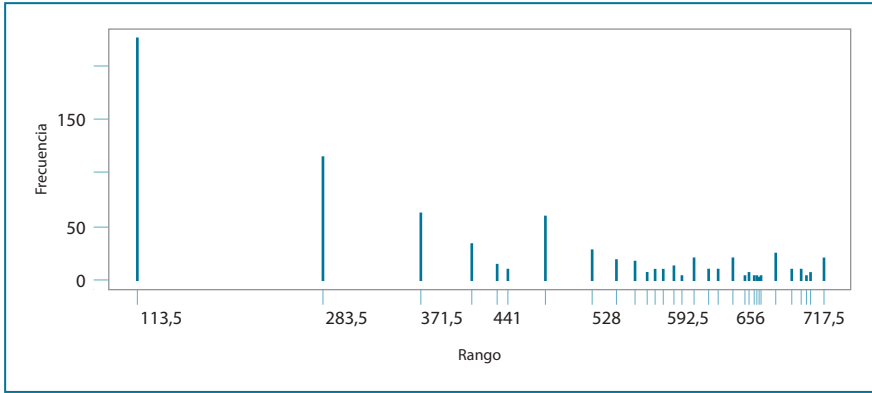


Fuente: elaboración propia.

CLASIFICACIÓN CON EL INDICADOR DE INNOVACIÓN GLOBAL (PRIMER COMPONENTE)

En el Gráfico 9 se observan 31 diferentes rangos para el primer componente. A medida que el rango se hace más bajo, mejor puntuación se tiene en el índice de innovación global. Del rango 620 en adelante, el establecimiento sólo cumple con dos de los cinco tipos de innovación, de esta manera se considerará que a partir de este rango la innovación en el establecimiento no es buena.

Gráfico 9
Rangos de los puntajes para el primer componente



Fuente: elaboración propia.

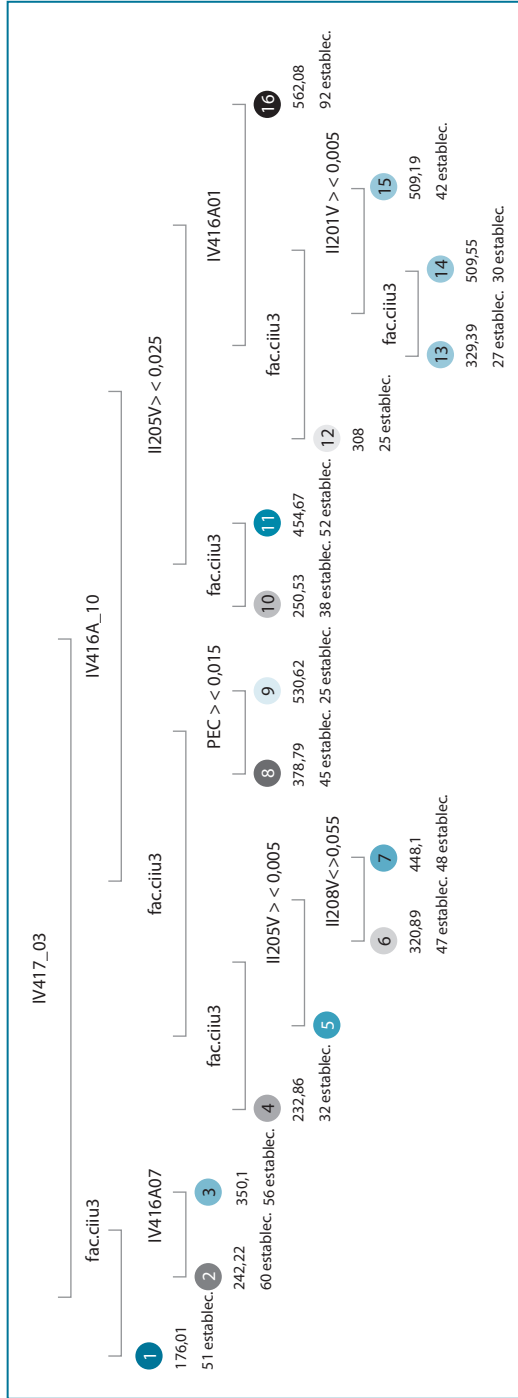
En el Gráfico 10 se presenta un árbol de regresión, pues la nueva variable respuesta es una variable continua dada por el rango del puntaje del primer componente. Se puede observar que el árbol es más complejo que los encontrados cuando se utiliza cada tipo de innovación por separado. Puesto que se sabe que a partir del rango 620 el establecimiento sólo cumple con dos de los tipos de innovación y el rango máximo que aparece en el árbol del Gráfico 10 es 562,08 (promedio de los datos del Cuadro 8), se procede a podar el árbol para obtener una clasificación más general.

Cuadro 8
Rangos del último grupo del Gráfico 10

Rango	113,5	283,5	371,5	418,5	451,5	485	528	550	566,5	577,5	584,5	592,5	602,5
Frecuencia	6	3	4	2	1	11	4	5	1	1	3	2	5
Rango	620	633,5	642	656	666,5	670,5	675	678	694	709,5	717,5	726	738
Frecuencia	7	1	2	2	2	3	1	1	9	4	4	1	7

Fuente: elaboración propia.

Gráfico 10
Clasificación de los establecimientos para el índice de innovación



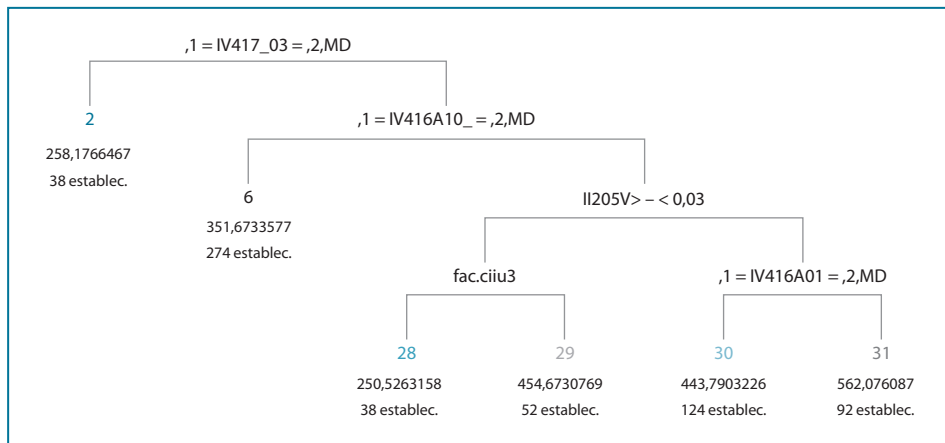
Fuente: elaboración propia.

En el Gráfico 11 se observa la clasificación luego de podar el árbol de clasificación presentado en el Gráfico 10, considerando un parámetro de complejidad $\alpha=0,02$. Las variables que determinan la clasificación son:

- IV417_03: el encargado de la ejecución de las actividades innovadoras en el interior de la empresa es la unidad o departamento de I+D.
- IV416A10: la fuente externa de la innovación está dada por cursos de entrenamiento y capacitación del personal.
- II205V: valor de las exportaciones sobre las ventas del establecimiento en 1995.
- CIU3: Código Industrial Internacional Uniforme a 3 dígitos.
- IV416A01: la fuente externa de la innovación son los clientes.

Gráfico 11

Clasificación de los establecimientos para el índice de innovación después de podar el árbol



Fuente: elaboración propia.

ESTABLECIMIENTOS INNOVADORES

Tomando en consideración el hecho de que entre más bajos sean los puntajes más innovador es el establecimiento, en el Gráfico 11 se tomaron como

establecimientos innovadores hasta el grupo etiquetado con el número 28, donde el rango del índice de innovación es 250,5263 (el promedio de los rangos de los puntajes de 38 establecimientos). Las reglas de clasificación obtenidas para establecimientos innovadores son:

1. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa es la unidad o departamento de I+D.
2. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa NO es la unidad o departamento de I+D; pero la fuente externa de la innovación está constituida por los cursos de entrenamiento y capacitación del personal.
3. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa NO es la unidad o departamento de I+D, la fuente externa de la innovación NO está constituida por cursos de entrenamiento y capacitación del personal, el valor de las exportaciones sobre las ventas del establecimiento en 1995 es mayor o igual a 0,025 y pertenece a alguno de los sectores industriales del Cuadro 9.

Cuadro 9

CIIU3 establecimientos innovadores, indicador

CIIU3	Sector
321	Fabricación de textiles
324	Fabricación de calzado y sus partes, excepto el de caucho o plástico
332	Fabricación de muebles y accesorios, excepto los que son principalmente metálicos
342	Imprentas, editoriales e industrias conexas
351	Fabricación de sustancias químicas industriales
352	Fabricación de otros productos químicos
353	Refinerías de petróleo
362	Fabricación de vidrio y productos de vidrio
369	Fabricación de otros productos minerales no metálicos
381	Fabricación de productos metálicos, excepto maquinaria y equipos

Continúa

CIU3	Sector
385	Fabricación de material profesional y científico, instrumentos de medida y control n.e.p., aparatos fotográficos e instrumentos de óptica
390	Otras industrias manufactureras

Fuente: elaboración propia.

ESTABLECIMIENTOS NO INNOVADORES

Tomado a partir del grupo etiquetado con el número 29 en el Gráfico 11. Las reglas aquí determinadas son:

1. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa NO es la unidad o departamento de I+D, la fuente externa de la innovación no está constituida por cursos de entrenamiento y capacitación del personal, el valor de las exportaciones sobre las ventas del establecimiento en 1995 es mayor o igual a 0,025 y pertenece a alguno de los sectores industriales del Cuadro 10.

Cuadro 10

CIU3 establecimientos no innovadores, indicador

CIU3	Sector
311	Fabricación de productos alimenticios, excepto bebidas
322	Fabricación de prendas de vestir, excepto calzado
323	Industrias del cuero y productos del cuero y sucedáneos del cuero y pieles, excepto el calzado y otras prendas de vestir
331	Industria de la madera y productos de la madera y el corcho, excepto muebles
341	Fabricación de papel y productos de papel
354	Fabricación de productos derivados del petróleo y del carbón
355	Fabricación de productos de caucho
356	Fabricación de productos de plástico
371	Industrias básicas de hierro y acero

Continúa

CIU3	Sector
382	Construcción de maquinaria, exceptuando la eléctrica
383	Fabricación de maquinaria, aparatos, accesorios y suministros eléctricos
384	Construcción de equipo y material de transporte

Fuente: elaboración propia.

2. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa NO es la unidad o departamento de I+D, la fuente externa de la innovación NO está constituida por cursos de entrenamiento y capacitación del personal, el valor de las exportaciones sobre las ventas del establecimiento en 1995 es menor de 0,025 y la fuente externa de la innovación está dada por los clientes.
3. El encargado de la ejecución de las actividades innovadoras en el interior de la empresa NO es la unidad o departamento de I+D, la fuente externa de la innovación NO está constituida por cursos de entrenamiento y capacitación del personal, el valor de las exportaciones sobre las ventas del establecimiento en 1995 es menor de 0,025 y la fuente externa de la innovación NO está dada por los clientes.

Según la información anterior, las variables determinantes a la hora de definir si un establecimiento es innovador o no son el sector industrial manufacturero al que pertenece y el valor de las exportaciones sobre las ventas del establecimiento.

CONCLUSIONES

Las conclusiones del trabajo tienen dos dimensiones, una metodológica y otra temática. Metodológicamente, respecto a la comparación entre CART y regresión logística para caracterizar a las empresas innovadoras desde los datos de la EDT, los resultados exponen claramente una ventaja de los CART sobre la regresión logística, pues la aplicación de esta última no lleva a resultados adecuados con la estructura de los datos disponibles.

Por otra parte, la aplicación experimental de los CART arroja resultados interesantes, que ilustran el potencial de esta técnica. En este sentido, las características de las empresas innovadoras y no innovadoras se examinan respecto a 33 variables de la EDT, utilizando un doble procedimiento: (a) considerando los tipos específicos de innovación (de producto, de proceso, de gestión y administración) y (b) considerando un indicador global de innovación construido vía análisis de los componentes principales. Las conclusiones relativas a ambos procedimientos se exponen a continuación.

Para empezar, las empresas innovadoras de producto resultan ser aquellas en las cuales la fuente de la innovación es interna y proviene de los directivos del establecimiento; en este grupo se ubicaron 458 de los 747 establecimientos de la muestra, de los cuales ninguno quedó mal clasificado. Esta es una característica que podría explicarse por el carácter informal y no sistemático de la innovación en las empresas de la muestra —aunque el análisis complementario de las variables caracterizadoras también identifica positivamente un conjunto de empresas que pese a ser minoritario es significativo—, las cuales utilizan como fuente interna de innovación la actividad permanente y organizada de I+D, ejecutan las actividades innovadoras en un departamento de I+D y han desarrollado proyectos de I+D en los años previos a la Encuesta.

Unas pocas empresas adicionales (18) no incluidas en el grupo anterior se suman al grupo de empresas innovadoras de producto, cuya característica representativa es que la fuente de la innovación es externa y proviene de los clientes. Aquí podría observarse un carácter reactivo de la innovación a las presiones de la competencia. Las demás empresas (271) no son innovadoras de producto.

Las empresas innovadoras de proceso se caracterizan por que su fuente de innovación son los directivos o los clientes. Aquí se sitúa la gran mayoría de empresas de la muestra, lo cual revela una mayor frecuencia de la innovación de proceso frente a la de producto, pero unas características comunes, asociadas de nuevo a la dependencia de la innovación frente a la iniciativa de los directivos o a la reacción a circunstancias inducidas por la competencia.

El análisis de las empresas innovadoras organizacionales y las que realizan cambios en el empaque y embalaje revela características similares a las de las empresas innovadoras de proceso.

Cuando se utiliza el indicador global de innovación, se obtienen las características de las empresas innovadoras (479 empresas según la definición adoptada) frente a las no innovadoras. En este caso, es interesante observar que las variables que definen a las empresas innovadoras son su desempeño exportador (dado por un mínimo del valor de las exportaciones sobre las ventas del establecimiento) y su pertenencia a un determinado grupo de sectores industriales, así: un buen desempeño exportador es suficiente para que las empresas de ciertos sectores sean innovadoras; en cambio, no es suficiente para empresas de otros sectores, que deben cumplir con otras condiciones para ser innovadoras —el encargado de la ejecución de las actividades innovadoras en la empresa es la unidad o departamento de I+D o la fuente externa de la innovación está dada por los cursos de entrenamiento y capacitación del personal). En cambio, las empresas de bajo desempeño exportador terminan siendo todas no innovadoras.

Esta última conclusión fortalece la asociación entre desempeño exportador e innovación, pero deja en discusión la naturaleza de la distinción entre los grupos de sectores para los cuales un determinado desempeño exportador no necesariamente está asociado a la innovación.

Finalmente, es importante señalar algunas limitaciones del trabajo, que quedan para resolver en un desarrollo futuro. En primer lugar, es necesario revisar y justificar cuidadosamente, frente a un análisis de la literatura pertinente, la selección de las variables incluidas en el análisis. Además, es muy importante buscar alternativas de definición de un indicador global de innovación, a partir de un análisis crítico del indicador aquí utilizado de manera exploratoria. Finalmente, queda el problema de la clasificación de los sectores industriales, pues la CIIU no parece la clasificación más adecuada para las dinámicas aquí examinadas. Quizás una exploración adicional de clasificaciones alternativas pueda arrojar luces sobre la relación sector industrial-innovación-desempeño exportador, que aquí no resulta tan clara.

REFERENCIAS

- Breiman, L.; Friedman, J.; Olshen, R. y Stone, C. (1984). *Classification and regression trees*. Boca Ratón, FL: Wadsworth International Group.
- Deconinck, E.; Zhang, M. H.; Coomans, D. y Heyden, Y. V. (2006). Classification tree models for the prediction of blood-brain barrier passage of drugs. *Journal of Quematical Information and Modeling*, 46 (3), 1410-1419.
- Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Godin, B. (2002). *The rise of innovation surveys: Measuring a fuzzy contest*. Montreal: Project on the History and Sociology of STI Statistics, CSIIC.
- Grinstein, A. y Goldman, A. (2006). Characterizing the technology firm: An exploratory study. *Research Policy*, 35 (1), 121-143.
- Kannebley Jr., S.; Porto, G. S. y Pazello, E. T. (2005). Characteristics of Brazilian innovative firms: An empirical analysis based on PINTEC-industrial research on technological innovation. *Research Policy*, 34 (6), 872-893.
- Koberg, C.; Sarason, Y. y Rosse, J. (1996). A taxonomic approach to studying high technology firms: deciphering the tower of Babel. *Journal of High Technology Management Research*, (20), 15-35.
- Observatorio de Ciencia y Tecnología (OCyT), (2001). *Resultados de innovación alcanzados por las empresas*. Recuperado el 27 de octubre de 2009, de <http://www.ocyt.org.co/ind/pfd/B2b.pdf>.
- Organisation for Economic Co-operation and Development (OCDE), (2000). *The measurement of scientific and technological activities: Proposed guidelines for collecting and interpreting technological innovation data*. Recuperado el 27 de octubre de 2009, de <http://www.oecd.org/dataoecd/35/61/2367580.pdf>.
- (2005). *Oslo Manual: guidelines for collecting and interpreting innovation* (3a ed.). Paris: OCDE Publications.
- Press, S. J. y Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73 (364), 699-705.
- Prieto Castellanos, K. A. (2005). *Regresión logística con penalidad ridge aplicada a datos de expresión genética*. Tesis de Maestría, Universidad de Puerto Rico, Mayagüez, Puerto Rico.
- R Development Core Team (s. f.). *The R project for statistical computing*. Recuperado el 1 de febrero de 2006, de <http://www.r-project.org>.
- Raveh, A. (1989). A nonmetric approach to linear discriminant analysis. *Journal of the American*, 84 (405), 176-183.
- Salazar, M. y Holbrook, J. A. (2004). A debate on innovation surveys. *Science and Public Policy*, 31 (4), 254-266.

- Salazar, M. y Vargas, M. (1998). *Encuesta sobre desarrollo tecnológico en la industria colombiana*. Bogotá: Departamento Nacional de Planeación.
- Sánchez, M. P. y Castrillo, R. (2006). La tercera edición del manual de Oslo: cambios e implicaciones. Una perspectiva de capital intelectual. *madri+d* (35), 1-16.
- Schumpeter, J. (1962). *Capitalism, Socialism and Democracy*. New York: Harper.
- Sistema Estratégico de Información (SEI), (1996). *Encuesta sobre desarrollo tecnológico en el establecimiento industrial colombiano: manual del encuestador*. Bogotá: autor.
- Suárez, O. M. (2004). Schumpeter: innovación y determinismo tecnológico. *Scientia et Technica*, 10 (25), 209-213.
- Timofeev, R. (2004). *Classification and regression trees (CART): Theory and applications*. Tesis de Maestría de la Humboldt, Berlín, Alemania.
- Vargas, M. y Malaver, F. (2004). Los avances en la medición del desarrollo tecnológico en la industria colombiana. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad*, 1 (2), 137-166.

Capítulo 11

Aplicación exploratoria de técnicas basadas en sistemas neurodifusos al estudio de la innovación a partir de encuestas realizadas en Colombia¹

*Luis Felipe Duque Álvarez
Jorge Manrique Henao
Jorge Robledo Velásquez*

Algunas técnicas provenientes de la estadística y de la inteligencia artificial pueden emplearse para transformar grandes cantidades de datos en reglas fácilmente interpretables por parte de un analista; sin embargo, estas técnicas poco se han aplicado al estudio de la innovación, utilizando los resultados de las encuestas nacionales. En el presente capítulo se explora la puesta en práctica de técnicas basadas en sistemas neurodifusos (Kasabov, 1996) al estudio de la innovación, utilizando datos de la Encuesta Anual Manufacturera de 2005 (EAM 2005). Así mismo, se presentan las principales limitaciones encontradas y se sugiere trabajo futuro

¹ Resultados preliminares de este trabajo pueden consultarse en Duque, Manrique y Robledo (2008).

para explorar posibles soluciones a tales limitaciones, a fin de mejorar y ampliar el potencial de la aplicación de tales técnicas al descubrimiento de conocimiento sobre la innovación, a partir de encuestas nacionales.

El presente capítulo se estructura de la siguiente forma: en la siguiente sección se presenta el proceso y los algoritmos constitutivos de las técnicas implementadas; luego se presentan y analizan los resultados de la aplicación experimental de las técnicas a la EAM 2005; después se plasman las propuestas de trabajo futuro, sugeridas a partir del análisis de la sección anterior, y, finalmente, se concluye respecto a la aplicación de las técnicas y se plantea una relación sucinta del trabajo futuro propuesto.

PROCESO Y ALGORITMOS IMPLEMENTADOS

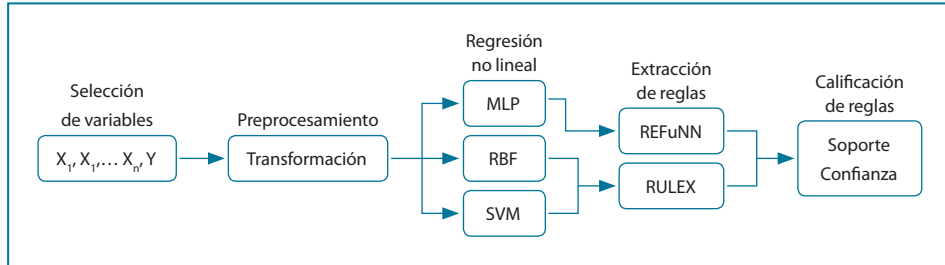
Hay tres algoritmos que sirven de base a la discusión planteada en los apartados siguientes, que se integran en el proceso esquematizado en el Gráfico 1. Tales algoritmos buscan obtener reglas difusas² a partir de los coeficientes arrojados por una regresión no lineal, de difícil interpretación. El primero emplea funciones regresoras del tipo *perceptrón multicapa*, de las que extrae reglas difusas simples empleando el procedimiento *Rule Extraction from Fuzzy Neural Networks* (REFuNN), introducido por Kasabov (1996). El segundo algoritmo parte de *funciones de base radial* para extraer reglas difusas empleando la técnica planteada por McGarry, Wermter y MacIntyre (1999). Finalmente, el tercer algoritmo es similar al segundo, pero emplea *máquinas de soporte vectorial* como funciones regresoras.

Como puede observarse en el Gráfico 1, si bien algunos de los algoritmos que se exponen son normalmente utilizados de manera independiente como técnicas distintas de la inteligencia artificial, sus similitudes permiten aquí su integración y aplicación experimental en un proceso de cinco etapas principales, a saber: (1) selección o construcción de variables, (2) preprocesamiento de variables, (3) regresión, (4) extracción de reglas y (5) calificación de reglas. Estas etapas se explican a continuación.

² Las reglas difusas se explican en el apartado “Extracción de reglas difusas”.

Gráfico 1

Proceso esquemático de aplicación de los algoritmos analizados



Fuente: elaboración propia.

SELECCIÓN O CONSTRUCCIÓN DE VARIABLES

En esta primera etapa se definen las variables dependientes y las variables independientes, de acuerdo con el sistema estudiado. En el caso del presente trabajo, estas se seleccionan a partir de la EAM 2005, teniendo en cuenta el análisis de la relación entre uso energético e innovación, planteado en Manrique y Robledo (2008).

En el Gráfico 2 pueden observarse las variables construidas para probar los algoritmos estudiados. Las variables X1, X2 y X3 están asociadas al subsistema de insumos, definido como “todos aquellos tangibles e intangibles que se requieren del entorno externo de la organización” (Manrique y Robledo, 2008: 29). Las variables X4, X5 y X6 están asociadas al subsistema de entorno interno, “conformado por los recursos propios de la empresa, como infraestructura, activos en general, capital y finanzas, personas y su capacitación, y donde a través de la adecuada combinación de estos recursos, se desarrollan las capacidades de producción, de I+D de gestión interna” (Manrique y Robledo, 2008: 29). Finalmente, la variable Y, que en el caso del presente estudio será la variable dependiente, hace parte del subsistema del entorno externo.

Gráfico 2

Variables construidas a partir de la EAM 2005

MP/CI (X1)	Total - materias primas y empaques consumidos Consumo intermedio
EEC/CI (X2)	Valor energía eléctrica (comprada-vendida) Consumo intermedio
KWH/PT (X3)	Calidad KWH consumidos Personal ocupado total
SPPP/ST (X4)	Sueldo y salarios - personal producción permanentes - Prof, téc y tec Sueldos y salarios total
M/AF (X5)	Valor maquinaria y equipo industrial Valores de activos fijos
IAF/AF (X6)	Inversión total en activos fijos Valores de activos fijos
VA/PT (Y)	Valor agregado Personal ocupado total

X(1) a X(6): variables independientes; Y: variable dependiente.

Fuente: elaboración propia.

La EAM 2005 presenta información de grupos y agrupaciones industriales de acuerdo con la Clasificación Industrial Internacional Uniforme (revisión 3), adaptada para Colombia (CIIU Rev. 3 AC). El conjunto empleado para probar los algoritmos estudiados en el presente trabajo se generó de la siguiente manera: en total se emplearon 314 subconjuntos, cada uno de los cuales tenía un área metropolitana y un grupo industrial asociado. Un ejemplo de este tipo de subconjuntos podría ser el que agrupa a las empresas dedicadas a la construcción de equipos y material de transporte (CIIU=384) ubicadas en el área metropolitana del Valle de Aburrá. A su vez, a cada uno de estos 314 subconjuntos se le asociaron las variables construidas en el Gráfico 2. De una manera similar, el conjunto de las variables independientes (X1, X2, X3, X4, X5 y X6) puede verse como un conjunto con 314 vectores, cada uno de los cuales tiene seis

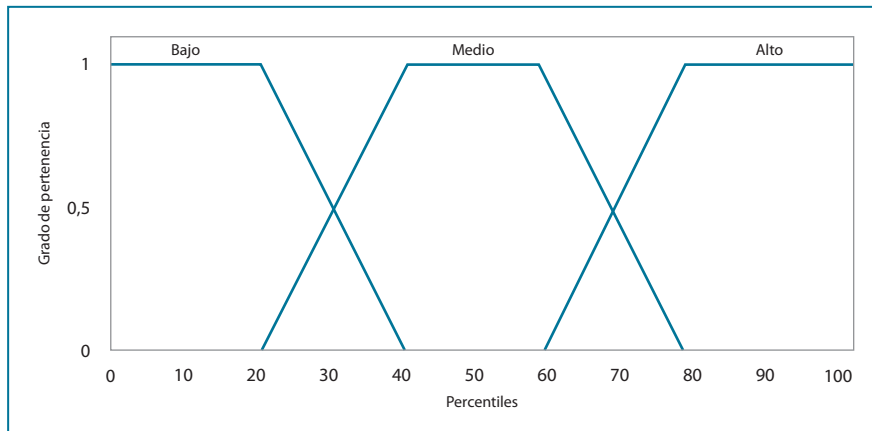
componentes, y el conjunto de variables dependientes puede verse como un vector con 314 componentes.

PREPROCESAMIENTO

En esta etapa se busca transformar el conjunto de variables seleccionadas de manera que puedan usarse de una forma apropiada en las etapas posteriores de regresión y extracción de reglas. Cuando se va a realizar un procedimiento de regresión empleando perceptrones multicapa se efectúa un emborronamiento de acuerdo con el Gráfico 3, con el fin de lograr una compatibilidad con los algoritmos de extracción de reglas a partir de este tipo de funciones (véase apartado “Extracción de reglas difusas”). En caso de usar funciones de base radial o máquinas de soporte vectorial para realizar la regresión, el procesamiento consta de una normalización de las variables de entrada con el fin de facilitar el procedimiento de aprendizaje no supervisado (véase apartado “Características de las regresiones no lineales”).

Gráfico 3

Emborronamiento usando tres funciones de membresía difusas



Fuente: Duque, Manrique y Robledo (2008: 88).

REGRESIÓN

A partir de los valores a los que se llega tras preprocesar las variables de interés, se llevan a cabo regresiones no lineales empleando familias de funciones del tipo perceptrones multicapa (ecuación 1), funciones de base radial (ecuación 2) y máquinas de soporte vectorial (ecuación 3), y se obtienen los coeficientes o parámetros de la familia de funciones que minimizan alguna medida de error (comúnmente puede ser el error absoluto medio o el error cuadrático medio). Un estudio detallado de la manera como estos tipos de funciones son entrenadas se puede encontrar en Kecman (2001), Martín del Brío y Sanz (2001), Gunn (1998) y Cristianini y Shawe-Taylor (2000).

$$f_{MLP} = \text{transfer} \left(\sum_{j=1}^p w'_j * \text{transfer} \left(\sum_{k=1}^N w_{kj} x_k - \theta_j \right) - \theta_0 \right) \quad (1)$$

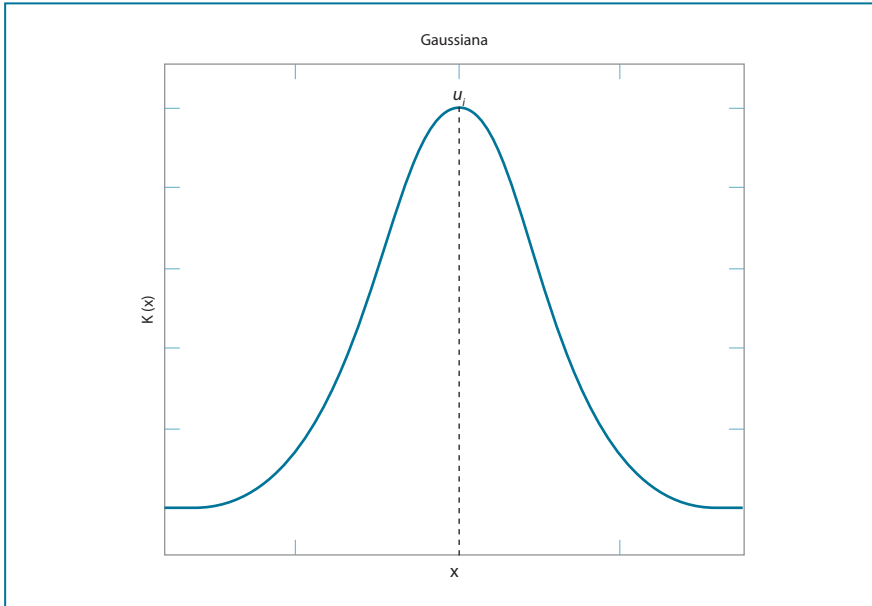
$$f_{RBF} = \sum_j w_j K(\bar{\mathbf{x}}, \bar{\mathbf{u}}_j, \sigma_j) \quad (2)$$

$$f_{SVM} = \sum_j w_j K(\bar{\mathbf{x}}, \bar{\mathbf{u}}_j, \sigma) \quad (3)$$

Como procedimiento de optimización numérica asociado a dicha minimización se pueden usar métodos numéricos basados en el descenso por el gradiente, optimización por enjambre de partículas y cuasi Newton (véase Kecman, 2001; Møller, 1993; Bishop, 1996). En las ecuaciones 1, 2 y 3, los valores w , θ , u , corresponden a parámetros de la familia de funciones que, como se explicó, deben ser hallados de tal manera que minimicen alguna medida de error determinada. La función *transfer()* es conocida como función de activación y comúnmente se emplea una función sigmoidea (Gráfico 4) para este fin. La función $K()$ es llamada función de Kernel, para la cual normalmente se usa una función gaussiana (Gráfico 5). Los valores $\bar{\mathbf{x}}$ son las variables independientes ($X_1 \dots X_6$) del Gráfico 2.

Gráfico 4

Funciones de base radial

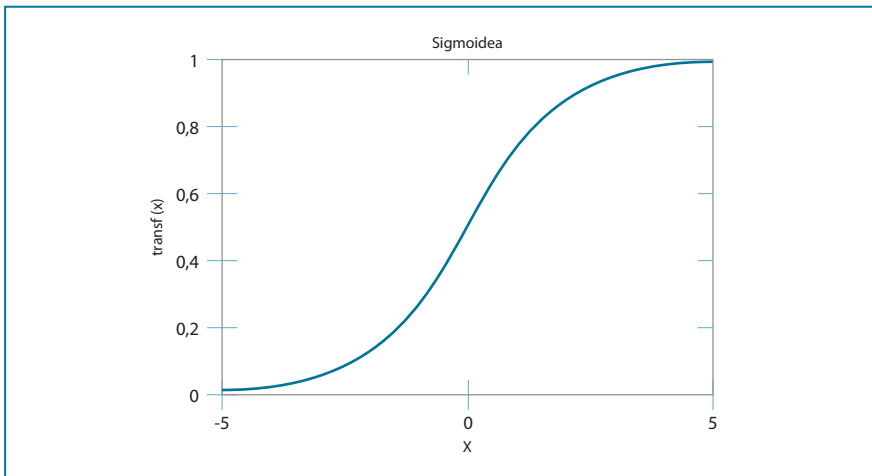


u_i : centroides de la RBF.

Fuente: Martín del Brío y Sanz (2001: 153).

Gráfico 5

Función de activación sigmoidea



Fuente: Martín del Brío y Sanz (2001: 62).

EXTRACCIÓN DE REGLAS DIFUSAS

A esta etapa del algoritmo entran las funciones y coeficientes encontrados en la etapa anterior y se obtienen reglas donde se plasma información interpretable por un analista. En el presente trabajo se exploran dos procedimientos: el REFuNN, introducido por Kasabov (1996), y la *extracción de reglas a partir de funciones de base radial*, expuesto por McGarry, Wermter y MacIntyre (1999).

El procedimiento REFuNN arroja como resultado reglas difusas simples, como la ejemplificada en la ecuación 4:

$$\text{Si } X_1 \text{ es alto y } X_2 \text{ es medio ENTONCES } Y \text{ es alto} \quad (4)$$

El procedimiento de extracción de reglas a partir de funciones de base radial que se seleccionó encuentra reglas como la expresada en la Ecuación 5.

$$\text{Si } a_1 \leq X_1 \leq b_1 \text{ y } a_2 \leq X_2 \leq b_2 \text{ ENTONCES } Y \text{ es bajo} \quad (5)$$

En el apartado “Características de la extracción de reglas” se dan ejemplos puntuales de estos tipos de reglas. En las ecuaciones 4 y 5 se observa que las reglas tienen dos partes: un antecedente, que se ubica antes del condicional *ENTONCES*, y un consecuente, que se ubica después de este. Es de anotar que, a su vez, el antecedente de las reglas puede tener varias partes enlazadas por la conjunción *y*. De esta manera, una regla genérica se podría expresar de la forma $A \rightarrow B$.

Hay que observar también que en las ecuaciones 4 y 5 podrían aparecer palabras como *alto*, *medio* o *bajo*, conocidas como conjuntos difusos. En el caso del presente trabajo, se dice que un valor pertenece al conjunto difuso alto cuando es mayor que el cuarto quintil; al conjunto difuso medio cuando se ubica entre el segundo y tercer quintil, y al conjunto difuso bajo cuando tiene un valor menor que el primer quintil.

En el Gráfico 3 se observa que entre los quintiles uno y dos, y los quintiles tres y cuatro no se puede concluir que haya una pertenencia

(o no pertenencia) “absoluta” a alguno de los conjuntos difusos. El procedimiento mediante el cual se encuentra el grado de pertenencia de un valor a un conjunto difuso se conoce como emborronamiento. Para una explicación más detallada sobre técnicas de emborronamiento se puede consultar a Martín del Brío y Sanz (2001). En el Cuadro 1 pueden observarse algunas de las reglas obtenidas empleando los algoritmos estudiados en el presente trabajo.

Cuadro 1
Algunos resultados obtenidos aplicando cada uno de los algoritmos

A. A partir de perceptrones multicapa	
R1: Si X3 es baja y X4 es baja, entonces la salida es baja	[S=6,3%, C=80%]
R2: Si X2 es media y X3 es baja, entonces la salida es baja	[S=3,8%, C=92,3%]
B. A partir de funciones de base radial	
R3: Si $p97 \leq X3 \leq p100$, entonces la salida es alta	[S=3,8%, C=100%]
R4: Si $p95 \leq X4 \leq p100$ y $p11 \leq X5 \leq p100$, entonces la salida es alta	[S=2,5%, C=57,1%]
C. A partir de máquinas de soporte vectorial	
R5: Si $p93 \leq X3 \leq p100$ y $p0 \leq X6 \leq p97$, entonces la salida es alta	[S=5,4%, C=85%]
R6: Si $p0 \leq X4 \leq p23$ y $p9 \leq X5 \leq p33$, entonces la salida es baja	[S=5,4%, C=70,8%]

Fuente: elaboración propia.

CALIFICACIÓN DE REGLAS DIFUSAS

Finalmente, se califican las reglas obtenidas. Para esto se emplean dos medidas llamadas *soporte* y *confianza*, que han sido comúnmente empleadas en el análisis de las canastas de mercado, tal y como lo explican Han y Kamber (2006). Si se tiene una regla del tipo $A \rightarrow B$, siendo A el antecedente, B el consecuente, y “ \rightarrow ” el operador implicación (ENTONCES), las medidas de confianza y soporte se definen en las ecuaciones 6 y 7.

$$Soporte(A \rightarrow B) = P(A \cap B) * 100\% \tag{6}$$

$$\text{Confianza}(A \rightarrow B) = P(B / A) * 100\% = \frac{P(A \cap B)}{P(A)} * 100\% \quad (7)$$

Así, el soporte de una regla es el porcentaje de las muestras que cumplen tanto con el antecedente como con el consecuente. Y la confianza está asociada a la probabilidad de que se cumpla el consecuente de una regla, habiéndose ya cumplido su antecedente.

Teniendo en cuenta lo anterior, la regla 1 (R1) del Cuadro 1 podría interpretarse de la siguiente manera: cuando hay un bajo consumo de kilovatios hora por persona y un bajo porcentaje de los sueldos está destinado a personal de producción permanente (técnico, tecnólogo o profesional), entonces hay una tendencia a que el grupo estudiado presente un bajo valor agregado. Los valores de S (soporte) y C (confianza) indicados después de la regla dicen que el 6,3% de los grupos estudiados cumple con el antecedente de la regla (X3 es baja y X4 es baja), y que de estos, el 80% cumple con el consecuente también (es decir, presenta bajos valores en su variable dependiente).

La forma de interpretar las otras reglas del Cuadro 1 es similar a la anteriormente expuesta, con la diferencia de que las reglas obtenidas a partir de funciones de base radial o de máquinas de soporte vectorial tienen en sus antecedentes referencias a percentiles (por ejemplo p97 hace referencia al percentil 97) y su interpretación se da en términos de intervalos.

ALGUNAS CARACTERÍSTICAS DE LOS ALGORITMOS ESTUDIADOS

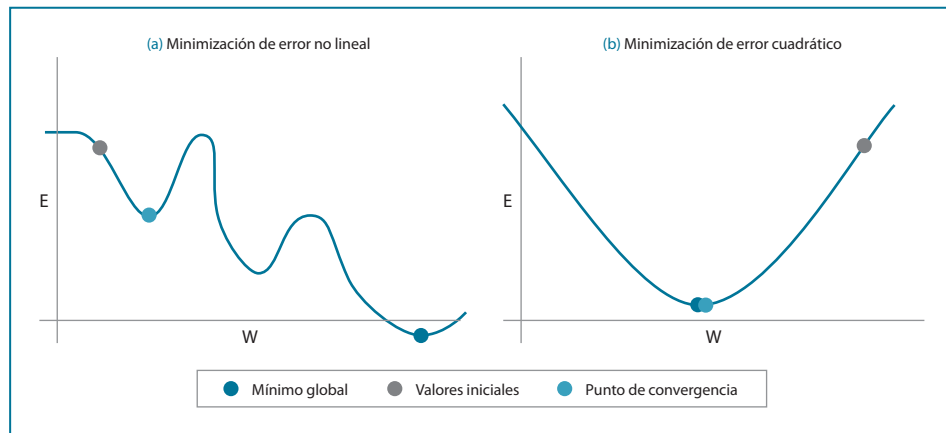
En el Gráfico 1 se observa el esquema general de aplicación de las técnicas empleadas en el Proyecto DESCUBRIMIENTO para extraer conocimiento sobre la relación entre el uso energético y la innovación a partir de la EAM 2005. En esta sección se explicitan algunas características asociadas a tal enfoque, las cuales dan pie a las propuestas de trabajo futuro que se plantean en la sección “Propuestas de trabajo futuro”. Las características que se exponen se agrupan en tres grupos: características de las regresio-

nes no lineales, características de la extracción de reglas y limitantes del procedimiento propuesto.

CARACTERÍSTICAS DE LAS REGRESIONES NO LINEALES

Para entrenar o encontrar los parámetros de una función del tipo perceptrón multicapa es necesario minimizar una función de error no lineal, algo que, en general, puede hacerse solamente usando métodos numéricos. En el Gráfico 6a se puede observar que si bien usando este tipo de algoritmos es posible llegar a minimizar una función de error no lineal localmente, en general no se llega al error mínimo global, es decir, no se garantiza convergencia.

Gráfico 6
Algoritmos numéricos de optimización



Fuente: elaboración propia.

Encontrar los parámetros de una función de base radial comúnmente se realiza en dos etapas, explicadas detalladamente en Martín del Brío y Sanz (2001): una etapa no supervisada (usando k -medias, por ejemplo) y una etapa supervisada, en la que bien podrían usarse métodos numéricos como los empleados para los perceptrones multicapa. Este enfoque

supervisado-no supervisado es igualmente un procedimiento en el cual no se garantiza llegar a un mínimo global.

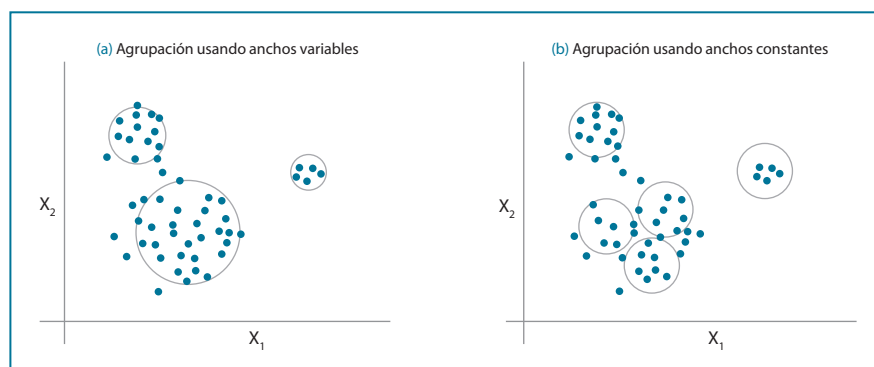
Entrenar los parámetros de una máquina de soporte vectorial es, por el contrario, un procedimiento de optimización de una función convexa y, aunque es un procedimiento numérico, garantiza una convergencia en el mínimo global de la función de error cuadrática asociada con una cierta tolerancia, como se expresa en el Gráfico 6b. Un tratamiento más detallado de estas técnicas puede encontrarse en Kecman (2001).

En las ecuaciones 2 y 3 se puede observar que hay una aparentemente leve diferencia entre una función de base radial y una máquina de soporte vectorial. Esta diferencia, asociada con el término σ , muestra que una función de base radial es una combinación lineal de funciones de Kernel con anchos variables (σ_j), mientras que una máquina de soporte vectorial es una combinación lineal de funciones de Kernel que tienen un ancho constante (σ).

Esto les da a las funciones de base radial un mayor grado de flexibilidad que, en teoría, llevaría a menores errores de regresión. En el Gráfico 7a se observa un ejemplo hipotético que surge al intentar agrupar muestras usando anchos variables (similar a como lo hacen las funciones de base radial), y en el Gráfico 7b un intento de agrupación usando anchos constantes (similar a como lo hacen las máquinas de soporte vectorial).

Gráfico 7

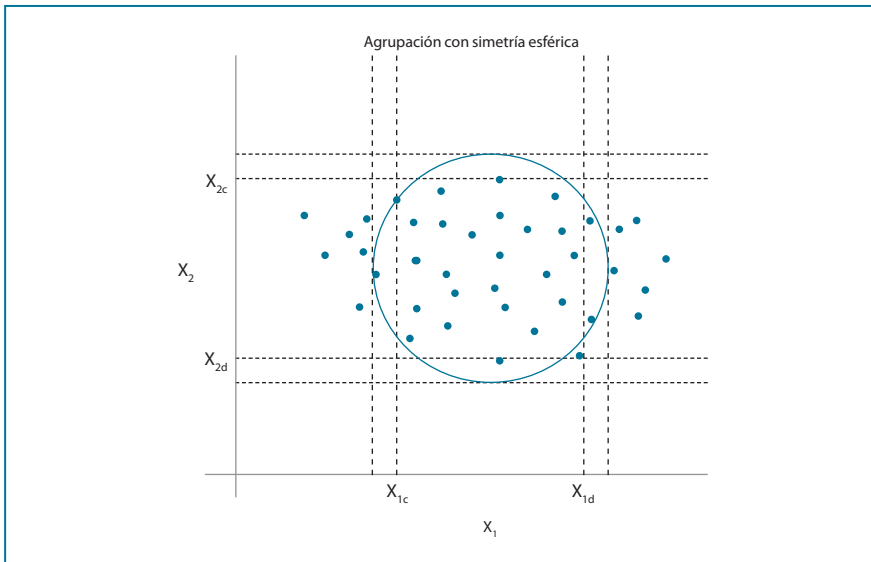
Agrupación usando anchos constantes y anchos variables



Fuente: elaboración propia.

Una característica que comparten tanto las máquinas de soporte vectorial como las funciones de base radial está asociada con las propiedades mismas de una función de Kernel, similar a como se expresa en el Gráfico 8. Esta simetría, como se verá más adelante, influye en el tipo de reglas que se extraen a partir de estas funciones al usar el algoritmo planteado por McGarry, Wermter y MacIntyre (1999). Del Gráfico 8 podría extraerse una regla similar a la ecuación 8, suponiendo que el conjunto mostrado representa elementos que coinciden en tener un alto valor de la variable Y.

Gráfico 8
Simetría de las funciones de base radial



Fuente: elaboración propia.

Si $X_{1c} \leq X_1 \leq X_{1d}$ y $X_{2d} \leq X_2 \leq X_{2c}$, ENTONCES Y es alto (8)

Esta simetría hace que $|X_{2d} - X_{2c}| = |X_{1c} - X_{1d}|$, pues se están agrupando puntos del espacio contenidos por una esfera.

CARACTERÍSTICAS DE LA EXTRACCIÓN DE REGLAS

En las ecuaciones 4 y 5 se pueden observar los tipos de regla que se obtienen a partir de perceptrones multicapa (ecuación 4) y a partir de funciones de base radial y máquinas de soporte vectorial (ecuación 5). El enunciado: “Si la temperatura es alta y la presión es muy alta, entonces hay un alto riesgo de explosión” es un ejemplo de regla de la ecuación 4; este tipo de enunciados es comúnmente utilizado en el lenguaje cotidiano y es conocido como una regla difusa simple.

Un enunciado similar como: “Si la temperatura está entre 90 y 100 grados centígrados y la presión está entre 4 y 6 atmósferas, hay un alto riesgo de explosión” es un ejemplo de las reglas expresadas por la ecuación 5. Este tipo de enunciados, si bien son un poco menos simples que los anteriormente expuestos, son más flexibles y pueden representar relaciones que no podrían expresarse con reglas difusas simples.

LIMITANTES DEL PROCEDIMIENTO PROPUESTO

El procedimiento esquematizado en el Gráfico 1 muestra que el punto de partida para aplicar los algoritmos estudiados es la selección de un conjunto de variables asociado a un fenómeno de interés y que el objetivo es llegar a un conjunto de reglas de asociación de las variables, cada una de las cuales debe estar cualificada por unos valores que miden su desempeño. Para llegar a estas reglas se debe pasar por cinco etapas que involucran una etapa de regresión y otra de extracción de reglas.

En general, este proceso demostró ser bastante robusto y útil como método analítico que integra varias técnicas y posibilita un trabajo sistemático y efectivo. Sin embargo, como se verá en la sección siguiente, es posible plantear algoritmos alternativos de tal manera que se disminuya el número de pasos necesarios entre las variables seleccionadas y las reglas obtenidas, a fin de mejorar la eficiencia del proceso.

PROPUESTAS DE TRABAJO FUTURO

USO DE ALGORITMOS DE OPTIMIZACIÓN POR ENJAMBRE DE PARTÍCULAS

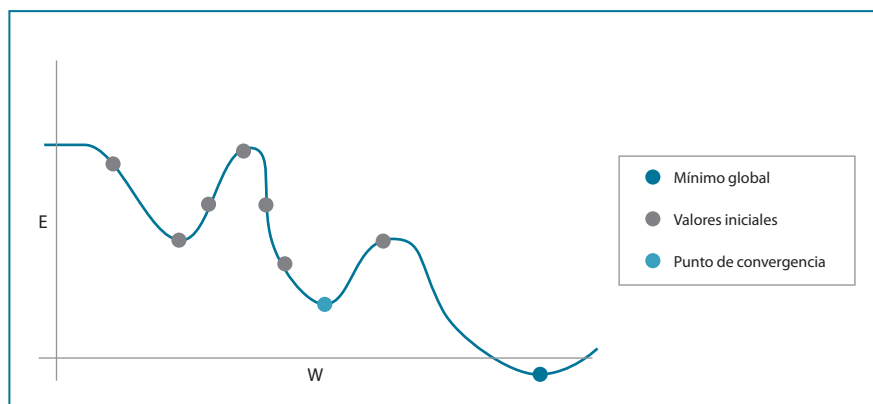
Como puede observarse en el apartado “Características de las regresiones no lineales”, encontrar los parámetros de un perceptrón multicapa o de una función de base radial es un procedimiento de optimización numérico que no garantiza convergencia en el mínimo global de la función de error minimizada (Gráfico 6a). Para intentar encontrar puntos más cercanos al mínimo global, es posible usar algoritmos bioinspirados, conocidos como de *optimización por enjambre de partículas*, presentados de una manera detallada en Barrera e Hincapié (2006).

En el Gráfico 6a se visualiza que minimizar una función de error no lineal empleando algoritmos basados en el descenso por el gradiente es un procedimiento que parte de un valor inicial que en cada iteración se va moviendo en la dirección en que hay un mayor decrecimiento de la función de error, hasta llegar a la convergencia en un mínimo local. Así, es posible que el algoritmo converja en mínimos locales diferentes cada vez que es ejecutado usando diferentes valores iniciales, como puede observarse en Bishop (1996).

Minimizar una función de error no lineal empleando algoritmos de optimización por enjambre de partículas (Gráfico 9) es, en resumen, un procedimiento que tiene en cuenta un conjunto de puntos iniciales que van recorriendo el espacio solución de una manera pseudoaleatoria. Cada punto tiene en cuenta los valores de error de los otros, y se mueve de acuerdo con esto. Con este tipo de técnicas se logra tener en cuenta un mayor subconjunto del espacio solución. Es de aclarar que este tipo de algoritmos tampoco garantiza una convergencia en el mínimo global de la función de error pero, como puede verse en el Gráfico 9, puede llegar a mínimos locales con menor error.

Gráfico 9

Minimización empleando optimización por enjambre de partículas



Fuente: elaboración propia.

FORMAS ALTERNATIVAS DE ENTRENAMIENTO DE FUNCIONES DE BASE RADIAL

La extracción de reglas a partir de funciones de base radial emplea información sobre los valores de los anchos (σ_j) y los centros \bar{u}_j de las funciones, información que fue obtenida en la etapa de aprendizaje no supervisado (ecuación 2). Un enfoque similar se podría implementar usando la etapa de entrenamiento no supervisado de la red para encontrar los valores de los centros \bar{u}_j y tomar los valores de los anchos σ_{j_0} , no como los anchos definitivos, sino únicamente como valores iniciales para la segunda etapa de entrenamiento, en la que se encontrarían los valores de los pesos \bar{w} y los valores definitivos de los anchos.

Es decir, dado un conjunto de entrenamiento $D = \{[\bar{x}_i, y_i] \in \mathfrak{R}^N \times \mathfrak{R}, i = 1, \dots, L\}$, se buscaría aplicar un algoritmo de aprendizaje no supervisado al espacio de entrada (x_i) para encontrar los valores de los centros \bar{u}_j y los valores iniciales de los anchos $\bar{\sigma}_{j_0}$; posteriormente, se realizaría un procedimiento de optimización de una función de error (ecuación 9) para encontrar los valores de $\bar{w}, \bar{\sigma}$.

$$S = \sum_{i=1}^L E(y_i - f_{RBF}(\bar{x}_i, \bar{w}, \bar{\sigma})) \quad (9)$$

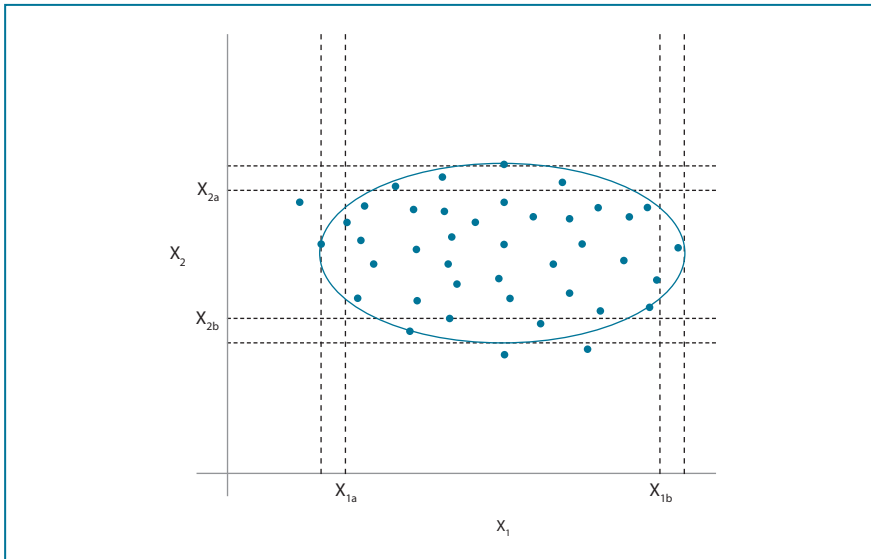
La principal ventaja que podría tener este enfoque es que los valores de los anchos de las funciones σ_j se optimizarían para obtener mejores regresiones, lo que hipotéticamente llevaría a obtener reglas con un mejor desempeño.

SUSTITUCIÓN DE LAS FUNCIONES DE BASE RADIAL

Otra característica asociada a las funciones de base radial que se observó en el apartado “Características de las regresiones no lineales” es su simetría esférica (Gráfico 8). Esta simetría hace que todas las reglas obtenidas a partir de este tipo de funciones (ecuación 8) tengan que cumplir con que $|X_{2D} - X_{2C}| = |X_{1C} - X_{1D}|$. Este problema podría eliminarse, en parte, sustituyendo las funciones de base radial por otro tipo de funciones. Como ejemplo de esto, obsérvese el Gráfico 10, de donde se podría extraer una regla en la que la restricción anteriormente planteada no existiría.

Gráfico 10

Agrupación sin simetría esférica



Fuente: elaboración propia.

MODIFICACIONES AL EMPLEO DEL ALGORITMO REFUNN

Debido a la relativa simplicidad de las reglas obtenidas a partir del algoritmo REFuNN (ecuación 4), una alternativa que presentaría un menor costo computacional sería la de evaluar todas las reglas posibles que se obtienen a partir de un determinado grupo de variables. Sin embargo, a medida en que el número de variables aumenta, la cantidad de posibles reglas aumentará, de acuerdo con la expresión $4^n \times 3$, en el caso de tres conjuntos difusos y n variables independientes. Así, para el caso de ocho variables independientes, se tendrán 196.608 posibles reglas, y para nueve, 786.432 posibles reglas.

En la medida en que el número de variables independientes aumente y se vaya volviendo más complicado probar todas las posibles reglas, se podría recurrir a una solución heurística, empleando algoritmos genéticos, que pueden estudiarse detalladamente en el trabajo de Haupt y Haupt (1997) y que definen una función de evaluación relacionada con las medidas de soporte (ecuación 6) y confianza (ecuación 7), para llegar a resultados explorando solamente un subconjunto de todas las posibles reglas.

OTRAS TÉCNICAS DE EXTRACCIÓN DE REGLAS A PARTIR DE MÁQUINAS DE SOPORTE VECTORIAL

En la actualidad existen técnicas relacionadas con extracción de reglas a partir de máquinas de soporte vectorial que podrían ser más adecuadas a las expuestas en el presente trabajo, donde básicamente se tratan las máquinas de soporte vectorial como casos particulares de las funciones de base radial. Algunas de estas técnicas se pueden encontrar en Chaves, Vellasco y Tanscheit (2005), así como en Chen, Li y Wei (2007).

APLICACIÓN DE LOS ALGORITMOS A OTRAS ENCUESTAS DISPONIBLES EN EL MARCO DEL PROYECTO DESCUBRIMIENTO

En el presente trabajo se aplicaron, de forma experimental, técnicas de inteligencia artificial a variables relacionadas con innovación y uso energético en la industria manufacturera colombiana a partir de la EAM 2005. El énfasis del trabajo es la viabilidad de poner en práctica tales técnicas, el análisis de sus potencialidades y limitaciones y la identificación de trabajo futuro. Habiendo constatado su viabilidad y potencial, el trabajo que sigue es la aplicación sistemática de las técnicas a la EAM y las encuestas de innovación, con un mayor rigor teórico en la selección de las variables y un mayor análisis de los resultados encontrados.

EXTRACCIÓN DE REGLAS DE OTROS SISTEMAS

Las técnicas de extracción de reglas a partir de redes neuronales podrían emplearse en cualquier tipo de sistema que se haya modelado con técnicas como perceptrones multicapa o funciones de base radial. Un caso particular de estos sistemas es el estudio de series de tiempo usando redes neuronales, tal y como lo plantean Jalil y Misas (2007), donde podrían aplicarse algoritmos de extracción de reglas para descubrir conocimiento relacionado con el pronóstico de los cambios, y que serían más fácilmente interpretables por parte de un analista.

CONCLUSIONES

El trabajo presentado explora la aplicación de diferentes técnicas de inteligencia artificial en la extracción de conocimiento a partir de bases de datos, a fin de construir reglas difusas que faciliten analizar relaciones no lineales entre múltiples variables asociadas a un fenómeno de interés. En el presente caso se utilizaron datos de la EAM de 2005 y un grupo de variables referidas a la relación entre innovación y uso energético.

La solución planteada en el Gráfico 1 para construir reglas difusas de una manera sistemática que, en esencia, consta de un algoritmo no supervisado y un algoritmo de extracción de reglas difusas, mostró ser un esquema modular y apropiado, pues permitió dividir una tarea compleja en varias tareas simples.

Se implementaron técnicas de extracción de reglas a partir de funciones regresoras no lineales previamente entrenadas (perceptrones multicapa, funciones de base radial y máquinas de soporte vectorial), y ello permitió inferir información de una manera sistemática a partir de funciones que en ocasiones son consideradas “cajas negras”, puesto que de ellas es difícil obtener conclusiones por simple inspección.

Se observa que algunas de las técnicas supervisadas estudiadas —particularmente perceptrones multicapa y funciones de base radial— llevan a emplear técnicas de optimización numéricas que no garantizan convergencia en el mínimo global; debido a esto, para trabajos futuros se plantea el uso de técnicas que involucren algoritmos de optimización por enjambre de partículas que podrían llevar a resultados más cercanos al óptimo.

Se observa también que entre la forma de una función de base radial y una máquina de soporte vectorial hay una diferencia que parece ser pequeña; sin embargo, resulta ser significativa, pues proporciona, por un lado, una mayor cantidad de grados de libertad a las funciones de base radial, lo que lleva a mejores aproximaciones; por el otro, hace que el entrenamiento de máquinas de soporte vectorial tenga asociado un procedimiento de optimización convexa, que garantiza convergencia cerca del óptimo global.

Las características asociadas a las funciones de Kernel empleadas dentro de las funciones de base radial y las máquinas de soporte vectorial dan pie a propuestas de investigación futura relacionadas con la sustitución de este tipo de funciones por otras con menor cantidad de simetrías, lo que podría llevar a reglas más flexibles y con mejores valores de soporte y confianza.

Al estudiar la forma de las reglas obtenidas usando las diversas técnicas empleadas, se observa que la extracción de reglas a partir de redes neuronales difusas (REFuNN) del tipo perceptrón multicapa, arroja reglas difusas simples y fáciles de interpretar; mientras que las reglas obtenidas a partir de funciones de base radial y máquinas de soporte vectorial son

más complejas y flexibles, y podrían expresar relaciones que no podrían lograrse con reglas difusas simples.

Al analizar el esquema de solución planteado en el Gráfico 1, se muestra que si bien este esquema modular permite comparar las técnicas implementadas, podría reemplazarse por un esquema con una menor cantidad de etapas que use de algoritmos genéticos bajo algunas circunstancias.

Finalmente, se puede observar que, debido a la flexibilidad y fácil implementación de las técnicas estudiadas en el presente trabajo, el estudio de otras encuestas similares a la utilizada y la exploración de relaciones de variables sobre diversos fenómenos asociados a la innovación es viable y de alto potencial analítico, lo que motiva adicionales exploraciones como trabajo futuro, más centradas en los asuntos temáticos y menos en la aplicación de las técnicas.

REFERENCIAS

- Barrera, J. e Hincapié, R. (2006). *Diseño de algoritmos Swarm Intelligence para solucionar problemas comunes en ingeniería*. Trabajo de grado de la Universidad Pontificia Bolivariana, Medellín, Colombia.
- Bishop, C. M. (1996). *Neural networks for pattern recognition*. Oxford: Oxford University Press.
- Chaves, A. C. F.; Vellasco, M. M. B. R. y Tanscheit, R. (2005). Fuzzy rule extraction from support vector machines. *Fifth International Conference on Hybrid Intelligent Systems*, 1 (6), 561-563.
- Chen, Z.; Li, J. y Wei, L. (2007). A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. *Artificial Intelligence in Medicine*, 41 (2), 161-175.
- Cristianini, N. y Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.
- Duque, L.; Manrique, J. y Robledo, J. (2008). Extracción de patrones de la encuesta anual manufacturera empleando inteligencia artificial. *Memorias: Primer Congreso Internacional de Gestión Tecnológica e Innovación* (1, p 88). Bogotá: s. e.
- Gunn, S. (1998). *Support vector machines for classification and regression* (reporte técnico). Southampton: University of Southampton. Recuperado el 2 de julio de 2008, de <http://citeseer.ist.psu.edu/7854.html>.
- Han, J. y Kamber, M. (2006). *Data mining: concepts and techniques*. San Francisco: Elsevier.

- Hartman, E.; Keeler, J. D. y Kowalski, J. M. (1990). Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation*, 2 (2), 210-215.
- Haupt, R. L. y Haupt, S. E. (1997). *Practical genetic algorithms*. New Jersey: Wiley-Interscience.
- Jalil, M. y Misas, M. (2007). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. *Revista Colombiana de Estadística*, 30, 143-161.
- Kasabov, N. K. (1996). *Foundations of neural networks, fuzzy systems, and knowledge engineering*. Cambridge: MIT Press.
- Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. Cambridge: MIT Press.
- Manrique, J. y Robledo, J. (2008). Relación entre el uso energético e innovación tecnológica en Colombia en el período 1992-2000. *Memorias: Primer Congreso Internacional de Gestión Tecnológica e Innovación* (1, p. 29). Bogotá: s. e.
- Martín del Brío, B. y Sanz, A. (2001). *Redes neuronales y sistemas difusos*. México: Alfaomega.
- McGarry, K. J.; Wermter, S. y MacIntyre, J. (1999). Knowledge extraction from radial basis function networks and multilayer perceptrons. *International Joint Conference on Neural Networks*, 4, 2494-2497.
- McGarry, K. J.; Tait, J.; Wermter, S. y MacIntyre, J. (1999). Rule-extraction from radial basis function networks. *Ninth International Conference on Artificial Neural Networks*, 2, 613-618
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6 (4), 525-533.
- Robledo, J.; Gómez, F. y Restrepo, J. (2008). Relación entre capacidades de innovación tecnológica y desempeño empresarial en Colombia. *Memorias: Primer Congreso Internacional de Gestión Tecnológica e Innovación* (1, p. 21). Bogotá: s. e.
- Serna, C. y Correa, J. (2008). Clasificación de las empresas colombianas: análisis comparativo de árboles de regresión y clasificación (CART) y regresión logística a la encuesta de innovación de 1996. *Memorias: Primer Congreso Internacional de Gestión Tecnológica e Innovación* (1, p. 22). Bogotá: s. e.

Capítulo 12

Tratamiento de datos y técnicas para analizar la relación entre el uso energético y la innovación a partir de encuestas oficiales colombianas¹

Jorge Manrique Henao
Jorge Robledo Velásquez

El presente trabajo busca contribuir a entender la relación entre uso energético e innovación tecnológica en Colombia. El capítulo se centra en el pretratamiento de los datos y la revisión de las posibilidades y limitaciones tanto de las bases de datos disponibles para el estudio como de las técnicas de análisis utilizadas. Para ello se emplean los datos sobre innovación y energía proporcionados por las dos encuestas colombianas de innovación realizadas hasta el momento de iniciar el estudio (EDT I y EDIT II), la Encuesta Anual Manufacturera (entre los años 1992 y 2005) y los reportes estadísticos anuales sobre energía del Ministerio de Minas

¹ Los autores expresan su agradecimiento al profesor Álvaro Lema Tapias de la Universidad Nacional de Colombia, sede Medellín, por sus aportes metodológicos para el desarrollo del trabajo que se reporta en esta capítulo.

y Energía de Colombia. La relación entre uso energético y crecimiento económico se utiliza como un antecedente ampliamente referenciado en la literatura para explorar el tipo de datos y las técnicas que pueden ser empleadas en el estudio de la relación entre uso energético e innovación.

Un primer acercamiento a los datos permite identificar algunas dificultades relacionadas con un porcentaje significativo de errores en la primera encuesta y una gran cantidad de datos faltantes en la segunda, lo cual conlleva descartar variables que podrían ser de importancia para el estudio. Por otro lado, la Encuesta Anual Manufacturera (EAM), además de que algunas de sus variables no aparecen en determinados años o se agregan o se desagregan entre sí en algunos períodos, no puede analizarse de manera continua en el período establecido (1992-2005) pues, a partir del 2000, la clasificación de los grupos industriales cambió a raíz de las directrices de una nueva revisión del Código Industrial Internacional Uniforme (CIIU), lo cual hace que las demás variables sólo sean comparables en los períodos 1992-2000 y 2001-2005.

A fin de tratar de abordar el estudio mediante un análisis *top-down*, inicialmente se revisó la literatura sobre la relación entre el uso energético y el crecimiento económico, donde se encontró que, hasta el momento, la mayoría de estudios sobre el tema se han enfocado en determinar a partir del análisis de series temporales, la causalidad existente entre el uso energético y el crecimiento económico medido a través del PIB, en diferentes países y para períodos relativamente largos (15 años o más).

Esta relación de causalidad, sin embargo, hasta el momento, ha sido evasiva a los investigadores, pues a pesar de la importancia que ha tenido, los resultados no han sido concluyentes (Guttormsen, 2004). También se han aplicado técnicas de estadística descriptiva, regresiones simples y, en otros casos, estadística multivariada para encontrar relaciones significativas entre variables de interés como el uso energético, el crecimiento económico, la producción y la eficiencia energética, entre otras.

En el presente trabajo, dadas las limitaciones de las características de los datos para el análisis de series temporales, se optó por aplicar distintas técnicas estadísticas que incluyen herramientas descriptivas y análisis multivariado, con las cuales se depuró gran cantidad de información

redundante y compleja asociada a un alto número de variables, a fin de construir en una primera instancia índices de valor de importancia, así como un modelo de las relaciones entre las variables depuradas, para llegar al agrupamiento y la caracterización de las diferentes empresas en unas encuestas y grupos CIIU en otras.

Los resultados preliminares muestran que la metodología empleada permite identificar muchas variables redundantes en las encuestas utilizadas como fuentes de datos, lo cual puede favorecer la identificación de dinámicas de innovación y desarrollo industrial y la determinación de índices de importancia, a partir de los cuales se pueden vislumbrar relaciones significativas entre la innovación y el uso energético en varias agrupaciones industriales en Colombia.

LAS ENCUESTAS

El trabajo se basa en varias fuentes de información disponibles con algunas restricciones impuestas por los organismos que las han aplicado. Se trata de las encuestas de innovación, efectuadas en 1996 y 2005; la EAM, entre 1992 y 2005, y estadísticas energéticas del Ministerio de Minas y Energía. Algunas de sus características y limitaciones se describen a continuación.

Las encuestas de innovación surgieron desde la década de los sesenta, por la necesidad de complementar la construcción de estadísticas e indicadores sobre investigación y desarrollo experimental (I+D) en los países de la Organización para la Cooperación y el Desarrollo Económico (OCDE), para el estudio de la innovación. Las definiciones y criterios han sido elaborados en varias versiones de los manuales de la familia Frascati, más específicamente las del *Manual de Oslo* (OECD, 2005), actualmente en su tercera versión.

En Colombia se han desarrollado varias encuestas que buscan mejorar el conocimiento sobre las dinámicas empresariales, sectoriales y regionales. En 1996, el Departamento Nacional de Planeación y Colciencias adelantaron la Primera Encuesta Colombiana de Innovación (EDT I) a una muestra de 885 establecimientos del sector manufacturero (sin el sector del tabaco),

de un universo de 4.501 establecimientos. Sin embargo, un 26,8% de la encuesta tiene la información incompleta relacionada en gran parte con las cifras de inversión en actividades de ciencia y tecnología. En general, se presentó una pérdida de muestra del 26,3%, con una tasa de respuesta de 88,5% (Salazar y Vargas, 1998), y que se evidencia durante el procesamiento de los datos, al tener que retirar del análisis variables que contienen gran cantidad de datos erróneos o por fuera de la escala de medición.

Algunas de las variables en cuestión se relacionan con el impacto de las inversiones en actividades innovadoras, maquinaria adquirida por la empresa, proyectos de desarrollo experimental, origen de la fuente de información de actualización tecnológica del establecimiento, propiedad industrial y otras variables que, aunque contienen datos, no están especificadas en el manual (véase Anexo). En el 2005 se realizó la Segunda Encuesta Colombiana de Innovación (EDIT II), para una muestra de 6.213 empresas. Sin embargo, al analizar su contenido, se encontraron datos completos sólo para 4.279 empresas (Robledo, Gómez y Restrepo, 2008).

Otro de los instrumentos de recolección de información es la EAM, que se aplica en Colombia desde 1955, con el objetivo de obtener información básica sobre el sector industrial que permita conocer su estructura, características y evolución. Ha tenido varias revisiones de actualización, como la de 1992 (CIU rev. 2), donde se adoptó el lenguaje contable del Plan Único de Cuentas (PUC), y la del 2000, donde se introdujo la CIU adaptada para Colombia (rev. 3 A. C.). Su universo de estudio y población objetivo para 2005 estaba conformado por un promedio anual de 9.816 establecimientos industriales que conforman 30 grupos industriales según la CIU rev. 3 A. C. La ficha técnica de la Encuesta y los cuadros de salida se encuentran publicados en DANE-DNP-Colciencias (2005).

La EAM presenta información generalmente de tipo contable sobre ingresos, valores de activos, costos y gastos. Debido a la clasificación diferente de grupos industriales (CIU) por las mencionadas revisiones, se dificultó analizar la información por agregación o desagregación sectorial, ya que ambas clasificaciones no son totalmente comparables. Esto obligó a hacer, de antemano, dos procesos diferentes para los períodos previo y posterior al 2000. Adicionalmente, algunas de las variables no aparecen

con regularidad en todos los años y otras se agregan o se desagregan entre sí en algunos períodos, lo cual también dificulta los análisis comparativos.

No obstante, a partir de esta encuesta se lograron identificar otros de los parámetros importantes que pueden reflejar las capacidades tecnológicas y sociales desarrolladas en el país, como son las empresas consideradas de alto conocimiento; pero para su definición todavía hay discusiones, pues depende de la organización que decide medirlas —OCDE, Instituto Nacional de Estadísticas de España (INE), etc.—. Para Colombia, el DANE incluye por CIIU rev. 2, los grupos 351: fabricación de sustancias químicas industriales; 353: refinerías de petróleo; 383: fabricación de maquinaria, aparatos y suministros eléctricos, y 384: equipo y material de transporte.

Otras de las fuentes de información para el estudio que suministran estadísticas sobre producción y consumos energéticos en el país son la Unidad de Planeación Minero Energética (UPME), entidad adscrita al Ministerio de Minas y Energía de Colombia (UPME, 2006), y la Comisión Reguladora de Energía y Gas (CREG).

DATOS Y TÉCNICAS PARA EL ANÁLISIS DE LA RELACIÓN ENTRE USO ENERGÉTICO Y CRECIMIENTO ECONÓMICO

Debido a las exigencias de aumento en la eficiencia e intensidad energética a raíz de la crisis del petróleo en los años setenta, se generó un cuestionamiento sobre la correlación de variables como el crecimiento económico y el uso energético en términos de cuál de ellos estimula el crecimiento del otro por canales indirectos de demanda agregada efectiva, capital humano, aumento en la efectividad y progreso tecnológico (Masih y Masih, 1997, citados por Ferguson, Wilkison y Hill, 2000). Recientemente, la literatura relacionada con el consumo energético y el crecimiento económico ha tomado mucho interés por dos nuevas razones: la primera, relacionada con el cambio climático y con el debate sobre si los países en vía de desarrollo tienden a incrementar sus emisiones de gases contaminantes, al ser el uso energético un requisito para su crecimiento económico. La segunda, relacionada con los desarrollos en metodologías econométricas, particu-

larmente la cointegración, que han permitido relacionar series temporales (Guttormsen, 2004).

Por ende, se han realizado muchos estudios que tratan de encontrar la causalidad entre el consumo o el uso energético y el crecimiento económico o ingreso, medido a través del producto interno bruto (PIB) o el producto nacional bruto (PNB), utilizando diversos métodos, pues si la causalidad unidireccional va del consumo energético a los ingresos (o crecimiento económico), implica que la reducción del consumo energético (más específicamente de electricidad) llevaría a disminuir los ingresos; mientras que si sucede lo contrario, o si no hay causalidad en ninguna dirección, las políticas de conservación energética no afectarían el ritmo del crecimiento económico o el crecimiento económico estimularía el consumo de electricidad (Narayan y Prasad, 2008). Por otro lado, la causalidad bidireccional sugiere que el consumo energético (generalmente de electricidad) y el PIB son interdependientes (Yoo y Ku, 2008). Adicionalmente, la literatura económica, por lo general, cita el consumo de electricidad por parte del sector industrial como el principal contribuyente al crecimiento (Shiu y Lam, 2004). Los resultados de esta relación de causalidad, sin embargo, no han sido concluyentes (Guttormsen, 2004).

En general, los trabajos se han abordado desde diferentes estrategias y enfoques metodológicos, o desde análisis y comparación de diferentes fuentes, algunas de las cuales se citan en Manrique y Robledo (2008). Según Guttormsen (2004), las metodologías utilizadas para el estudio del uso energético y el crecimiento económico han tenido cierta evolución desde los primeros trabajos hechos a finales de los años setenta, pues iniciaron con la metodología VaR y pruebas de causalidad de Granger,² las cuales asumen que las series de datos son estacionarias. Luego se consideraron las variables no estacionarias y se aplicaron herramientas de cointegración para estudiar las relaciones; posteriormente, las que se han trabajado con estimadores multivariados.

² El test o prueba de Wiener-Granger fue desarrollado por el premio Nobel de Economía (2003) Clieve W. J. Granger y consiste en comprobar si los resultados de una variable (como serie temporal) pueden predecir el comportamiento de otra y si tiene carácter unidireccional o bidireccional. Véase Granger (1969).

Entre las metodologías más usadas se encuentran diferentes técnicas de series temporales, como análisis de correlación, causalidad bivalente, cointegración multivariante, paneles de cointegración, modelamientos de corrección de errores vectoriales y descomposiciones de varianza, que se han aplicado para determinar la fuerza de la causalidad entre el consumo energético y el PIB (Climent y Pardo, 2007). Se señala, en particular, un estudio realizado por Hannesson (2002), basado en estadística descriptiva y regresiones simples, en el cual se divide el rango de tiempo abarcado en subperíodos, según la evolución de los precios del petróleo, para concluir que la relación entre el crecimiento del uso de energía y el del PIB es bajo luego de la primera crisis del petróleo y aumenta hacia 1986, especialmente en los países ricos (véase Guttormsen, 2004).

Este mismo trabajo arrojó como conclusiones principales que, a diferencia de los primeros estudios realizados por varios investigadores, la mayoría de estudios recientes indican una relación bidireccional entre el uso de la energía y el crecimiento económico; que las metodologías usadas y los períodos seleccionados pueden influir en los resultados de los estudios, y que se utiliza el procedimiento de Johansen para relacionar más de dos series temporales.

Una revisión de los resultados de los estudios efectuados hasta ahora, basados en análisis de series temporales y de diferentes fuentes, muestra que arrojan direcciones de causalidad en varios sentidos, es decir, para algunos países la causalidad unidireccional va del consumo energético al PIB; en otros, se encuentra causalidad bidireccional; mientras en otros no se encuentra ninguna causalidad. En algunos casos se ha encontrado causalidad en un sentido, si se estudia el corto plazo, y en sentido contrario, cuando se analiza el largo plazo. También se encuentran trabajos que relacionan, además, otros factores de producción, como los de Wolde-Rufael (2009) y Narayan y Smyth (2005). En general, los trabajos sobre el uso energético y el crecimiento económico se han abordado desde diferentes estrategias y enfoques metodológicos, principalmente econométricos, y en algunos casos estadísticos o desde análisis y comparación de diferentes fuentes —véanse referencias de la revisión en Manrique y Robledo (2008)—.

Por otro lado, existen algunas referencias sobre indicadores que relacionan la energía con las capacidades tecnológicas de un país, como el caso del consumo de electricidad (kWh) per cápita del Banco Mundial, mencionado por Archibugi y Coco (2005), mientras que Shiu y Lam (2004) establecen que el consumo de electricidad por el sector industrial es el principal contribuyente al crecimiento. En cuanto al uso de energías renovables, Sadorsky (2009) ha encontrado recientemente a través de paneles de cointegración que el aumento en los ingresos reales per cápita en economías emergentes, dentro de las que se ha estudiado la colombiana, genera en el largo plazo un impacto positivo sobre el consumo per cápita de energías renovables.

Los estudios energéticos también se han abordado desde la determinación de sectores clave y plantean, desde el modelo de Leontief—que busca un índice de producción que pueda equilibrar la demanda total—, un resultado que expresa la energía final total utilizada por el sistema productivo y un vector que indica la energía final por unidad de *output* sectorial (Alcántara y Padilla, 2003). Por otro lado, Kumar Ray y Sudhakara Reddy (2008) desarrollaron indicadores de intensidad energética para subsectores del sector industrial en la India (hierro y acero, aluminio, textiles, papel y cemento) utilizando datos en unidades físicas de producción y uso energético. También se encuentran trabajos en los que se utilizan técnicas de estadística multivariante, como análisis de conglomerados o *clusters*, para clasificar los sectores industriales en función de variables relacionadas con los *inputs* y *outputs* del proceso innovador (Delgado Alaminos y Sánchez Vizcaíno, 1999).

El consumo energético en el sector industrial es diferente al de los otros sectores porque, entre otras razones, este puede llegar a producir parte de la electricidad que necesita; por lo tanto, su consumo es igual a la electricidad generada más la energía comprada. El petróleo, el gas y la biomasa forman una parte importante de los insumos de este sector como combustibles o como materias primas. El carbón se ha ido reduciendo como combustible directo en las industrias, pero su consumo se ha incrementado en plantas térmicas generadoras de electricidad para todos los sectores, excepto el del transporte (Kumar Ray y Sudhakara Reddy, 2008).

Recientemente, los debates sobre el calentamiento global y las emisiones de dióxido de carbono han llevado, además de los análisis meramente econométricos, a analizar la intensidad energética con el fin de elaborar estrategias que permitan disminuir tales efectos (Kumar Ray y Sudhakara Reddy, 2008). Adicionalmente, el desarrollo de indicadores específicos de consumo energético basados en unidades físicas parecen reflejar mejor la realidad que los basados en unidades monetarias, y favorecen el contraste internacional y la formulación de políticas sectoriales (Kumar Ray y Sudhakara Reddy, 2008).

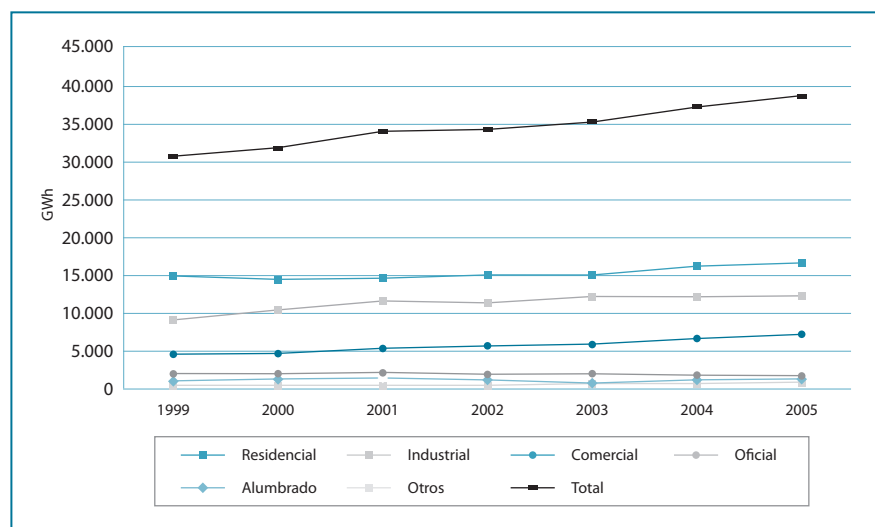
Los análisis se deben centrar en los subsectores industriales que más energía demandan: hierro y acero, aluminio, textiles, papel y cemento (Worrell *et al.*, 2008), y en los principales factores que determinan el consumo energético en una economía, como la producción global, la estructura de la economía y el producto o actividad por unidad de energía empleada, componente que se refiere a la intensidad energética, cuya disminución ocurre cuando aumenta la cantidad de producto, servicio o actividad por unidad de energía, y puede indicar que se han hecho implementaciones de eficiencia energética.

Un indicador muy usual de intensidad energética es la relación entre el uso energético sectorial/subsectorial y el producto o actividad del sector o subsector (Kumar Ray y Sudhakara Reddy, 2008). En un trabajo de clasificación sobre el consumo de energías tradicionales y modernas en 20 países latinoamericanos se mostraba que, aunque individualmente el comportamiento es heterogéneo, la correlación entre el consumo de energías y el PIB per cápita es positiva, cambia en el tiempo y es mayor cuando el crecimiento económico es más rápido; adicionalmente, el consumo de energías modernas es mayor que el consumo de energías totales en todos los casos (González, 2006).

En el caso de los países de más bajo consumo de energías per cápita, la más alta correlación ocurre en el período 1981-2003 debido, entre otros factores, a una mayor cantidad de países que han completado la transición energética de las energías tradicionales a las modernas y por los malos resultados de las políticas económicas aplicadas. Los cuatro grupos o patrones identificados de uso aparente de energía en el último siglo, relacionados

con el consumo y el crecimiento, permiten ubicar a Colombia en el patrón de los países con bajo consumo aparente de energías modernas inicial y tasas decrecientes a largo plazo, entre otras razones, por débiles experiencias industriales y debilidad institucional histórica (González, 2006). El consumo de energía eléctrica en el sector industrial colombiano ha sido cercano al 30% del total e inferior al consumo residencial, como puede apreciarse en el Gráfico 1.

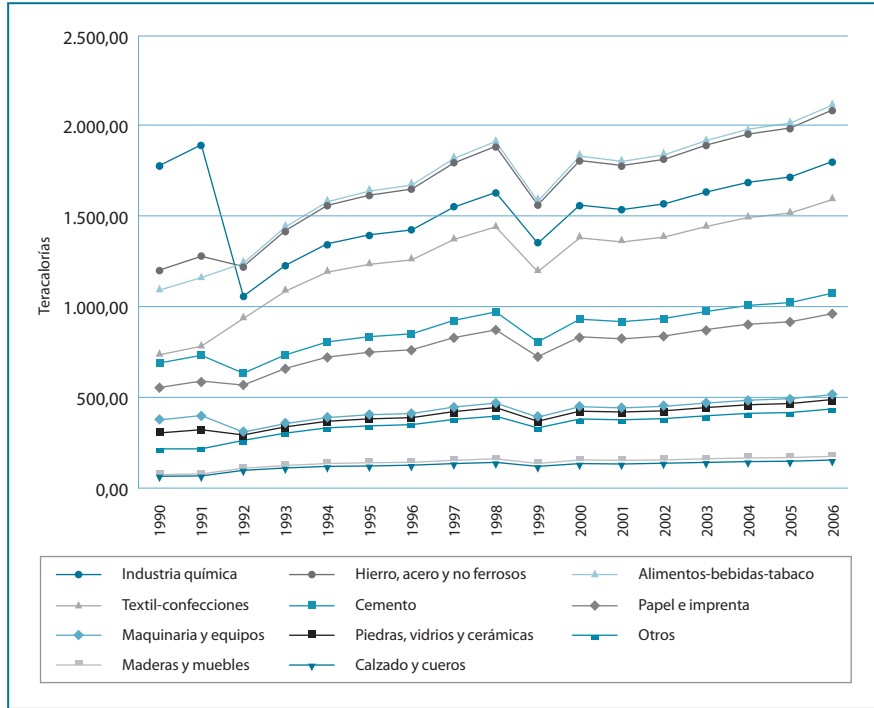
Gráfico 1
Consumo de energía eléctrica por sectores



Fuente: elaboración propia con base en datos de la UPME.

El sector industrial está conformado por subsectores, cuyo consumo de energía eléctrica desde 1990 ha tenido tendencia a permanecer constante, con excepción del año de recesión 1999, cuando hubo una importante disminución en todos los subsectores, como se muestra en el Gráfico 2.

Gráfico 2
Consumo de energía eléctrica en Colombia por sector industrial

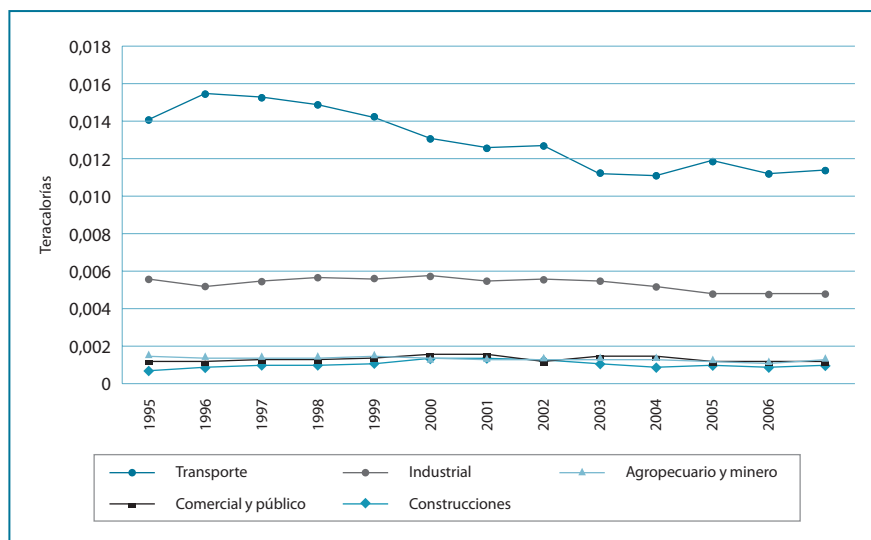


Fuente: elaboración propia con base en datos de la UPME.

La intensidad energética por sectores es el cociente entre el consumo final de energía por sector y el PIB sectorial, expresado en valor constante, y refleja las tercalorías requeridas para obtener un millón del producto. En Colombia, la intensidad energética presenta una significativa caída en las últimas décadas, lo cual indica, según la UPME (2006), que en el sector productivo, a partir de 1992, disminuyeron los consumos al mejorar la eficiencia energética, no sólo por las políticas de apertura económica, sino por los años de crisis a finales de la década de los noventa (Gráfico 3).

Gráfico 3

Intensidad energética por subsectores (teracalorías/millones de pesos)



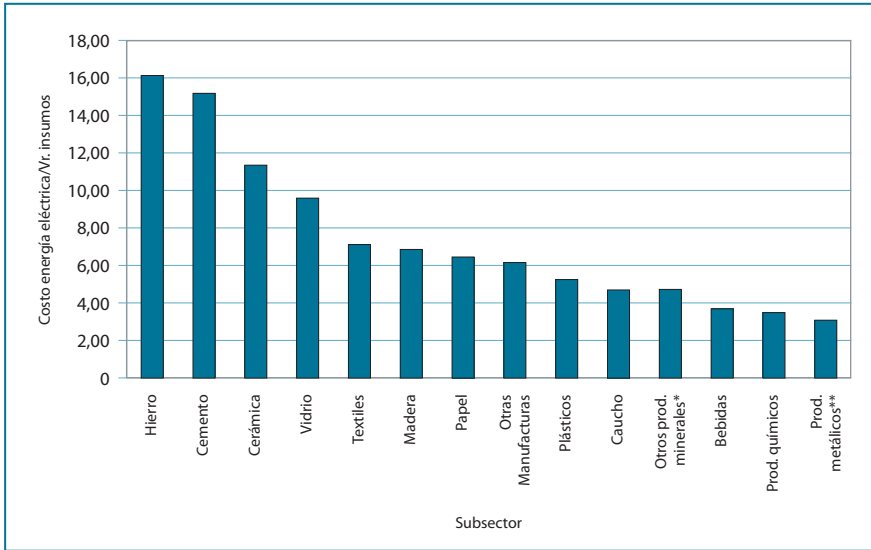
Fuente: Departamento Nacional de Planeación (DNP); UPME–Ministerio de Minas y Energía.

En Colombia se han hecho esfuerzos mediante propuestas encaminadas a emplear racionalmente el recurso energético en la industria, fundamentados en algunas comparaciones con otros países, como lo muestra una presentación realizada en 2006 por el Ministerio de Comercio, Industria y Turismo (Salazar, 2006), donde se menciona que una de las razones para promover el uso racional de la energía es porque en las pymes el consumo de energía eléctrica representa hasta un 25% de la producción total industrial. La participación del costo de la energía eléctrica en el costo total de los insumos es mayor en los sectores del hierro, el cemento y la cerámica, lo cual coincide en parte con los sectores mencionados por Worrell *et al.* (2008) y Kumar Ray y Sudhakara Reddy (2008) (Gráfico 4).

En el 2003, por ejemplo, los subsectores que demandan un mayor insumo de energía eléctrica respecto al total de su producción son el textil, el de vidrio y productos de vidrio y el de los materiales metálicos, cada uno con una participación cercana al 10% (Cuadro 1).

Gráfico 4

Participación de la energía eléctrica en el valor total de los insumos



* No metálicos; ** Sin incluir maquinaria ni equipos.
Fuente: elaboración propia con base en Salazar (2006).

Cuadro 1

Sectores con mayor demanda de energía eléctrica sobre el total de su producción en 2003

Subsector	% sobre el total de la producción
Hilados e hilos; tejidos de fibras textiles	10,9
Vidrio y productos de vidrio	9,9
Metales comunes y productos metálicos	8,2
Productos de caucho y productos plásticos	7,2
Productos de petróleo refinado	6,9
Artículos textiles (excepto prendas de vestir)	6,7
Pasta de papel, papel y cartón	5,2
Impresos y artículos análogos	5,2
Otra maquinaria y suministro eléctrico	4,5
Azúcar	4,4
Tejidos de punto o ganchillo; prendas de vestir	4,1
Café transformado	4,0

Continúa

Subsector	% sobre el total de la producción
Aceites, grasas animales y vegetales, borras y tortas	3,5
Productos de madera, corcho, paja y materiales trenzables	3,3
Bebidas	3,2
Productos químicos básicos y elaborados	3,0
Cuero y productos de cuero; calzado	3,0

Fuente: elaboración propia con base en Salazar (2006).

MÉTODOS DE ANÁLISIS

Al momento de iniciar el estudio se contaba con los datos de las dos encuestas de innovación (EDT I y EDIT II) y datos anuales de consumo energético y de valores contables de la EAM; también estaban disponibles los reportes estadísticos sobre energía del Ministerio de Minas y Energía. Sin embargo, la falta de series temporales sobre innovación hizo que el análisis no pudiera abordarse por los métodos mencionados antes para el estudio de la causalidad entre uso energético y desarrollo económico. En su lugar, se realizó un análisis previo de las variables de las diferentes bases de datos para identificar las dificultades iniciales de contraste de los datos en el período establecido para el estudio y se aplicaron técnicas de estadística descriptiva para identificar las características individuales de las variables. Posteriormente, se aplicaron técnicas de estadística multivariante, como análisis de componentes principales, que, junto con un índice de valor de importancia asignado a las variables en cuestión por parte de expertos conocedores del fenómeno estudiado, buscaron colinealidades en las variables para eliminar la redundancia en la información del sistema.

Luego, un análisis de conglomerados a *clusters* permitió identificar y caracterizar grupos de empresas con comportamientos o desempeños similares. A continuación se describen de manera breve los principales métodos estadísticos utilizados; no obstante, los detalles técnicos se encuentran en la bibliografía referenciada. El análisis completo de los resultados de la aplicación de las técnicas y su interpretación a la luz de las teorías existentes es un trabajo en curso a la fecha, no incluido en el capítulo.

ANÁLISIS EXPLORATORIO

Cuando se tienen datos provenientes de diversas fuentes, es importante realizar un análisis exploratorio en varias etapas, cuyo objetivo es evaluar la calidad y la consistencia de los datos; presentar un análisis exploratorio y gráfico de los datos, para determinar la distribución de las variables; aplicar el tratamiento de datos ausentes; detectar y tratar datos atípicos (*outliers*); comprobar los supuestos subyacentes en los métodos multivariantes: normalidad, linealidad, homoscedasticidad y multicolinealidad; resumir los datos mediante el uso conjunto de medidas de resumen y sus gráficos, y explorar formas de categorizar variables a través de búsqueda de puntos de corte (Díaz de Rada, 2002).

De acuerdo con las relaciones entre variables, el análisis exploratorio de datos puede ser univariado, bivariado o multivariado, y este último es el caso específico del presente análisis. En cuanto a su forma de visualización, el análisis exploratorio de datos puede ser descriptivo o inferencial y se puede realizar gráficamente y mediante pruebas de hipótesis y de contraste. Al principio del apartado “Resultados preliminares” se muestran algunos de los resultados obtenidos del análisis exploratorio.

ANÁLISIS DE COMPONENTES PRINCIPALES

El análisis de componentes principales (ACP) es una técnica de análisis multivariado que busca descubrir unos factores más importantes que pudieran explicar, si existen, ciertos patrones de comportamiento a través de las variables originales. Su principal uso se relaciona con la explicación de las estructuras de varianza y covarianza de una serie de variables originales, mediante unas pocas combinaciones lineales de ellas, con lo cual se logra reducir la cantidad de datos, disminuir la dimensionalidad del problema, simplificar los procesos y mejorar la interpretación de los resultados (Lema, 2005).

Esto se consigue al seleccionar aquellas componentes cuyos valores propios sean iguales o superiores a uno y al determinar cuál es el porcen-

taje de variabilidad explicada en los datos originales. El procesamiento de datos en un programa de aplicaciones estadísticas como Statgraphics® permite visualizar mediante una gráfica de criba (*cree plot*) el número de combinaciones lineales de las variables estudiadas, que se encuentran por encima o por debajo del valor uno. Las redundancias se eliminan al sacar las variables que presentan multicolinealidad (características similares a la de otra variable) o que introducen ruido al sistema y disminuyen el porcentaje de explicabilidad.

Por lo general, al realizar sucesivamente el proceso por subgrupos de variadas³ y, posteriormente, con las variables resultantes se encuentra un porcentaje de explicabilidad mucho mayor. Esta es una de las potencialidades del método de ACP (Lema, 2005), que puede aplicarse para conocer la relación entre los elementos de una población cuando se sospecha que en dicha relación influyen de manera desconocida unas variables o propiedades de dichos elementos (Pla, 1986).

En un análisis de datos ecológicos se considera que con una explicación de variabilidad superior al 30% se logra una buena percepción de los fenómenos altamente aleatorios. Es decir, se busca obtener combinaciones lineales de componentes cuyos porcentajes de varianzas sean superiores a este valor (ter-Braak *et al.*, 2002, citado en Lema, 2005). En el presente caso, el proceso de reducción de dimensionalidad por el método de ACP se aplica a las variables relativizadas de cada subvariada, en procesos sucesivos hasta llegar a cada variada.

ÍNDICE DEL VALOR DE IMPORTANCIA RELATIVIZADO

El índice de valor de importancia relativizado (IVIR) es un indicador adimensional que muestra el grado de importancia estadística que tiene un individuo en un grupo o contexto, según la totalidad de las características de las variables estudiadas (Lema, 2005). En el caso estudiado, puede in-

³ Se llamará *variada* a cualquier factor de importancia que pueda desagregarse en otras variadas en varios niveles, hasta llegar a las variables finales.

dicar el potencial de innovación de las empresas en la muestra encuestada. El procedimiento empleado se basa en el manejo de la tendencia, que consiste en generar datos para descubrir las variables más importantes a través de calificación y ponderaciones de variables que van del mejor al peor puntaje. Se calcula mediante varios procesos, el primero de los cuales consiste en seleccionar las variables del fenómeno que se va a estudiar y ponderarlas a través de un panel de expertos, al asignar un valor de prioridad o ponderador a cada variable que se encuentra dentro del grupo o variada de orden superior.

Las variables seleccionadas se relativizan dentro de cada variada de siguiente orden, según los pesos previamente asignados. Se realiza luego una disminución de dimensionalidad del sistema mediante el método de ACP aplicado a cada subvariada, a cada variada y a la totalidad del sistema resultante con el fin de identificar colinealidades entre las variables y eliminar las redundancias entre ellas, es decir, aquellas variables que ofrecen información que ya se encuentra en el sistema. Con estas variables se calcula el IVIR para las variables pertenecientes a cada subvariada y a cada variada, como la sumatoria de los valores relativizados de las variables en cada subvariada. El IVIR de cada variada de orden superior será la suma ponderada de los IVIR de cada subvariada, multiplicada por el peso asignado de la variada correspondiente, y dividido por la sumatoria de los índices de valor de importancia (IVI) de las subvariadas que componen la variada.

$$IVI = \frac{\sum x_i P_i}{\sum P_j}$$

Donde x_i es el valor de la variable luego de haberse convertido a escala de valores, P_i es el peso asignado por los expertos a la variable con respecto al fenómeno estudiado y P_j es el peso de la variada, compuesto por la sumatoria de los pesos de las variables que la componen. Con los valores resultantes de IVI para cada observación se conforman los IVI totales o IVITO, los cuales se someten luego a un análisis de conglomerado (*cluster*) según varios métodos o estrategias de fusión y así se clasifica cada observación según el *cluster* al que pertenezca. Mediante el IVITO se comparan las diferentes encuestas y se identifican aquellos casos de mayor relevancia para el sistema.

ANÁLISIS DE CONGLOMERADOS O CLUSTER

Por conglomerado o *cluster* se acepta la definición de un conjunto de objetos similares (Mollineda, 2001). El análisis de conglomerados consiste en agrupar un conjunto de datos multidimensionales (filas de la matriz de datos X) en un conjunto de grupos homogéneos. Para ello se utilizan funciones de similitud entre ellos. Así mismo, el análisis de conglomerados puede emplearse para agrupar variables (columnas de la matriz de datos X). Estos métodos se conocen como métodos de clasificación automática o no supervisada. Se busca incluir en un mismo grupo a individuos con fuertes similitudes entre ellos (alta homogeneidad interna dentro del *cluster*), que sean disímiles a su vez con individuos de grupos distintos (alta heterogeneidad externa entre *clusters*) respecto a algún criterio de selección predeterminado.

En una representación geométrica de una clasificación exitosa, los objetos dentro de un conglomerado estarán muy cercanos unos de otros, y los conglomerados diferentes estarán muy separados (Cortijo, 2001). El análisis de conglomerados se puede combinar con el ACP, pues mediante este se pueden homogeneizar los datos y posteriormente se puede realizar un análisis de *cluster* sobre los componentes obtenidos. Una técnica de agrupamiento debe encontrar y caracterizar los agrupamientos para luego etiquetar los patrones con base en estos (Cortijo, 2001). Cada patrón se asocia con un agrupamiento y cada agrupamiento consiste en un conjunto de patrones, y se formalizan de la siguiente manera: sean M patrones X_1, X_2, \dots, X_M de P .

Un proceso de agrupamiento consiste en buscar K grupos (conjuntos de patrones) S_1, S_2, \dots, S_K tales que todo $X_i, i=1, 2, \dots, M$ pertenece a uno y sólo uno de estos grupos.

$$S_1 \cup S_2 \cup \dots \cup S_K = S_i \cap S_j = 0 \quad \forall i \neq j$$

Usualmente, cada agrupamiento se caracteriza por un centroide (media) al que se denota por $Z_i, i=1, 2, \dots, K$. Puesto que el propósito del conglomerado es agrupar objetos similares, se necesita alguna medida para evaluar las diferencias y las similitudes entre objetos. El concepto de similitud

es una medida de correspondencia o semejanza entre los objetos que van a ser agrupados. La estrategia más común consiste en medir la equivalencia en términos de la distancia entre los pares de objetos, agrupando en un mismo conglomerado aquellos que tienen distancias reducidas entre ellos y que, por lo tanto, son más parecidos entre sí.

Existen tres tipos de medidas de similitud, dependiendo de los objetivos y los datos: de *correlación*, de *distancia* y de *asociación*. Las primeras y segundas requieren datos métricos, mientras que las medidas de asociación requieren datos no métricos (Figueras y Gargallo, 2003). El criterio de homogeneidad más simple se basa en la distancia entre ejemplares de un mismo agrupamiento respecto a su prototipo o centroide (promedio, mediana, etc.) que debe ser significativamente menor a la distancia entre ejemplares de agrupamientos diferentes ubicados, por ejemplo, en un plano cartesiano. El término *distancia* se refiere a una función δ que cumple con los siguientes axiomas métricos (Mollineda, 2001):

1. Propiedad de no negatividad: $\delta(x, y) \geq 0$.
2. Propiedad del cero: $\delta(x, y) = 0$ si y sólo si $x=y$.
3. Simetría: $\delta(x, y) = \delta(y, x)$ para todo x, y .
4. Desigualdad triangular: $\delta(x, y) + \delta(y, z) \geq \delta(x, z)$ para todo x, y, z .

Las medidas de proximidad que se usan para variables cuantitativas son el coeficiente de congruencia o el de correlación (Figueras y Gargallo, 2003). Para datos binarios, el coeficiente de Jacard y el de acuerdo simple. Las medidas de distancia para variables cuantitativas más usuales son: la distancia euclídea y distancia euclídea al cuadrado, la métrica de Chebyshev, la de Manhattan, la de Minkowski y la de Mahalanobis.

También hay medidas para tablas de frecuencias basadas en la χ^2 de Pearson y medidas para datos binarios como la distancia euclídea al cuadrado, la de Lance y Williams. Finalmente, existen medidas para datos de tipo mixto. Los métodos de análisis de conglomerados o *clustering* también se conocen como métodos de clasificación automática o no supervisada. Existen dos tipos de procedimientos: *métodos no jerárquicos*, que asumen un modelo definido para los datos, y *métodos jerárquicos*, que pueden ser

acumulativos (por aglomeración o *bottom-up*) o divisivos (*top-down*) y que no asumen ningún modelo estadístico para los datos. Los datos se ordenan en niveles de manera que los niveles superiores contienen a los inferiores y la jerarquía construida permite obtener una partición de los datos en grupos. En este caso se utiliza la matriz de distancias o similitudes entre elementos de la matriz de datos.

El conglomerado por aglomeración empieza con cada objeto en un grupo separado y luego los objetos se van agrupando en conjuntos cada vez más grandes hasta formar un solo grupo. Dentro de los procedimientos de agrupación (existen varios) se encuentra, por ejemplo, el *procedimiento de Ward*, cuyo objetivo es que en cada paso se formen grupos lo más homogéneos posible internamente. En cada agrupamiento se calculan las medias para todas las variables. Luego, para cada objeto, se calcula la distancia euclídea al cuadrado para las medias de los grupos; estas distancias se suman a todos los objetos. En cada etapa, se combinan los dos agrupamientos con el menor incremento en la suma total de los cuadrados de las distancias dentro de los *clusters*.

De los métodos jerárquicos, el método de enlace promedio y el procedimiento de Ward han demostrado un mejor desempeño que otros (Gondar, 2000). La representación gráfica del proceso de agrupación se denomina dendograma, que consiste en un gráfico cartesiano en el cual se ubican los objetos y las distancias a las que estos se agrupan. Se lee de izquierda a derecha cuando es vertical. Las líneas verticales representan los grupos. La posición de la línea en la escala indica las distancias en las que se unen los grupos.

METODOLOGÍA

Las diferentes metodologías de cálculo exigen que los datos sean debidamente preparados, para lo cual se procede a identificar las variables según los manuales suministrados para ello, a seleccionar aquellas de importancia para el estudio, a encontrar variables con datos erróneos y a limpiar, estandarizar y relativizar los datos de trabajo. Se evita así la influencia negativa en el análisis de la presencia de diferentes unidades de medición;

de diferentes formatos en los registros, como cifras decimales y cifras en porcentaje; de errores en toma de datos, y de uso de diferentes escalas de valores como $(0 \text{ y } 1)$, $(1, \dots, n)$, $(0, \dots, n)$, entre otros problemas.

Adicionalmente, se requiere que los datos del análisis se encuentren en una misma tabla. La preparación para el análisis se realiza mediante un proceso de selección de variables pertinentes de las bases de datos por parte de los investigadores, pues es posible que existan variables no relacionadas con el proceso, que posteriormente retrasarían el análisis. Deben estandarizarse también las diferentes escalas de calificación y valores de las respuestas, por otras en las cuales se considere el cero como el valor de una actividad que no se realiza, no se tiene o no aplica. Para el cálculo de los índices de valor de importancia, un grupo de expertos en los conceptos que se van a analizar —en este caso las capacidades de innovación presentes en cada variable de la encuesta— asigna un valor numérico de importancia o peso p_i a cada variable seleccionada, entre m y 1 , siendo m el número de variables de la variada procesada y 1 el valor mínimo de dicha variada, en orden descendente. Luego se aplica el mismo procedimiento por cada grupo de subvariadas que componen una variada, asignando el valor de mayor importancia o prioridad desde k para la subvariada más importante de la variada estudiada hasta 1 como menor valor.

Mediante el ACP se abordaron las encuestas de innovación de 1996 y 2005, desde una perspectiva global, con el fin de indagar, inicialmente, si existía información redundante en dichas encuestas y qué tan importante era la información resultante al eliminar las redundancias o ruidos. Este proceso se logró reduciendo la dimensionalidad del sistema de variables mediante el ACP y, posteriormente, clasificando según un IVI. Las variables resultantes serán componentes de algunas de las capacidades de innovación, y de manera preliminar, pueden permitir identificar aquellas capacidades comunes en ambas encuestas. Posteriormente, se contrastarán estos resultados y el uso energético en estos años, para determinar si existe alguna relación importante entre estos resultados.

Se seleccionaron las variables que presumiblemente ofrecieran una mejor información para identificar capacidades de innovación tecnológica, según parámetros planteados en la encuesta y en la literatura. Puede

trabajarse con todas las variables, si no se tienen elementos para seleccionar algunas de ellas. Se ilustra, a manera de ejemplo, el proceso desarrollado con alguna de las variadas de la EDT I, que corresponden a los capítulos, los cuales se dividen en subcapítulos o subvariadas, y estas, a su vez, en variables, para un total de tres niveles. Un grupo de expertos en el tema asignó valores de ponderadores a cada variable, a cada subvariada y a cada capítulo o variada, según la relevancia que la variable o variada tuviera en el fenómeno que se va a estudiar.

Posteriormente, se introdujeron nuevas escalas de valores de la tabla (entre 0 y n) para todas las variables que lo requirieran y se relativizó la tabla de datos, multiplicando el valor seleccionado de la variable por el ponderador y dividiéndolo por la sumatoria de ponderadores de la subvariada correspondiente. Estos valores se sometieron a un ACP, según el procedimiento descrito antes, y se redujo la dimensionalidad del sistema. Los datos resultantes del anterior proceso de ACP son el insumo para calcular los IVI para cada subvariada y cada capítulo. Con estos se calculó el IVITO.

Finalmente, con los datos resultantes de los IVI e IVITO se realizó un análisis de *cluster* utilizando como medida de distancia la euclídea al cuadrado y varios métodos de fusión, que buscaban producir agrupamientos con diferentes grados de incertidumbre, pues no se sabía para las variables utilizadas cuál era la mejor. Se produjo así un conjunto de datos con los cuales es posible calcular los respectivos límites de confianza para la separación entre los grupos y poder encontrar las fronteras de separación entre un grupo y otro. Mediante el análisis de *cluster* se buscaba definir un número de grupos igual al de las categorías en que se dividiría la población y encontrar los valores entre los cuales se sitúan los IVITO para el agrupamiento final de las observaciones (empresas o agrupaciones CIU en este caso) en la jerarquía definida.

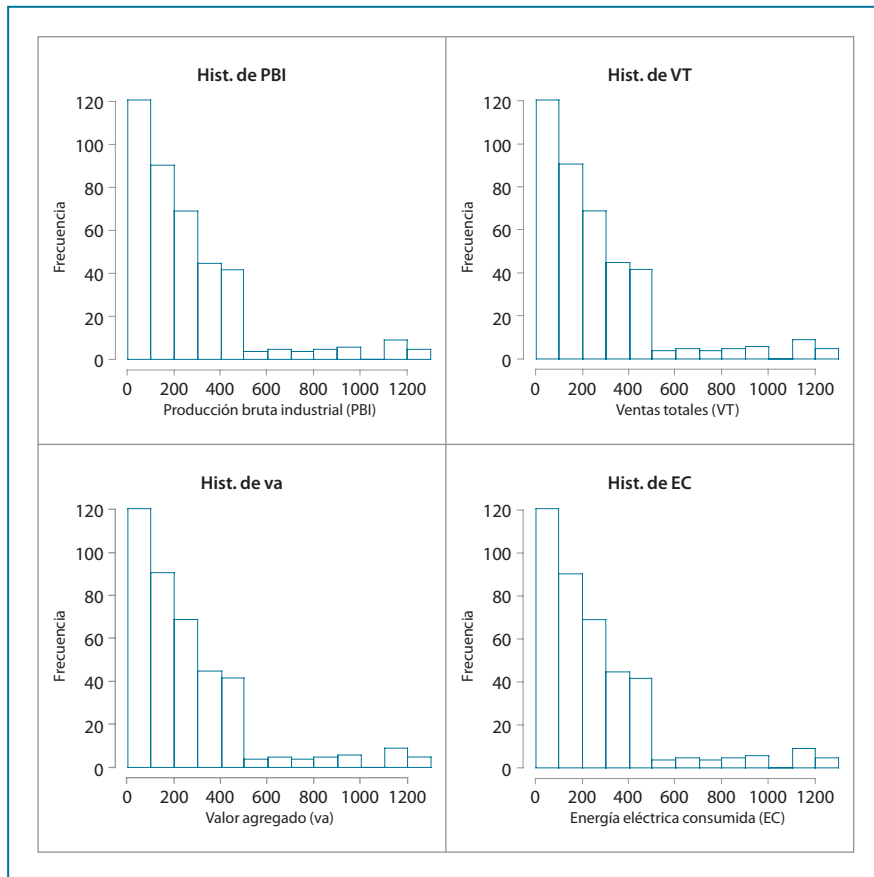
Para las variables que no son comparables debido a su propia definición, clasificación o rangos de tiempo diferentes se hace una descripción gráfica de los comportamientos que pueden visualizarse, con el fin de presentar de manera general comportamientos de utilidad en análisis posteriores.

RESULTADOS PRELIMINARES

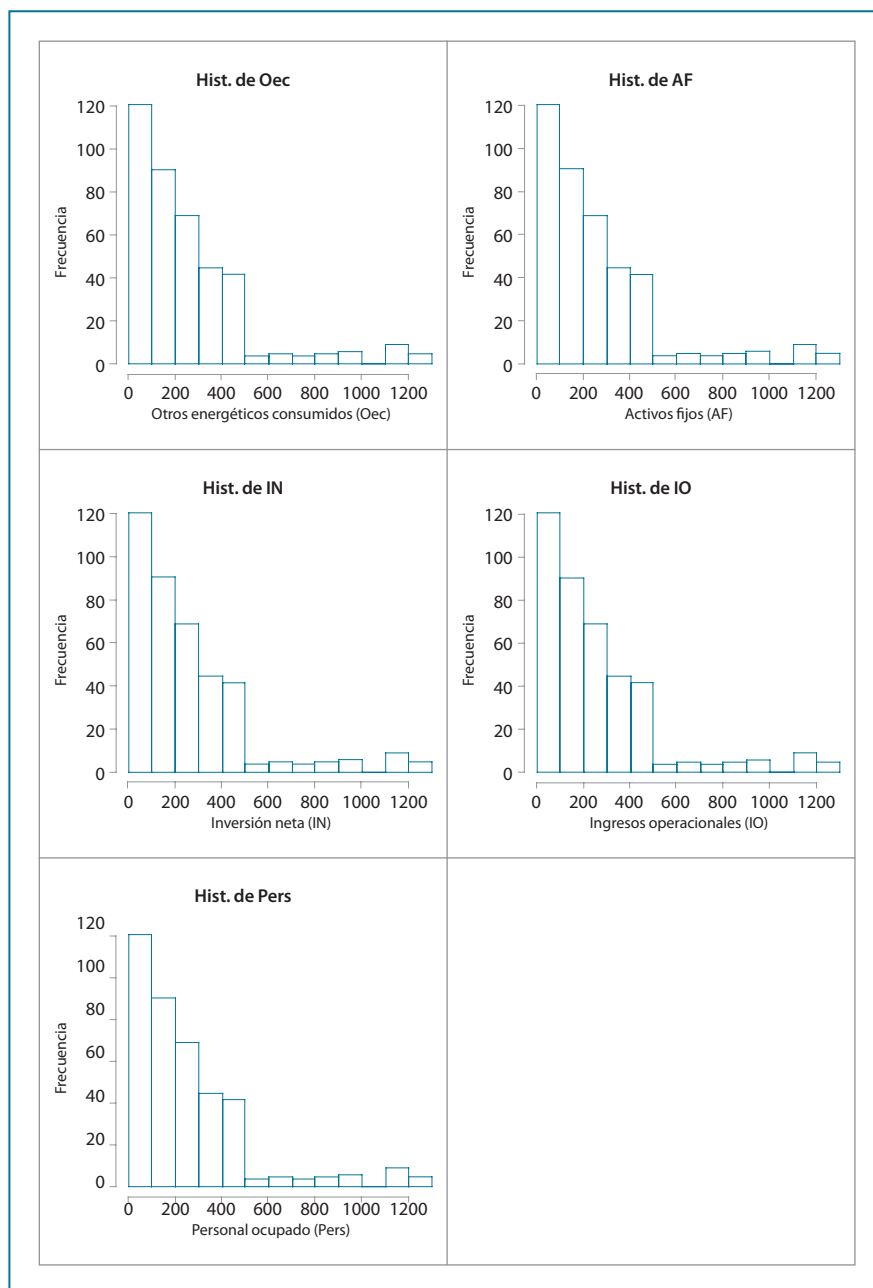
Los resultados obtenidos en cada una de las encuestas analizadas por los diferentes métodos mencionados permiten ilustrar algunas situaciones. En el análisis descriptivo se encontraron varias situaciones relacionadas con la distribución de las variables. En la EAM, por ejemplo, todas presentaban sesgo a la derecha como se aprecia en los histogramas del Cuadro 2.

Cuadro 2

Histogramas de datos de la EAM, 1992-2000



Continúa



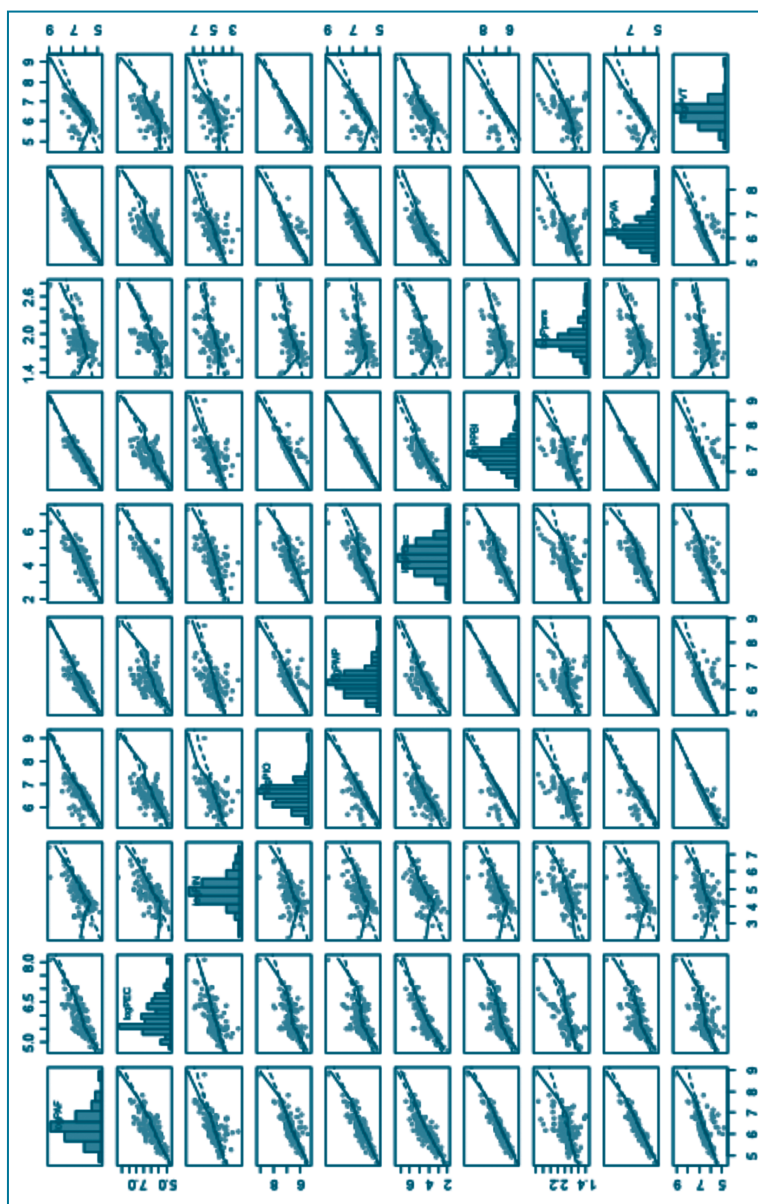
Fuente: elaboración propia.

Al analizar las relaciones entre las variables de la EAM, en la matriz de dispersiones (Gráfico 5), se observa que, en vista de los sesgos y la gran dispersión de los datos entre las variables relacionadas, se ha requerido una transformación logarítmica. Además, se aprecian (aunque de una manera poco legible por el tamaño de la figura) las formas de relación entre las variables promediadas por el número de empresas; por esto se identificaron los nombres de las variables con una letra *P* inicial.

Las variables consideradas son: activos fijos: AF (PAF es el promedio de los activos fijos y así sucesivamente para las demás variables); energía eléctrica consumida: EC; inversiones: IN; ingresos operacionales: IO; valor de materias primas: MP; otros energéticos consumidos: Oec; producción bruta industrial: PBI; número de personas: Pers; valor agregado: VA, y ventas totales: VT. De manera preliminar, se pueden inferir relaciones de tipo lineal entre los logaritmos de algunas variables, mientras en otras no es tan clara la relación inicial.

Para ilustrar otro tipo de análisis al que se han sometido los datos de la EAM se muestra el dendograma correspondiente al análisis de *cluster* de las variables consumo de energía eléctrica (EC) y otros energéticos consumidos (Oec) en 1992 (Gráfico 6). Los diferentes agrupamientos de los grupos CIIU muestran cuáles de ellos tienen características similares frente al consumo energético en ese año. Cabe anotar que a través de los años siguientes los agrupamientos no se mantienen de igual manera, en términos generales.

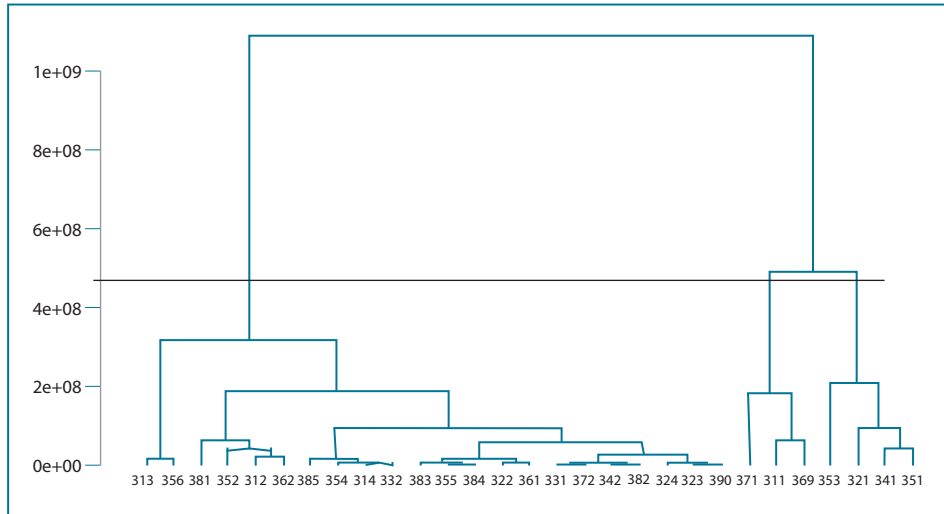
Gráfico 5
Matriz de dispersiones de los logaritmos de las variables de la EAM



Fuente: elaboración propia.

Gráfico 6

Dendograma de agrupación por CIU del consumo energético EC y Oec para 199



Fuente: elaboración propia.

Para 1992, el consumo energético promedio por empresa de cada grupo CIU tendía a agruparse en varios niveles según la cantidad de kilovatios de energía eléctrica consumida y los pesos consumidos en otros energéticos (Gráfico 6). Así, en el nivel aproximado de $2.0e+0,8$ presentan características similares los grupos 353 (refinerías de petróleo), 371 (industrias básicas de hierro y acero) y el grupo conformado por el 313 (industrias de bebidas) y el 356 (fabricación de productos plásticos).

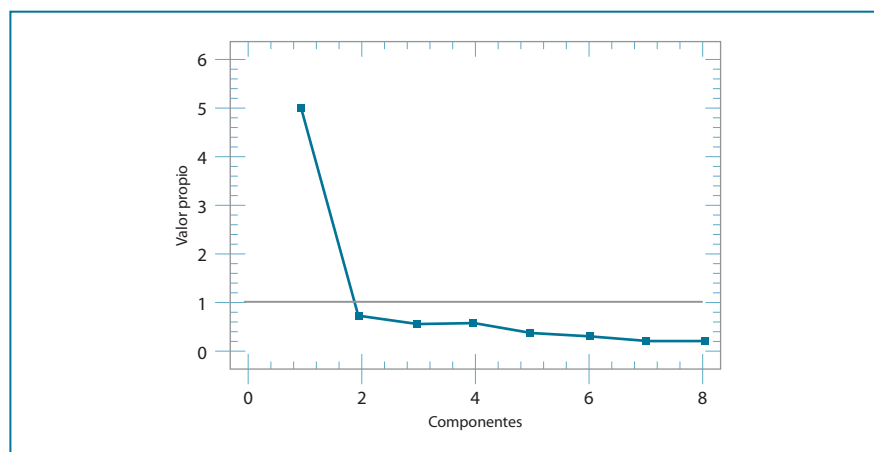
En las encuestas, dentro del proceso de identificación de variables relacionadas con el aspecto energético, desde el punto de vista de consumo o ahorro, se requiere conocer agrupamientos que permitan delimitar los rangos de valores de importancia dentro de los cuales se encontrarán los establecimientos que presentan algunos patrones de desempeño. En el capítulo 4 de la EDT I, sobre tipificación de la innovación tecnológica, se hallaron variables que indagan sobre la importancia que se da en el establecimiento industrial a los objetivos de la I+D, y que van dirigidos a mantener o acrecentar participación en el mercado, aumentar márgenes de utilidad, acceder a mercados masivos o con demanda de rápido cre-

cimiento, mejorar condiciones de trabajo y seguridad industrial, reducir daños en el entorno, disminuir costos de producción, mejorar calidad mediante implementación de control de calidad de productos o a través de un sistema de aseguramiento.

Dichas variables se identifican como IV114_01 a IV414_08. Puesto que el objetivo de disminuir costos de producción es por ahora la única variable que podría tener incluido el aspecto de consumo energético, este se analizará de manera particular (variable IV414_06). El ACP inicial arroja como resultado que una componente de las ocho variables originales (Gráfico 7) permite explicar un 62,84% de la variabilidad, mientras que dos componentes permiten una explicabilidad del 71,86%, lo cual es muy buen indicador según lo expresado anteriormente.

Gráfico 7

Diagrama de criba

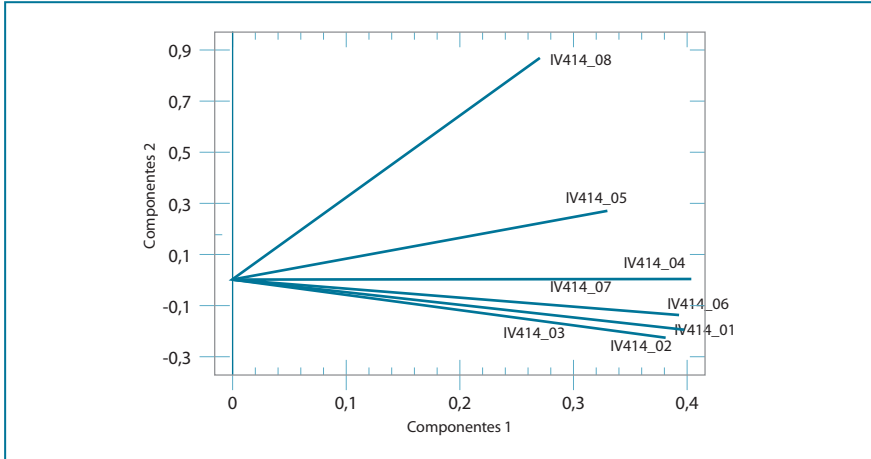


Fuente: elaboración propia.

No obstante, en la parte inferior del Gráfico 8 se observa un grupo de líneas que representan variables colineales y que ofrecen información redundante, entre las que se encuentran las variables IV414_01 (mantener o acrecentar participación en el mercado) e IV414_06 (disminuir costos de producción). Al continuar el proceso de disminución de dimensionalidad aparecen nuevas colinealidades.

Gráfico 8

Componentes principales de la variada 4.8

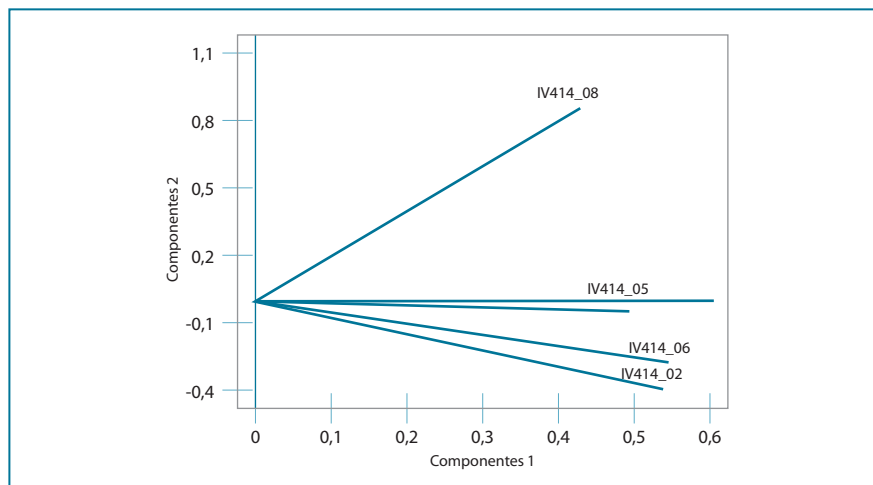


Fuente: elaboración propia.

En vista de nuestro interés por analizar la variable IV414_06, se retira la IV414_01 del proceso, y se obtienen como resultado cuatro variables que, con una componente, ofrecen una explicabilidad del 64,63% de la variabilidad del sistema; y alcanzan con dos componentes un 81,22%. De ahí que sea la variable sobre disminución de costos de producción una de las más importantes. Los objetivos que han resultado ser más importantes se refieren al aumento de márgenes de utilidad, la reducción de daños en el entorno, la disminución de costos de producción y el mejoramiento de la calidad mediante la implementación del sistema de aseguramiento de la calidad (Gráfico 9).

Gráfico 9

Variables resultantes en las componentes principales de la variada 4.8



Fuente: elaboración propia.

El cálculo del IVI de esta variada arroja valores resultantes entre 0,0 y 2,0. Para el análisis de *cluster* se realizan seis estrategias de agrupamiento que se encuentren disponibles (como los métodos del centroide, de la mediana, del vecino más próximo, del vecino más alejado, de los promedios y el de Ward) y se estiman los de intervalos de confianza correspondientes para identificar los agrupamientos.

Se encontraron cuatro agrupamientos con los siguientes valores: el grupo 1, que contiene los establecimientos con IVITO superiores a 1,59; el grupo 2, que contiene los establecimientos con IVITO entre 0,85 y 1,59; el grupo 3, que está entre 0,50 y 0,85, y el grupo 4, que contiene establecimientos con IVITO por debajo de 0,50. Los establecimientos del grupo 1 se caracterizan por tener un tamaño más grande. Se encuentran allí las empresas de petróleo y empresas que pertenecen a todos los subsectores industriales y la mayor parte de las industrias altas en conocimiento. Cabe anotar que estas industrias se encuentran en todos los agrupamientos. De manera similar se procede con todas las variadas. No obstante la importancia de las variables de la variada que se acaba de describir, cuando se somete la totalidad de variables resultantes a un nuevo ACP, el capítulo 4

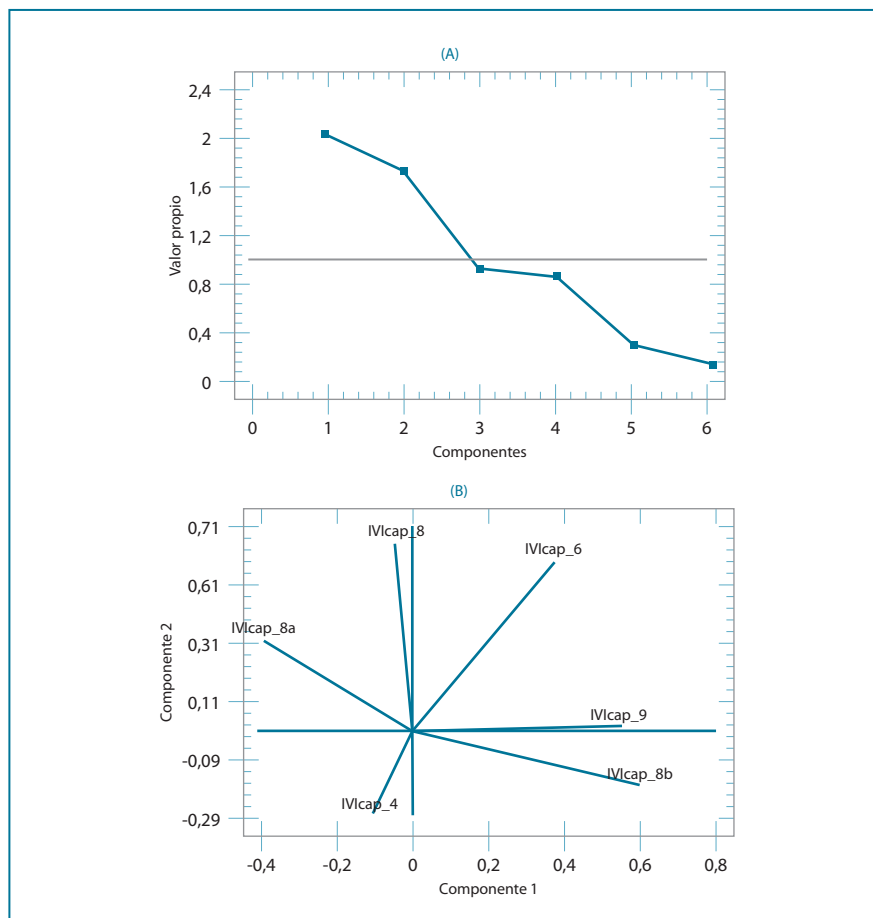
sólo queda representado por las variables IV401A, sobre la introducción de nuevos productos en el mercado, y la variable IV419, sobre variación de la inversión en actividades innovadoras.

Al finalizar el proceso se lograron identificar 16 variables de las 245 consideradas, las cuales combinadas linealmente en una, dos o tres componentes, cuyos valores propios son mayores que uno, ofrecen una mejor explicabilidad de todo el sistema, así: 39,932% de explicabilidad con una combinación lineal, un 61,886% con dos componentes principales y 72,241% con tres componentes principales. Posteriormente, se calcula el IVI para los capítulos resultantes, a los cuales también se les hace de nuevo un ACP, de donde resulta redundante uno de los capítulos, el cual se ha extraído para el resultado final de reducción de dimensionalidad. En el Gráfico 10b se muestran, de los nueve capítulos analizados, aquellos cuyas variables (algunas de ellas) han permanecido para explicar el sistema original en un porcentaje importante.

Puede decirse que la información ofrecida por la EDT I está ampliamente caracterizada por los capítulos de las variables del Gráfico 10b. Aunque varios capítulos no aparecen representados en los resultados, las variables resultantes logran una descripción general, pues se relacionan con la realización de proyectos de I+D, la capacitación tecnológica en procesos productivos y administrativos, la variación de la inversión en actividades innovadoras, el mejoramiento tecnológico de los procesos como objetivo de proyectos de I+D, el uso de fuentes internas y de cooperación internacional para financiar proyectos de I+D, las inversiones en plantas piloto y en prototipos, la preocupación por estar al día en la tecnología y las limitaciones en capacidades para ello, las dificultades en el mercado de capitales para proyectos de I+D, el conocimiento de algunos incentivos para I+D y el conocimiento de algunas normas de certificación.

Gráfico 10

Diagramas de ACP de capítulos resultantes de la EDT I: (a) diagramas de criba de valores propios y (b) pesos de componentes para los IVI de los capítulos



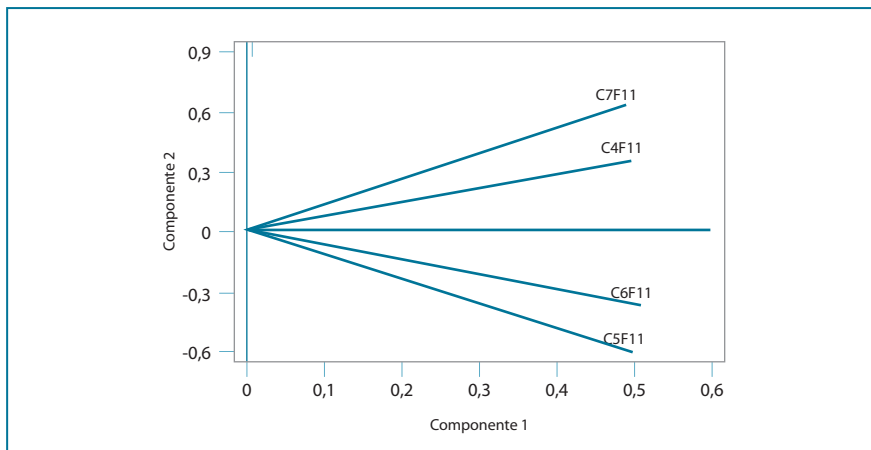
Fuente: elaboración propia.

En la EDT II se explicitan más detalles sobre la importancia de los objetivos de la innovación y cuáles se plantean específicamente; incluso se indaga por objetivos dirigidos a la reducción del consumo de energía, en términos de introducción de bienes o servicios para la empresa, el mercado nacional o el mercado internacional; la introducción de procesos nuevos para la línea de producción principal o complementarias, o la introducción de nuevas formas organizativas o nuevas formas de comercialización.

Al aplicar esta vez el proceso de ACP a las ocho variables relativizadas de la variada de esta encuesta (Gráfico 11), se obtiene una componente de cuatro variables que logra explicar el 77,27% de la variabilidad, y dos componentes que logran una explicabilidad del 86,06%.

Gráfico 11

ACP de la variada sobre objetivos de la innovación en la EDIT II



Fuente: elaboración propia.

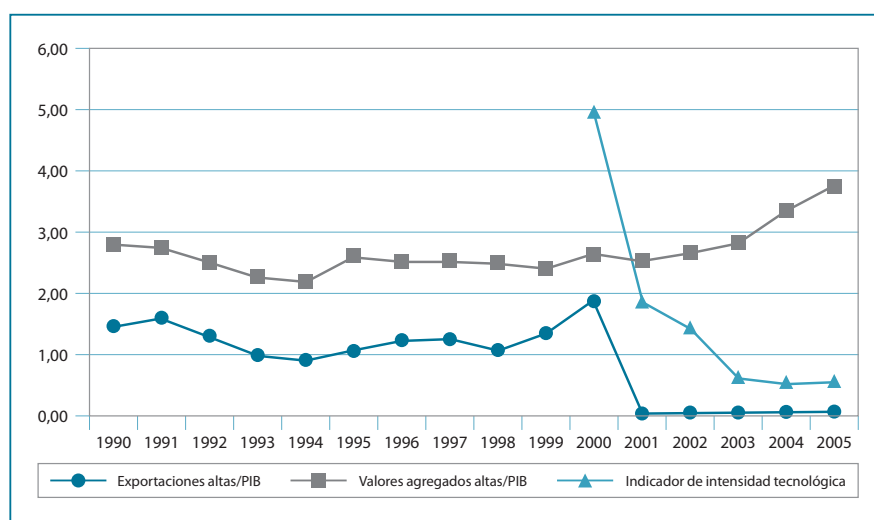
Esta variada es explicada por el comportamiento de las variables C4F11, C5F11, C6F11 y C7F11, que se refieren a la introducción de productos al mercado internacional, a procesos en las líneas de producción principal y secundarias, a la introducción de nueva organización y a nuevas formas de comercialización, respectivamente.

Por otro lado, respecto al comportamiento del sector industrial, en la EAM se puede identificar el caso de las exportaciones de las empresas de alto conocimiento. Entre 2006 y 2008, las exportaciones de este sector se incrementaron un 39% y “se destacaron las partes de aviones distintas de hélices, trenes de aterrizaje, pilas y acumuladores eléctricos, transformadores eléctricos, suturas quirúrgicas y vacunas” (Proexport, 2009: 6). Esta cifra, que pareciera muy importante, sólo refleja el incipiente crecimiento del sector —desde prácticamente cero, en 2000— a 1,4% del total exportado en 2008, según reportan los informes de Proexport y el Ministerio de

Comercio, Industria y Turismo (2008 y 2009). En el Gráfico 12 aparece, además, el comportamiento de la intensidad tecnológica, que se define como el gasto en tecnología sobre producción bruta, y se ha registrado únicamente a partir del 2000. Muestra un comportamiento fuertemente descendente.

Gráfico 12

Comportamiento de las exportaciones, valor agregado e intensidad tecnológica en el sector de alto conocimiento

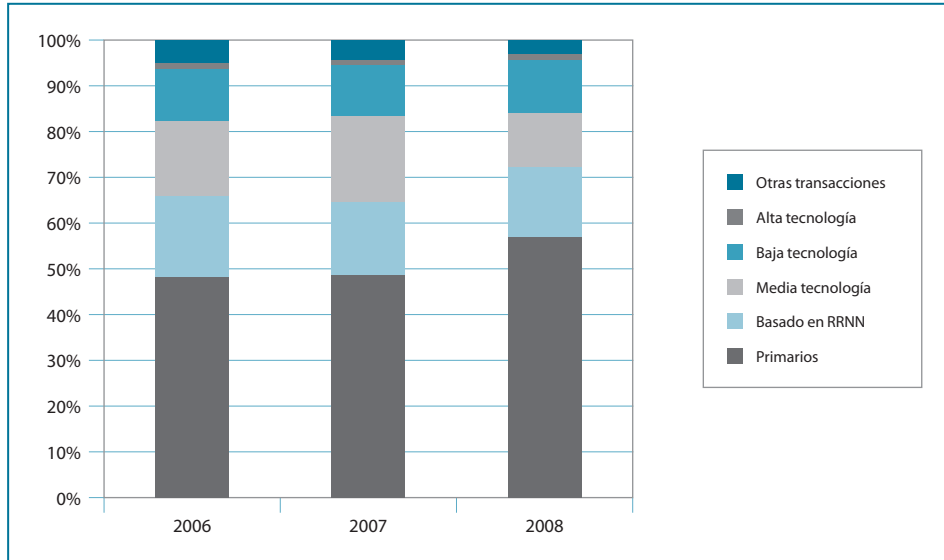


Fuente: elaboración propia con base en datos de DANE.

El comportamiento de las exportaciones de las industrias altas en conocimiento poco ha cambiado en los últimos años y se ilustra en el Gráfico 13 (Proexport, 2008), el cual se destaca por su incipiente participación en el total de las exportaciones, como la delgada área oscura en la parte superior del gráfico, luego del área correspondiente a “otras transacciones”.

Gráfico 13

Exportaciones colombianas por intensidad tecnológica



Fuente: Proexport (2008), basado en DANE-DIAN. Cálculos del OEE, Mincomercio.

CONCLUSIONES

La cantidad y la calidad de la información disponible en los países desarrollados y de la OCDE, más específicamente, han posibilitado una gran cantidad de estudios sobre desarrollo, innovación y crecimiento y han permitido formular teorías al respecto. La información disponible en Colombia no ofrece datos regulares suficientes para construir series temporales que puedan abordarse desde las metodologías econométricas. En su lugar debe recurrirse a técnicas estadísticas de carácter descriptivo y de análisis multivariado.

Las diferencias estructurales en las bases de datos de las encuestas colombianas de innovación dificultan de gran manera el procesamiento, la comparación y el análisis de los datos, lo cual ocasiona retardos importantes en los procesos de investigación y disminuye la producción de conocimiento pertinente, en comparación con estudios similares en otros

países (especialmente de la OCDE), donde la estructura de las bases de datos y el acceso a la información están diseñados para facilitar la investigación permanente.

Las características de la información disponible en las encuestas de innovación realizadas en el país no permiten cruzar directamente las variables ni los establecimientos encuestados, con la consecuente pérdida de información. Esto hace que los establecimientos deban agruparse para su análisis y que se pierda mucha de la información que se ha recolectado.

Una de las condiciones importantes para reducir la brecha de conocimiento es generar información pertinente y facilitar el acceso a microdata. Este es un asunto que ha sido resuelto en gran medida en varios de los países de la OCDE. En Colombia, por el contrario, este es todavía un problema por resolver, con impactos importantes en la cantidad y calidad de la investigación y el conocimiento producido a partir de los resultados de las encuestas oficiales.

Las bases de datos de la EDT I y de la EDIT II presentan gran cantidad de errores como datos por fuera de rango, datos mal registrados y registros con información no disponible. En la EAM, muchas variables no presentan continuidad en el tiempo, algunas se agregan en ciertos períodos y otras se desagregan. Adicionalmente, la diferencia de clasificaciones de los grupos CIIU rev. 2 y rev. 3, afecta a todas las encuestas y hace que los datos tengan que ser procesados de manera independiente, pues ambas clasificaciones no son totalmente comparables.

Otro asunto que se revela durante el procesamiento de la información consiste en que las encuestas presentan mucha información redundante, lo cual, a su vez, introduce ruido al proceso y dificulta el análisis. Para superar esta dificultad, se ha recurrido a una herramienta como el IVI, que acompañada de un ACP, elimina el ruido introducido por variables que portan información redundante y permite identificar grupos de empresas con ciertos rasgos característicos dentro de una encuesta. En este caso, se detectó el grupo de empresas que caracterizan en gran parte la muestra de la EDT I. No se concluye aún sobre su potencial para hacer comparaciones entre diferentes encuestas. No obstante, también en su aplicación a la EDIT II, el ACP demostró ser una herramienta muy útil

en la disminución de la dimensionalidad de un sistema de variables, sobre todo si se combina con el IVI.

La EAM proporciona algunas variables relacionadas directamente con el desempeño del negocio, como el valor agregado, las ventas y la producción bruta industrial, con las que se puede visualizar el proceso completo de innovación; sin embargo, no contiene información importante sobre la I+D, ni sobre otros indicadores de desempeño del negocio que son valiosos para analizar el impacto de la innovación, como las exportaciones.

Los datos sobre el uso energético en la industria colombiana se encuentran agregados en subsectores, mas no en grupos del CIIU ni de empresas. Tampoco se encuentran disponibles por producto ni por proceso de producción. Esto dificulta el análisis de descomposición de factores como la eficiencia energética, planteados por Kumar Ray y Sudhakara Reddy (2008).

Finalmente, se observa que la metodología del IVI ofrece un significativo potencial para analizar relaciones de interés entre variables a partir de los datos de las encuestas disponibles, y ello contribuye a superar algunos de los problemas de las encuestas identificados en el trabajo.

REFERENCIAS

- Alcántara, V. y Padilla, E. (2003). “Key” sectors in final energy consumption: an input-output application to the Spanish case. *Energy Policy*, 31 (15), 1673-1678.
- Archibugi, D. y Coco, A. (2005). Measuring technological capabilities at the country level: A survey and a menu for choice. *Research Policy*, 34 (2), 175-194.
- Climont, F. y Pardo, A. (2007). Decoupling factors on the energy-output linkage: The Spanish case. *Energy Policy*, 35 (1), 522-528.
- Cortijo, F. (2001, noviembre). *Técnicas no supervisadas: métodos de agrupamiento*. Notas del curso Reconocimiento de Formas, Universidad de Granada, España.
- Delgado Alaminos, J. y Sánchez Vizcaíno, G. (1999). Caracterización de las ramas industriales en España según su comportamiento innovador. *Ekonomiaz: Revista Vasca de Economía*, (44), 256-277.
- Departamento Administrativo Nacional de Estadística (DANE), Departamento Nacional de Planeación y Colciencias (2005). *Innovación y desarrollo tecnológico en la industria manufacturera: Colombia 2003-2004*. Bogotá.

- Díaz de Rada, V. (2002). *Análisis exploratorio de datos*. Recuperado el 2 de septiembre de 2007, de <http://www.udc.es/dep/ecoapl2/esteco1/tecnicas/bea/R-ANALISIS%20EXPLORATORIO.PDF>.
- Ferguson, R.; Wilkinson, W. y Hill, R. (2000). Electricity use and economic development. *Energy Policy* (28), 923-934.
- Figueras, S. y Gargallo, P. (2003). *Análisis de conglomerados o cluster*. Recuperado el 3 de abril de 2007, de <http://www.5campus.org/leccion/cluster>.
- Gondar, J. E. (2000). *Análisis de cluster*. Recuperado en octubre 2007, de <http://www.estadistico.com/arts.html?20001023>.
- González, J. J. (2006). *Patrones de consumo aparente de energías modernas y actividad económica en América Latina durante el siglo XX*. Recuperado el 2 de noviembre de 2009, de <http://www.helsinki.fi/iehc2006/papers3/Jofre.pdf>.
- Granger, C. W. J. (1969). Investigating causal relationship by econometric models and cross-spectral methods. *Econometrica*, 37, 424-438.
- Guttormsen, A. G. (2004). *Causality between energy consumption and economic growth*. Agricultural University of Norway, Department of Economics and Resource Management. 1-31.
- Hannesson, R. (2002). Energy use and GDP growth, 1950-97. *OPEC Review*, 26 (3), 215-233.
- Kumar Ray, B. y Sudhakara Reddy, B. (2008). *Understanding industrial energy use: Physical energy intensity changes in Indian manufacturing sector*. Mumbai: Indira Gandhi Institute of Development Research (IGIDR).
- Lema, Á. (2005). *Elementos de estadística multivariada* (2a ed.). Medellín: Universidad Nacional de Colombia-sede Medellín.
- Manrique, J. y Robledo, J. (2008, 14-15 de agosto). *Relación entre el uso energético e innovación tecnológica en Colombia en el período 1992-2000*. Documento presentado en el I Congreso Internacional de Gestión Tecnológica e Innovación, Bogotá, Colombia.
- Masih, A. M. M. y Masih, R. (1997). On the temporal causal relationship between energy consumption, real income, and prices: Some new evidence from Asian-energy dependent NICs Based on a multivariate cointegration/vector error-correction approach. *Journal of Policy Modeling*, 19 (4), 417-440.
- Mollineda, R. (2001). *Técnicas de agrupamiento jerárquico para la selección de prototipos y clasificación basadas en distancias: uso en cadenas cíclicas*. Valencia: Universidad Politécnica de Valencia.
- Narayan, P. y Prasad, A. (2008). Electricity consumption-real GDP causality nexus: Evidence from a bootstrapped causality test for 30 OECD countries. *Energy Policy*, 36 (2), 910-918.
- Narayan, P. y Smyth, R. (2005). Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. *Energy Policy*, 33 (9), 1109-1116.

- Pla, L. E. (1986). *Análisis multivariado: método de componentes principales*. Washington: OEA.
- Proexport y Ministerio de Comercio Industria y Turismo de Colombia (2008). *Informe de exportaciones, turismo e inversión octubre de 2008* [presentación]. Bogotá: Ministerio de Comercio Industria y Turismo de Colombia. 58 diap.
- (2009). *Informe de exportaciones febrero de 2009* [Presentación]. Bogotá: Ministerio de Comercio Industria y Turismo de Colombia. 32 p. Recuperado el 20 de septiembre de 2009, de <http://www.google.com.co/search?hl=es&q=exportaciones+por+intensidad+tecnologica+2006+proexport&btnG=Buscar&meta=&aq=f&oq=>.
- Robledo, J.; Gómez, F. A. y Restrepo, J. F. (2008, 14-15 de agosto). Relación entre capacidades de innovación tecnológica y desempeño empresarial en Colombia. En *Memorias del Primer Congreso Internacional de Gestión Tecnológica e Innovación*, Universidad Nacional de Colombia, Bogotá, Colombia.
- Sadorsky, P. (2009). Renewable energy consumption and income in emerging economies. *Energy Policy*, 37 (10), 4021-4028.
- Salazar, D. C. (2006). *Programa uso eficiente de la energía en pymes*. Recuperado el 2 de noviembre de 2009, de http://www.si3ea.gov.co/Si3ea/Documentos/ciure/Documentos/Sexta%20sesion/Presentacion_MCIT.pdf.
- Salazar, M. y Vargas, M. (1998). *Encuesta sobre desarrollo tecnológico en la industria colombiana*. Bogotá: Departamento Nacional de Planeación, Unidad de Desarrollo Empresarial, División de Desarrollo Tecnológico.
- Shiu, A. y Lam, P.-L. (2004). Electricity consumption and economic growth in China. *Energy Policy*, 32 (1), 47-54.
- Unidad de Planeación Minero Energética (UPME), (2006). *Boletín estadístico de minas y energía 1999-2005*. Bogotá: Ministerio de Minas y Energía de la República de Colombia. Recuperado el 2 de noviembre de 2009, de http://www.upme.gov.co/Docs/boletin_estadistico_1999_2005.pdf.
- Wolde-Rufael, Y. (2009). Coal consumption and economic growth revisited. *Applied Energy*, 87 (1), 160-167.
- Worrell, E. et al. (2008). *World best practice energy intensity values for selected industrial sectors*. Madison: Ernest Orlando Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division.
- Yoo, S.-H. y Ku, S.-J. (2008). Causal relationship between nuclear energy consumption and economic growth: A multi-country analysis. *Energy Policy*, 37 (5), 1905-1913.

ANEXO

EJEMPLOS DE VARIABLES CON DATOS ERRÓNEOS EN LA EDT I

Variable	Descripción
IV420_01	Impacto de la inversión en actividades innovadoras: los costos unitarios de producción (sí=1; 2=no)
IV420_02	Impacto de la inversión en actividades innovadoras: la productividad del establecimiento (sí=1; 2=no)
IV420_03	Impacto de la inversión en actividades innovadoras: la participación en el mercado nacional (sí=1; 2=no)
IV420_04	Impacto de la inversión en actividades innovadoras: las exportaciones (sí=1; 2=no)
IV420_05	Impacto de la inversión en actividades innovadoras: la participación de productos innovados en las ventas totales de productos elaborados por el establecimiento (sí=1; 2=no)
IV420_06	Impacto de la inversión en actividades innovadoras: la participación de productos innovados exportados en el total de exportaciones del establecimiento (sí=1; 2=no)
V511A_03	Estado de la última maquinaria adquirida (1=nueva; 2=usada)
V511B_02	Sistema de operación de la anterior maquinaria (1>manual; 2=mecánico; 3=eléctrico)
V511B_03	Estado de la anterior maquinaria adquirida (1=nueva; 2=usada)
VI609C2	Proyecto de desarrollo experimental realizado por otro establecimiento perteneciente a la misma empresa: nacional o extranjera (nacional=1; extranjera=2)
VI609E2	Proyecto de desarrollo experimental realizado por una firma consultora: nacional o extranjera (nacional=1; extranjera=2)
VI609G2	Proyecto de desarrollo experimental realizado por el cliente: nacional o extranjero (nacional=1; extranjera=2)
VI609H2	Proyecto de desarrollo experimental realizado por otra empresa: nacional o extranjera (nacional=1; extranjera=2)
VIII820B02	Origen de la fuente de información de actualización tecnológica del establecimiento (1=nacional; 2=extranjera): firma consultora (sí=1; no=2)
VIII820B03	Origen de la fuente de información de actualización tecnológica del establecimiento (1=nacional; 2=extranjera): proveedor (sí=1; no=2)

Continúa

Variable	Descripción
VIII820B04	Origen de la fuente de información de actualización tecnológica del establecimiento (1=nacional; 2=extranjera): otra empresa (sí=1; no=2)
VIII820B05	Origen de la fuente de información de actualización tecnológica del establecimiento (1=nacional; 2=extranjera): ferias, exposiciones (sí=1; no=2)
VIII820B06	Origen de la fuente de información de actualización tecnológica del establecimiento (1=nacional; 2=extranjera): bancos de datos, publicaciones técnicas (sí=1; no=2)
IX900	No especificada en el manual
IX907	¿Qué imagen tiene de la delegatura para la propiedad industrial de la SIC? (1=buena; 2=regular; 3=mala; 4=no sabe)
IX915B01	No especificada en el manual
IX915B02	No especificada en el manual
IX915B03	No especificada en el manual
IX915B04	No especificada en el manual
IX917B01	No especificada en el manual
IX917B02	No especificada en el manual
IX917B03	No especificada en el manual
IX917B04	No especificada en el manual

Fuente: elaboración propia basándose en la Encuesta sobre Desarrollo Tecnológico en el Establecimiento Industrial Colombiano 1996, Colciencias y DNP, Colombia.

Capítulo 13

Los datos, la información y el conocimiento sobre la innovación en Colombia y sus perspectivas de desarrollo

*Jorge Robledo Velásquez
Florentino Malaver Rodríguez*

En este capítulo se sintetizan los principales hallazgos, conclusiones y propuestas que emergen de los aprendizajes y reflexiones generados por el Proyecto DESCUBRIMIENTO. Para empezar, debe reiterarse que el dato es el insumo básico del proceso de *Knowledge Discovery in Databases* (KDD); las técnicas de análisis, su herramienta fundamental, y el conocimiento, su producto. Describir el proceso de descubrimiento en estos términos es fácil; su aplicación exitosa, enormemente compleja.

En torno a este proceso se viene acumulando conocimiento y experiencia, sobre todo en contextos empresariales, en lo que usualmente se conoce como *inteligencia de negocios* (*business intelligence*), gracias a las grandes posibilidades abiertas por las tecnologías de la información y la comunicación para la captura, transmisión, depuración, contextualización, almacenamiento, procesamiento, análisis y visualización de los datos, la información y el conocimiento. En el Proyecto DESCUBRIMIENTO se abordó el desafío de avanzar en la aplicación de este proceso, en terrenos

poco explorados y utilizados para alimentar la formulación de políticas y estrategias a partir de los datos (cada vez más abundantes) de las encuestas¹ disponibles con información relativa a los estados y dinámicas de los sistemas de innovación colombianos (regional, sectorial y nacional).

El trabajo mostró que el desafío es enorme e ineludible; no sólo porque el país y sus instituciones incurren en altos costos por la aplicación, procesamiento y distribución de las encuestas y sus resultados, sino porque la innovación y sus procesos sistémicos de generación y transformación del conocimiento científico y tecnológico en riqueza económica, bienestar social y desarrollo humano adquieren cada vez mayor relevancia en lo que se ha dado llamar la economía y la sociedad del conocimiento.

El Proyecto DESCUBRIMIENTO es de naturaleza exploratoria de un conjunto relativamente alto de modelos, métodos e instrumentos aplicados a varios temas de análisis relativos a la innovación; es interdisciplinario, por las contribuciones provenientes de distintas disciplinas de la ingeniería, la economía, la administración, la estadística, la física y la sociología, por mencionar las más relevantes, y es interinstitucional, por el origen de los esfuerzos y apoyos que permitieron integrar el equipo investigativo y adelantar las actividades propuestas. Los resultados y conclusiones, en concordancia con el carácter exploratorio del Proyecto, son tentativos y abiertos; sin embargo, generan percepciones y propuestas de trabajo futuro que contribuyen a abrir la discusión y posibilidades de ampliación y profundización investigativa. También permiten señalar debilidades tanto en el diseño y aplicación de las encuestas como en la entrega de sus resultados, que conducen a algunas recomendaciones para considerar en versiones futuras.

En este capítulo se presentan las conclusiones, algunas recomendaciones y perspectivas de trabajo futuro. Para ello el contenido se divide

¹ Aquí se incluye, por brevedad en la expresión, la base de datos ScienTI, que no tiene las características de encuesta que poseen la Primera Encuesta Colombiana de Innovación (EDT), la Segunda Encuesta Colombiana de Innovación (EDIT II), la Encuesta Anual Manufacturera (EAM) y la Encuesta de Innovación y Desarrollo Tecnológico de Bogotá y Cundinamarca (EIByC).

en tres acápitos: (i) sobre los datos, su pretratamiento y almacenamiento; (ii) sobre las técnicas de análisis, y (iii) sobre la innovación en Colombia.

DATOS, PRETRATAMIENTO Y ALMACENAMIENTO: UN LARGO CAMINO POR RECORRER

Por ser el insumo básico del proceso de KDD, es necesario aproximarse a la noción de dato, siendo conscientes de su complejidad y estando alerta sobre a su aparente simplicidad y el descuido con que muchas veces se usa. Esto es más importante aún si se considera que en el Proyecto DESCUBRIMIENTO construimos una bodega de datos con datos provenientes de fuentes múltiples y heterogéneas.

Para empezar, el dato sólo se puede entender como un elemento inserto, pero diferenciado de un contexto dado. Entonces, en la interacción dato-contexto, el dato se dota de significado, sentido y alcance. Esto, por supuesto, no puede suceder sin referencia a los acuerdos sociotécnicos *de jure* o *de facto* que llamamos normas y estándares, tema desarrollado por Charum en el Capítulo 1. El trabajo emprendido en el Proyecto DESCUBRIMIENTO está apoyado en objetos y procedimientos informáticos; por lo tanto, las normas y estándares son una necesidad fundamental toda vez que, como afirma Charum, sin normas y estándares es imposible asegurar el funcionamiento técnico adecuado de los sistemas de información. Conscientes de ese hecho, las comunidades internacionales que trabajan sobre encuestas y análisis de la innovación se esfuerzan por definir normas y estándares que sustenten un trabajo colectivo y colaborativo coherente. Sus resultados se evidencian, por ejemplo, en el *Manual de Oslo*, de la Organización para la Cooperación y el Desarrollo Económico (OCDE), y en las propuestas de manual de la Red de Indicadores de Ciencia y Tecnología (RICYT), relativas a la innovación.

Al respecto, una conclusión que emerge del Proyecto es el incipiente desarrollo de referentes normativos para el trabajo con los datos de las encuestas seleccionadas. Estos referentes parecen no existir, al menos de manera explícita, lo cual dificulta enormemente el trabajo con los datos.

Por ello cobra especial relevancia el esfuerzo por identificar y dotar de metadatos los datos de la bodega de datos del Proyecto, como datos sobre los datos originales que permiten construir contenidos a partir de estos.

En el mismo sentido, las relaciones entre los datos de una base de datos, interpretadas a la luz de los estándares informáticos sobre los que fueron construidas —incluidos los diagramas entidad-relación, las reglas de integridad y los atributos de las tablas, registros y campos de la base de datos—, contribuyen a definir el contexto que da significado, sentido y alcance a los datos. En el país, de nuevo, los investigadores usualmente enfrentan el reto de reconstruir estas relaciones debido a la forma desestructurada como normalmente les son entregados los datos. Hay casos en los cuales, como sucede con la base de datos ScienTI —componente del Grupo Latinoamérica y el Caribe (GrupLAC) y el Currículum Vitae Latinoamericano y del Caribe (CvLAC)— disponible para el Proyecto, la carencia de normas y estándares impidió casi por completo su utilización efectiva.

La solución adoptada en el Proyecto la presentan Jiménez, Villa y Rico en términos del metamodelo de la bodega de datos desarrollado en el Capítulo 2. El modelo permite, para cada variable de interés, incluir los metadatos necesarios para trabajar con los datos de manera confiable. Esto incluye datos simples del tipo: unidades de medida, valores máximos y mínimos, promedios y otros descriptores estadísticos básicos, así como metadatos más elaborados, en términos de descripciones gráficas de su distribución, a fin de facilitar una comprensión inicial del conjunto de datos disponibles y orientar la selección de las técnicas de análisis más adecuadas. De ese modo, el metamodelo propuesto apoya completamente el proceso de KDD en todas sus fases.

En el Proyecto se logra construir un objeto tecnológico que permite almacenar, pretratar y disponer de grandes cantidades de datos y metadatos, tanto de primer nivel como de segundo, en términos de las categorías descritas por Charum en el Capítulo 1. La bodega de datos permite almacenar, entonces, datos de primer nivel, provenientes de las fuentes originales y sus descriptores simples; metadatos o datos de segundo nivel, en forma de agregaciones, síntesis, clasificaciones y descripciones de sus

propiedades estadísticas más elaboradas; incluso metadatos más complejos, en la forma de documentos que presentan resultados de análisis más especializados hechos sobre los datos.

Una etapa importante del proceso, el pretratamiento de los datos (que incluye, entre otras operaciones, la remoción de ruido o errores y la identificación y manejo de datos faltantes y de *outliers*), distó de ser sencilla, dadas las grandes cantidades de datos y su heterogeneidad. Esto llevó a explorar el desarrollo y aplicación de técnicas específicas de depuración de datos (véase Jiménez y Amón, 2009). No obstante, aquí hay un frente de trabajo susceptible de desarrollo ulterior y de un alto impacto potencial sobre la productividad y calidad del trabajo realizado.

Con los datos depurados y almacenados, a la bodega de datos se le dotó de un mecanismo gestor y de estándares técnicos apropiados y explícitos, que le brindaron un atributo fundamental: la posibilidad para los investigadores de acceder a los datos y metadatos de manera conveniente y a distancia. En esta dirección, el Proyecto identificó la necesidad de contar con una interfaz que permitiera la consulta de la bodega a investigadores carentes del dominio de los lenguajes especializados de consulta que usualmente están disponibles. Este es otro de los trabajos futuros identificados por el Proyecto, que se justifica tanto por la necesidad de acceso descentralizado y a distancia a la bodega por parte de las personas autorizadas como porque el trabajo directo de los investigadores sobre la bodega le permite a esta, mediante los mecanismos de captura apropiados, mejorar la calidad y cantidad de los datos y metadatos disponibles.

El proceso de KDD va más allá de las posibilidades de almacenamiento y acceso a datos y metadatos confiables. En un segundo nivel incluye la aplicación de técnicas de análisis y la visualización de las relaciones y patrones subyacentes a los datos, sobre lo cual se presentarán las conclusiones y propuestas de trabajo futuro más adelante. Un tercer nivel se alcanza cuando los datos se convierten en conocimiento, al ser puestos en relación en un contexto formado por teorías que permiten, finalmente, su interpretación. Aquí es donde puede surgir el nuevo conocimiento o, al menos, la formulación de nuevos problemas, preguntas e hipótesis que conduzcan a nuevos descubrimientos. Este tema es abordado en el tercer

acápite del capítulo, donde se presentan las conclusiones que emergen de análisis y contextualización teórica de los resultados del trabajo.

El proceso completo, no obstante, no termina acá, como bien lo señala Charum. En sus palabras, “estas interpretaciones [que resultan del trabajo teórico] vuelven a poner en tensión a los datos y al sujeto que interpreta”. Como resultado, es posible una segunda mirada sobre los datos, su completitud y su capacidad de representación del objeto de análisis desde el marco teórico utilizado, al igual que sobre las normas y estándares que permiten dotarlos de significado, sentido y alcance.

Por ello, como resultado del ciclo completo, que lleva de los datos al conocimiento, pasando por las normas, los estándares y las teorías, y que regresa de nuevo a los datos para iniciar otra iteración en la espiral de construcción de conocimiento, en determinados momentos es importante que se replanteen y ajusten las normas y los estándares que intervienen en el proceso, e incluso en las encuestas o instrumentos específicos de recolección de datos sobre la innovación. Esto explica, por ejemplo, la evolución que ha tenido el *Manual de Oslo*, que ya va en su tercera versión desde su primera publicación en 1992, y las mejoras incrementales que se han hecho sobre las encuestas de innovación de algunos de los países y grupos de países de la OCDE.

Este proceso evolutivo y necesario parece no existir, sin embargo, en las encuestas colombianas de innovación, por lo menos en lo que respecta a las diferencias entre la primera y segunda encuestas.² Las diferencias en el diseño de las encuestas son disruptivas y, con frecuencia, generan serias dificultades en los análisis comparativos que, en principio, son uno de los principales objetivos de las encuestas periódicas. Esto se evidencia claramente en el trabajo de Robledo, Gómez y Restrepo, presentado en el Capítulo 4. Aun así, como este mismo trabajo revela, las dificultades pueden

² En el caso de la EIByC se intenta dar continuidad a la primera encuesta e incorporar los avances conceptuales y metodológicos surgidos de los distintos estudios de caracterización de la innovación en Colombia. En particular, se incorporaron como hipótesis algunos de los principales resultados del estudio de casos adelantado por investigadores de 10 universidades colombianas (Vargas, Malaver y Zerda, 2003).

ser superables en cierto grado, introduciendo marcos teóricos que permitan comparar los resultados a niveles relativamente altos de abstracción.

El carácter evolutivo de los datos, metadatos, normas, estándares y modelos teóricos también incide en el diseño de la bodega de datos. En este sentido, el metamodelo desarrollado en el Proyecto DESCUBRIMIENTO incluye consideraciones de diseño que configuran una bodega de datos flexible para posibilitar tanto su mejoramiento como su expansión, de forma que se puedan corregir y completar los datos y metadatos, e integrar nuevos datos y metadatos procedentes de las más recientes aplicaciones de las encuestas o de nuevas encuestas no incluidas anteriormente.

El Proyecto encuentra que la imposibilidad de cruzar los datos entre las distintas encuestas sigue siendo un problema sin resolver en Colombia y sobre el cual no se conoce todavía una propuesta de solución. Esta imposibilidad limita enormemente las posibilidades de análisis de los datos, al perderse importantes relaciones entre ellos. La solución generalmente adoptada, de agregar los datos al nivel de objetos de análisis posibles de construir en las diferentes encuestas, a partir de las clasificaciones normalizadas —típicamente, la propuesta del Código Industrial Internacional Uniforme (CIIU)—, es una solución parcial que impide un trabajo eficiente y genera pérdidas considerables de valiosa información.

A continuación se señalan algunos de los principales problemas identificados en el Proyecto, a partir del trabajo con las encuestas y bases de datos, y que Jiménez, Rico y Jiménez analizan con detalle en el Capítulo 3; también se presentan los retos identificados y las recomendaciones específicas para el diseño, aplicación y disposición de versiones futuras de las encuestas. Estos problemas y retos se concentran en seis aspectos de las encuestas y sus datos, así:

Estructura y semántica. Se detectó falta de validación de los datos o ausencia de estos, así como falta de integridad referencial respecto a algunas preguntas que no tenían respuestas asociadas y la situación inversa.

Metadatos. Con frecuencia, las bases de datos carecen de los metadatos necesarios para identificar las características básicas de los datos, por lo que fueron necesarios esfuerzos significativos para reconstruir los metadatos

faltantes a partir de los formularios de las encuestas y de material documental, no siempre disponibles.

Incumplimiento de reglas de integridad. Durante el proceso de depuración de datos se encontraron problemas que revelan falta de rigor en el cumplimiento de reglas de integridad, lo que ocasionó demoras en la construcción de la bodega de datos y, en algunos casos, imposibilidad para trabajar con la integridad referencial planteada en el diseño de esta.

Inconsistencias. El análisis de inconsistencias se llevó a cabo, inicialmente, mediante una revisión manual de los datos y, luego de la carga de los datos en la bodega de datos, realizando consultas mediante sentencias de lenguaje de consulta estructurado (SQL, por su sigla en inglés). Aquí se identificaron códigos no definidos o aplicados de forma incorrecta, así como preguntas sin respuestas y respuestas sin preguntas.

Complejidad de los datos y valores atípicos. La depuración de los datos mostró problemas de completitud de las bases de datos, originados ya sea en la recolección de los datos o en su transferencia al equipo de investigación. También se identificaron valores atípicos que pudieron haber tenido efectos significativos sobre los análisis, de no haberse identificado y corregido.

Problemas con ScienTI. Esta base de datos en particular presentó serias dificultades para su integración a la bodega de datos. No se pudo avanzar significativamente en esta dirección y los análisis con sus datos fueron muy limitados. Los problemas principales se refieren a: (1) la falta del modelo entidad-relación y de diccionarios de variables; (2), con frecuencia, nombres de tablas y campos ininteligibles, lo que imposibilitaba su identificación, y (3) la forma como ScienTI captura la información, que conduce a problemas de validez de los datos.

Los retos y recomendaciones, surgidos de la experiencia de los investigadores durante el desarrollo del Proyecto, y particularmente del análisis de Jiménez, Rico y Jiménez, se identifican en siete frentes de trabajo principales:

1. En el mejoramiento del proceso de gestión de los datos e información desde la etapa de diseño de las encuestas y su aplicación

hasta la entrega de los datos, metadatos y documentación complementaria a los investigadores. Al respecto, una posible solución podría ser generar interfaces de captura de datos con definiciones claras de las reglas de integridad de dominio, referencial, de valores no nulos y demás, necesarias para la construcción correcta de las bases de datos. Como complemento, deben utilizarse formatos y mecanismos de entrega de los datos y metadatos que garanticen su adecuada recepción y uso. Esto puede implicar mayores costos, pero se justifica plenamente por la mayor eficiencia del proceso y por la calidad de los análisis que se pueden realizar.

2. El mejoramiento de los esquemas de acceso a los metadatos e información complementaria por parte de los investigadores. Para contribuir a fortalecer la capacidad de análisis de los resultados de las encuestas, para que estas cumplan a cabalidad su propósito, se requiere mejorar el acceso de los investigadores no sólo a los datos, sino también a los metadatos, y a otra documentación relevante para su análisis.
3. El acceso a microdata y la posibilidad de cruzar los resultados de las distintas encuestas (que merece especial consideración).³ El acceso a microdata es una práctica posible que garantiza preservar la confidencialidad de la información —que no puede ser revelada a terceros por compromisos legales y éticos—, pues sin la garantía de confidencialidad, podría disminuir la confianza y la cooperación de quienes responden las encuestas, lo cual impactaría en la calidad de las estadísticas. Técnicamente, el acceso a microdata es posible a través de varios esquemas, de los cuales los más usados son la publicación de microdata anónima; el acceso a microdata semianónima, mediante un acuerdo de licencia interinstitucional; la provisión de servicios de ejecución de análisis remotos, y la provisión de servicios de “enclave” o en las instalaciones del proveedor de los datos y bajo su supervisión directa. De estas alternativas,

³ Al respecto hay que aclarar que la primera y segunda encuestas entregan microdata, pero no se pueden cruzar; la EAM no entrega microdata ni tampoco se puede cruzar.

evaluamos como la más conveniente para el equipo investigador el acceso a las bases de datos bajo licencia, pues de esta forma se posibilitaría no solamente el acceso a microdata, sino también la posibilidad técnica de cruzar la información de varias encuestas.

4. El acceso a microdata y la posibilidad de cruzar los datos debe también extenderse a otras encuestas que presenten datos de interés para el estudio de la innovación, por ejemplo, la Encuesta Anual de Servicio y la Encuesta de Innovación de Servicios. Esto posibilitaría un análisis con mayor cobertura sectorial y más rico en variables y, por lo tanto, más útil para el descubrimiento de conocimiento y la generación de diagnósticos y recomendaciones de política y gestión.
5. El mejoramiento del contenido y la regularización de la frecuencia de las encuestas de innovación en Colombia. Sobre la frecuencia, se observa un acortamiento del período transcurrido entre la aplicación de la segunda y tercera⁴ encuestas de innovación, que sería deseable se mantuviera en el tiempo. Idealmente, se deberían seguir las recomendaciones de la OCDE a este respecto.
6. Sobre el contenido, tres recomendaciones específicas surgen a la luz de la experiencia adquirida en el Proyecto: (a) la inclusión en la encuesta de variables que permitan una mejor medición de las capacidades de innovación, tanto en el sentido de la completitud (que todos los ámbitos de las capacidades estén cubiertos) como de la consistencia teórica (que las variables incluidas midan dimensiones relevantes de las capacidades). (b) La inclusión y/o acceso a variables de desempeño empresarial relacionadas con las ventas, las utilidades (pérdidas), las exportaciones y el mercado. (c) Una clara definición de las innovaciones logradas por la empresa durante la ventana de observación de la Encuesta, para que se pueda hacer una mejor caracterización de ellas.

⁴ Realizada en el 2007 por el Departamento Administrativo Nacional de Estadística (DANE), el Departamento Nacional de Planeación (DNP) y Colciencias, cuyos resultados aún no se han divulgado.

7. Finalmente, el mejoramiento de los canales de comunicación entre proveedor y usuario. Establecer canales directos y ágiles de comunicación con el proveedor de los datos puede facilitar el que los investigadores reciban orientación clave para conocer los datos y metadatos, con los potenciales beneficios para la investigación que ello entraña.

TÉCNICAS DE ANÁLISIS: UNA EXPLORACIÓN INCIPIENTE Y UN FUTURO PROMISORIO

Desde su concepción, el Proyecto DESCUBRIMIENTO tuvo un propósito exploratorio de técnicas de análisis novedosas o poco aplicadas al estudio de los estados y dinámicas de la innovación. En este sentido, constituye un intento por descubrir las limitaciones y potencialidades de técnicas fecundas en posibilidades en otros campos de análisis, para ampliar el abanico de herramientas que puedan contribuir a fortalecer la incipiente capacidad de análisis de la innovación en el país. De paso, es necesario reconocer (también hay que decirlo) que los grupos que se unieron en torno al Proyecto buscan acrecentar o adquirir conocimiento y experiencia directa en técnicas no muy familiares en el país. En esta dirección, el Proyecto tuvo también el carácter de un proceso de aprendizaje metodológico, al que contribuyeron la interdisciplinariedad y múltiples competencias de los integrantes del equipo. De hecho, algunos capítulos (10, 11 y 12) hicieron mayor hincapié en la exploración de técnicas que en la búsqueda de resultados temáticos.

Este acápite del capítulo presenta una corta reflexión acerca de los resultados de este ejercicio exploratorio sobre las técnicas. El propósito no es hacer descripciones exhaustivas de carácter procedimental, para lo cual la literatura es prolífica, sino plantear un balance de la experiencia del equipo del Proyecto, relativa a la aplicación de las técnicas. Para ello, nos referiremos a las principales técnicas utilizadas y que ya han sido descritas con el detalle necesario en los capítulos precedentes.

TÉCNICAS DE ANÁLISIS DE ASOCIACIÓN ENTRE VARIABLES ORDINALES

Uno de los desafíos técnicos de las encuestas colombianas de innovación es la mezcla de tipos de variables, con predominio de las variables categóricas y ordinales. En general, esto es natural en encuestas de este tipo, pero introduce dificultades cuando se trata de construir variables latentes a partir de diversas variables observables que incluyen ambas clases de estas. Tal fue la situación que enfrentaron, por ejemplo, Robledo, Gómez y Restrepo (Capítulo 4) al intentar medir y analizar la asociación entre variables latentes del tipo “capacidad” y “desempeño” a partir de conjuntos de variables cuantitativas y ordinales provistas por las encuestas. La solución explorada fue calcular el coeficiente gamma (γ) de Goodman-Kruskal, puesto que su campo óptimo de aplicación es la relación entre dos variables ordinales de dos o más niveles.

Debido a la mezcla de tipos de variables observables a partir de las cuales se construyeron las variables latentes, fue preciso llevar todas las variables observables a una escala ordinal dicotómica. Esto obligó, en el caso de las variables cuantitativas, a determinar los valores que definen el límite entre valores ordinales, con base en el criterio de expertos. Este procedimiento, si bien mostró ser relativamente simple, no necesariamente es el que mejores resultados arroja, pues introduce la subjetividad del experto y causa pérdidas de información valiosa al reducir las variables cuantitativas a variables ordinales más simples. Como alternativa a este procedimiento de construcción de variables latentes, se propone explorar métodos basados en la lógica difusa, que tienen la ventaja adicional de estar diseñados para lidiar con las características borrosas de las variables cuando estas se cualifican mediante procedimientos evaluativos.

COMPONENTES PRINCIPALES Y ANÁLISIS DE CLUSTER

La caracterización de grupos de interés analítico es un objetivo investigativo frecuente. En el Proyecto DESCUBRIMIENTO fue abordado mediante una combinación de análisis factorial y análisis de *cluster* por Arbeláez,

Gómez y Tamayo (Capítulo 5), para el caso del relacionamiento empresa-instituciones de educación superior (IES), y por Gómez y Robledo (Capítulo 7), para la identificación de perfiles y características de empresas en relación con la internacionalización de la financiación de las actividades de desarrollo tecnológico (ADT). La aplicación de esta combinación de técnicas ha sido reportada en la literatura para el análisis de algunos fenómenos de gestión de la tecnología y la innovación, pero no abunda su aplicación al tratamiento de los datos de las encuestas de innovación y, en nuestro medio, es posiblemente nula.

En las aplicaciones que hicieron Arbeláez, Gómez y Tamayo, por una parte, y Gómez y Robledo, por la otra, las técnicas probaron ser de alto potencial analítico, aun considerando las limitaciones de información que representa una encuesta pública frente a la completitud y especificidad de la información obtenida mediante el diseño y aplicación de instrumentos a la medida, que es la situación encontrada en la literatura.

Estas técnicas son útiles para implementar estrategias no supervisadas o de minería de datos. En este caso, el principal desafío encontrado es la escogencia de las variables que participarán en el análisis factorial y en la caracterización de los *clusters*. Para ello, en el Proyecto se utilizaron dos aproximaciones: (i) la de Arbeláez, Gómez y Tamayo, que se basa en la identificación de las variables relevantes reportadas en la literatura y la elaboración de modelos conceptuales básicos a partir de los cuales se seleccionan las variables pertinentes disponibles en las encuestas, y (ii) la aproximación de Manrique y Robledo (Capítulo 12), que emplea una variante en la escogencia de las variables para el análisis factorial, que busca aprovechar al máximo la información disponible en la encuesta, por medio de criterios de expertos. Esta aproximación, sin embargo, requiere una evaluación adicional para aclarar sus posibles ventajas frente a aproximaciones más tradicionales.

ÁRBOLES DE CLASIFICACIÓN

Otra de las necesidades analíticas que frecuentemente emerge en la investigación estadística es la obtención de reglas de asociación entre variables

para descubrir patrones estructurales de los datos. Para estas aplicaciones se recurre a varias técnicas de minería de datos, entre las cuales están los árboles de clasificación (CART). En el Proyecto DESCUBRIMIENTO fueron utilizados los árboles de clasificación por parte de Malaver, Vargas y Ardila (Capítulo 6), para estudiar la asociación entre capacidades de absorción y desempeño tecnológico empresarial, y por parte de Serna, Correa y Robledo (Capítulo 10), para caracterizar las empresas innovadoras.

Las anteriores aplicaciones demostraron sus potencialidades y limitaciones. Por una parte, los CART, en la aplicación de Serna, Correa y Robledo, resultaron valiosos para los análisis de caracterización de las empresas innovadoras frente a las no innovadoras mediante reglas de asociación, considerando distintas tipologías y medidas de innovación. Adicionalmente, los CART demostraron su versatilidad, al poderse aplicar en casos de análisis donde la regresión logística no fue viable por problemas de convergencia por separación completa o cuasi completa.

Por otra parte, puede suceder que las reglas de asociación resultantes de la aplicación de los árboles de clasificación no ofrezcan una interpretación útil o coherente, como fue en parte el problema encontrado por Malaver, Vargas y Ardila. Esto, sin embargo, no necesariamente es un problema de la técnica misma, sino que también puede surgir de las limitaciones de la información con que se cuenta para el análisis e, incluso, de la insuficiencia de modelos teóricos para interpretar los resultados.

MODELOS ECONÓMICOS LOGIT Y PROBIT

En general, los modelos econométricos son técnicas de análisis que permiten explicar el comportamiento de una o varias variables dependientes o endógenas, en función de una o más variables independientes o exógenas que se configuran como *variables explicativas*. Las ventajas de los modelos econométricos están asociadas a su amplia utilización y las múltiples experiencias similares que permiten al investigador transitar por terrenos metodológicamente firmes. Sus principales desventajas están relacionadas con las simplificaciones introducidas a los modelos que les sirven de fun-

damento, por lo que pueden ser muy limitados para abordar fenómenos de alta complejidad.

No obstante, los modelos econométricos constituyen una alternativa técnica valiosa. En el Proyecto DESCUBRIMIENTO, los modelos *logit* y *probit* fueron utilizados por Sierra, Malaver y Vargas (Capítulo 8) para analizar la financiación de la innovación a partir de la EIByC. Dada la naturaleza discreta de la variable dependiente (carácter innovador o no innovador de la empresa), estos modelos permiten establecer valores confiables de la probabilidad de que una u otra variable utilizada como explicativa influya en el resultado binario propuesto, y con ello instituir relaciones entre un factor y un resultado con fuerza de significatividad estadística en los intervalos de confianza asumidos como válidos para hacer afirmaciones.

La aplicación hecha por Sierra, Malaver y Vargas de los modelos *logit* y *probit* fue exitosa, considerando las variables explicativas disponibles y los modelos de análisis utilizados. Sin embargo, en gracia a la complejidad del fenómeno, es importante avanzar en la contrastación o la complementación de estos resultados con los obtenidos al aplicar otras metodologías y consideraciones teóricas, que amplían las posibilidades de análisis, como las presentadas por Otálora, Hurtado y Quimbay en el Capítulo 9, en relación con el mismo tema (financiación de la innovación), y por Serna, Correa y Robledo en el Capítulo 10 sobre las características de las empresas innovadoras.

TÉCNICAS DE ANÁLISIS DE REDES COMPLEJAS

Las redes complejas se caracterizan por poseer rasgos no triviales, del tipo encontrado usualmente en el estudio de las redes sociales y las redes de computadores. En ellas, los patrones de conexión entre sus nodos no son totalmente regulares o aleatorios. En este tipo de redes se estudian características asociadas a la distribución de grado, al coeficiente de aglomeración de nodos, a la preferencia o no de ciertos nodos de establecer conexiones con otros, a la estructura comunitaria y a la estructura jerárquica de la red. Gracias a su habilidad para modelar fenómenos en los que múltiples agen-

tes interactúan de forma compleja, las redes complejas son vistas como una potente alternativa para allegar conocimientos nuevos y fundamentales sobre la estructura y dinámica de las redes económicas.

Considerando la innovación desde la perspectiva de los sistemas complejos, es natural llegar a las redes complejas como una posibilidad de análisis de las relaciones establecidas entre sus agentes. En el Capítulo 9, Otálora, Hurtado y Quimbay analizan la relación entre empresas e instituciones financieras para financiar las actividades de investigación y desarrollo (I+D) e innovación. El análisis se centra en dos parámetros cuantificables para establecer el interés de los actores de un grupo por interactuar con actores de otro grupo (parámetro “interés”) según la visibilidad de los últimos (parámetro “visibilidad”). Sobre estos dos parámetros del modelo, de fácil interpretación, se analizan las tendencias de la red.

El estudio de la financiación de la innovación como un caso de red compleja es una novedosa forma de abordar el problema. Los resultados de Otálora, Hurtado y Quimbay muestran las potencialidades de esta perspectiva aplicada al análisis de las relaciones entre diversos agentes de la innovación. En el caso particular de empresas y entidades financieras, se trata de una red bipartita en la que los nodos del mismo tipo no se vinculan entre sí, conformada por 6.212 empresas industriales manufactureras y 39 fuentes de financiación, con datos de la EDIT II.

Los resultados se resumen en las características básicas de conectividad de la relación de conocimiento entre empresas y fuentes de financiamiento, a partir de las propiedades de adyacencia que determinan la distribución de grado para los nodos; además, se encuentran los parámetros “interés” y “visibilidad” y se interpretan sus valores y distribución según los datos de la EDIT II.

TÉCNICAS DE ANÁLISIS BASADAS EN SISTEMAS NEURODIFUSOS

En el Proyecto DESCUBRIMIENTO, Duque, Manrique y Robledo (Capítulo 11) acudieron a los sistemas neurodifusos para explorar su potencial de aplicación al estudio de la innovación a partir de las bases de datos,

aprovechando la posibilidad que brindan en la extracción de reglas fácilmente interpretables con grandes cantidades de datos, sobre las cuales se aplican algoritmos no supervisados de regresión no lineal. El trabajo exploró la aplicabilidad y limitaciones de tres técnicas de análisis. La primera emplea funciones regresoras del tipo *perceptrón multicapa* (MLP, por su sigla en inglés), de las que extrae reglas difusas simples empleando el procedimiento *Rule Extraction from Fuzzy Neural Network* (REFuNN); la segunda emplea *funciones de base radial* como funciones regresoras y el procedimiento RULEX para extraer reglas difusas; finalmente, la tercera es similar a la segunda, pero emplea *máquinas de soporte vectorial* como funciones regresoras.

El fenómeno escogido como caso de aplicación de esas técnicas fue la relación entre uso energético e innovación a partir de variables seleccionadas de la EAM 2005. La aplicación de las técnicas se hizo a partir de un proceso integrado de cinco etapas, compuesto por la selección de las variables, su preprocesamiento, la aplicación de los algoritmos de regresión no lineal, la extracción de reglas y su calificación mediante las medidas de soporte y confianza. Este proceso mostró ser un esquema modular y apropiado que permitió la aplicación de las técnicas de manera sistemática y confiable. No obstante, se identificó una posibilidad de aumentar la eficiencia del proceso, utilizando algoritmos genéticos que podrían reducir sus etapas bajo ciertas circunstancias, lo cual se plantea como uno de los frentes de trabajo futuro que continuarán la exploración de las técnicas.

En general, las tres técnicas arrojaron resultados útiles, aunque la primera lleva a reglas difusas más simples y fáciles de interpretar, mientras que las dos últimas permiten la obtención de reglas complejas y flexibles, las cuales podrían expresar relaciones no posibles de construir mediante las reglas más simples. Por otra parte, si bien las técnicas demostraron su alto potencial para los análisis de los fenómenos complejos de la innovación, también se identificaron posibilidades adicionales de trabajo futuro para mejorar su desempeño. En particular, los algoritmos de optimización por enjambre de partículas se vislumbran como opciones prometedoras para garantizar convergencias más cercanas al óptimo global en los procedimientos de minimización de la medida de error de la regresión.

Finalmente, las características asociadas a las funciones de Kernel, empleadas dentro de las funciones de base radial y las máquinas de soporte vectorial, dan pie a propuestas de investigación futura, relacionadas con la sustitución de este tipo de funciones por otras con menor cantidad de simetrías, lo que podría llevar a reglas más flexibles y con mejores valores de soporte y confianza.

LA INNOVACIÓN EN COLOMBIA: AMPLIAS POSIBILIDADES POCO EXPLOTADAS

La innovación es un fenómeno complejo, como es usual que sean los fenómenos sociales sistémicos. Los estudiosos del tema, luego de proponer sucesivamente modelos lineales, articulados, integrados y en red, coinciden hoy en la naturaleza sistémica de la innovación, como fenómeno social dinámicamente complejo y de características no lineales, cuyos resultados funcionales son inciertos, y en el que participan múltiples agentes con respuestas y comportamientos no conocidos a priori. Su éxito es multicausal y específico. No tiene causas simples ni generalizables. Por ello es difícil establecer relaciones causa-efecto, y las extrapolaciones en el tiempo y en el espacio son altamente problemáticas.

La innovación es también un fenómeno histórico y geográfico: se sostiene sobre una base de capacidades específicas acumuladas y renovadas a través del tiempo en firmas, sectores, *clusters*, regiones y países. Decisiones pasadas afectan sensiblemente las posibilidades presentes y futuras. Las dimensiones espaciales imponen limitaciones o abren oportunidades de desarrollo innovador.

Para enfrentar con alguna mínima garantía de éxito el reto de generar aportes de conocimiento sobre esta realidad compleja, en el Proyecto DESCUBRIMIENTO se optó por una aproximación diversa tanto en los ámbitos temático, conceptual y teórico, como en los instrumentos técnicos de análisis. Esto condujo a una dinámica de trabajo en red, más que a la conformación de un único equipo con una única agenda investigativa. Dos hilos conductores fundamentales mantuvieron la cohesión de la red:

el interés en los estados y las dinámicas de la innovación en Colombia y el trabajo sobre los datos de las encuestas, explorando las múltiples posibilidades técnicas de la estadística y la inteligencia artificial integradas en el proceso de KDD.

El ejercicio exploratorio del Proyecto DESCUBRIMIENTO, utilizando los datos de las encuestas de innovación como insumo de procesos KDD, revela una doble realidad en cuanto a la producción de conocimiento sobre la innovación en Colombia. Por un lado, se vislumbra la alta potencialidad de esta aproximación para explorar relaciones, clasificaciones y reglas de asociación entre variables; por el otro, se constata que medir, evaluar, relacionar y caracterizar *constructos* tan complejos como “capacidad” (tecnológica, de innovación, de absorción, etc.) y “desempeño” (tecnológico, innovador, competitivo, de negocio, etc.), a partir de los datos de las encuestas de innovación, es un asunto sumamente problemático en el que todavía abundan conceptos similares con diferencias no claras, confusiones terminológicas, múltiples aproximaciones teóricas y una proliferación de paradigmas.

Un contexto de datos circunscrito a los resultados de una única encuesta es insuficiente si se considera la complejidad de la innovación como fenómeno sistémico y evolutivo. Previendo esta situación, el Proyecto DESCUBRIMIENTO se propuso enriquecer las posibilidades de análisis articulando datos provenientes de diferentes fuentes (encuestas y bases de datos) y recurriendo, por una parte, al conocimiento y experiencia desarrollados en la práctica empresarial de la inteligencia de negocios y, por la otra, al desarrollo de los procesos de KDD aplicados a las bodegas de datos.

Sin embargo, las posibilidades de enriquecer los análisis de los resultados de las encuestas de innovación con datos provenientes de la EAM y la base de datos ScienTI fueron muy limitadas, por las dificultades de acceso a los microdata y el cruce entre encuestas. Por lo tanto, aunque no arranca de cero, ante la magnitud del desafío investigativo, el trabajo apenas comienza. Aun así, es posible señalar algunas direcciones en que la exploración realizada en el Proyecto produjo aportes de interés temático: (a) en la relación entre capacidades y desempeño, (b) en el financiamiento de la innovación y el desarrollo tecnológico, (c) en el relacionamiento entre empresas e IES, y (d) en la caracterización de las empresas innovadoras.

RELACIÓN ENTRE CAPACIDADES Y DESEMPEÑO

El estudio de las capacidades organizacionales a partir de las encuestas de innovación era claramente, desde el comienzo, un asunto de alto interés, no solamente por la atención que tanto en ámbitos investigativos como de la práctica política y gerencial han recibido tales capacidades, sino porque este concepto está explícitamente identificado como fundamental en documentos conceptuales que sustentan el diseño de las encuestas. En el Proyecto se realizaron dos exploraciones a partir de encuestas distintas y utilizando métodos de análisis distintos: la exploración de la relación entre capacidades tecnológicas de innovación y desempeño empresarial, presentada por Robledo, Gómez y Restrepo en el Capítulo 4, y la exploración de la relación entre capacidades de absorción y desempeño tecnológico presentada por Malaver, Vargas y Ardila, en el Capítulo 6.

En cuanto al primer trabajo, el interés del equipo investigativo era, en un principio, explorar la asociación estadística entre las capacidades de innovación y el desempeño empresarial, entendido este desde la doble perspectiva de la innovación (desempeño innovador) y del negocio, y su evolución en el período comprendido entre la primera y la segunda encuestas de innovación. Sin embargo, las dificultades asociadas a los cambios de diseño entre ambas encuestas, junto con la imposibilidad de cruzarlas (a nivel de microdata) y con la EAM, para obtener información sobre variables relativas al desempeño del negocio (ventas, exportaciones, utilidades, etc.) en una ventana temporal convenientemente amplia, impidieron explorar la relación entre el desempeño innovador y el desempeño del negocio.

A pesar de esos inconvenientes, los resultados del trabajo de Robledo, Gómez y Restrepo muestran que en las empresas los resultados del análisis estadístico de las encuestas colombianas de innovación aportan evidencia empírica significativa sobre la validez del modelo teórico que explica el desempeño innovador empresarial a partir de la base de capacidades de innovación tecnológica acumuladas por la empresa. Al respecto, es relevante el hecho de que estos resultados sean consistentes en los análisis sobre las dos encuestas colombianas de innovación, a pesar de sus diferencias de diseño y casi una década de separación en el tiempo.

Esto es una clara indicación de que las capacidades de innovación tecnológica son componentes estructurales de la organización productiva cuando se trata de explicar su comportamiento innovador. Como parte de un ejercicio exploratorio, estos resultados deberán consolidarse en aspectos conceptuales, procedimentales y analíticos, pero los indicios arrojados por la exploración realizada son alentadores y sugieren que el trabajo apunta en una dirección prometedora que deberá ampliarse y profundizarse en el futuro.

Uno de los aspectos en que el trabajo revela necesidad de mayor exploración es el análisis de los estados y dinámicas sectoriales de la innovación. Al respecto, el Proyecto DESCUBRIMIENTO acudió a la CIIU, intentando encontrar patrones sectoriales significativos y de interés. Sin embargo, los resultados no fueron muy alentadores. Esto se debe, posiblemente, a que esta sectorización no es adecuada para los fenómenos bajo análisis, y es necesario acudir a propuestas más afines con las dinámicas tecnológicas, como la taxonomía propuesta por Pavitt (1984) y ajustada por varios autores posteriormente.

Por las dificultades mencionadas, fue imposible explorar la relación entre capacidades tecnológicas de innovación y desempeño del negocio. Como la literatura especializada lo señala, esta es una relación compleja, establecida en el contexto de sistemas dinámicos en los que operan importantes retardos y que exhiben un comportamiento altamente contingente, y aquello que en ciertas épocas o bajo ciertas condiciones es válido, no lo es bajo otras. Sin embargo, esta dificultad no debe desalentar el análisis; por el contrario, debe ser un aliciente. La cuestión es que, en último término, la tecnología y la innovación y el desarrollo tecnológico no son fines en sí mismos, sino medios para obtener desempeños empresariales superiores en el mercado. Y si bien es justo reconocer que algunas encuestas realizadas en el país abordan explícitamente esta relación (Malaver y Vargas, 2006; Calderón, Álvarez y Naranjo, 2009), mientras los problemas de acceso a microdata y cruce entre encuestas y bases de datos que se señalaron arriba no se superen, las encuestas de innovación del orden nacional seguirán siendo instrumentos muy limitados para la consecución de los fines que se proponen.

En cuanto al trabajo de Malaver, Vargas y Ardila, el interés se centra en la asociación entre los recursos (humanos), las prácticas en que se despliega el núcleo de las capacidades para adquirir las tecnologías exógenas a la empresa y el desempeño tecnológico en que se expresa la capacidad de absorción tecnológica, en su componente de adquisición de tecnología. Los datos provienen de la EIByC. Los resultados obtenidos no muestran una relación clara entre el nivel de formación de los recursos humanos y el desempeño tecnológico vía adquisición de tecnología. Esto indica, como lo señalan los autores, que lo que en otros contextos es una buena premisa de partida, en la industria de Bogotá y Cundinamarca, “es un fenómeno que debe estudiarse con mayor profundidad, si se quieren formular políticas realistas”.

Por el contrario, la búsqueda de reglas de asociación mediante arboles de decisión mostró, aunque no de manera concluyente, que las capacidades de absorción tecnológica acumuladas y puestas en acción en las prácticas para adquirir las tecnologías exógenas a la empresa inciden positiva y significativamente en el grado de cambio tecnológico incorporado.

Los resultados sugieren, por último, que la relación entre *inputs* y *outputs* no es directa y que en los procesos de incorporación de las tecnologías median capacidades acumuladas y aprendizajes que influyen de manera significativa en los desempeños tecnológicos alcanzados. Esta conclusión confirma, una vez más, la complejidad del fenómeno estudiado y las limitaciones de los datos disponibles, e indica la necesidad de efectuar estudios y utilizar herramientas metodológicas que permitan ahondar en su caracterización. Debido a lo exógeno del grueso del cambio técnico, esta profundización es vital para el desarrollo tecnológico de países y regiones en vía de desarrollo.

FINANCIAMIENTO DE LA INNOVACIÓN Y DESARROLLO TECNOLÓGICO

En esta dirección de la exploración, un primer interés del Proyecto se enfoca en examinar la importancia de las fuentes internacionales en el financiamiento de las ADT, según se desprende de los resultados de la EDIT II.

Al respecto, en el capítulo 7, Gómez y Robledo analizan dos aspectos del problema: (i) la visibilidad de las fuentes financieras⁵ (banca de inversión y de comercio internacional, casa matriz, organismos internacionales de cooperación técnica internacional, programas de cooperación técnica internacional y Programa CARANA) y (ii) la relación entre la inversión extranjera directa a través de empresas catalogadas como extranjeras (capital social mayoritariamente extranjero) y la inversión en actividades de I+D y capacitación tecnológica (CT). Para el efecto, se utilizaron descripciones estadísticas y técnicas de análisis multivariante (particularmente técnicas de análisis de componentes principales y análisis de *cluster*).

Respecto del primer asunto, los resultados revelan un alto desconocimiento de las fuentes internacionales para financiar las ADT, aun cuando es la banca la fuente más visible, lo cual podría explicarse por la natural asociación existente entre la banca y las operaciones financieras. Las fuentes de cooperación internacional (exceptuando CARANA) son prácticamente invisibles. En cualquier caso, cuando se trata de acceder a los recursos, la inspección de los resultados es concluyente: solamente 14 de las 6.212 empresas encuestadas usaron una fuente internacional para financiar sus ADT. De todas maneras, el acceso casi nulo a los recursos financieros internacionales para ADT es un resultado similar al que alcanzan Sierra, Malaver y Vargas en el Capítulo 8, utilizando una fuente de información distinta y un modelo *probit* de análisis econométrico.

Una vez confirmada la insignificante participación de fuentes internacionales para el financiamiento de ADT, emerge la pregunta sobre la asociación entre inversión extranjera directa y las ADT, con énfasis en I+D y CT. Para este análisis se recurrió a las técnicas de análisis multivariante, considerando las variables asociadas a las inversiones en I+D y CT y el origen del capital (nacional o extranjero) de las empresas. La aplicación de las técnicas de análisis revela la existencia de dos *clusters* de establecimientos conformados a partir de dos factores que explican el 76% de la varianza: los

⁵ Es importante llamar la atención sobre la falta de rigor de la clasificación y definición de estas fuentes, según la EDIT II, lo que seguramente incide de manera importante en la exactitud de las respuestas y, por lo tanto, en la validez de los análisis.

establecimientos investigadores y los establecimientos capacitadores. Para ambos *clusters*, la financiación de las actividades de I+D y de CT proviene principalmente de fuentes nacionales.

Los establecimientos investigadores se caracterizan por su esfuerzo en I+D y el tamaño grande de sus empresas. Este *cluster*, sin embargo, está integrado por un número muy pequeño de empresas (0,32% del total), mientras el *cluster* de establecimientos capacitadores está integrado casi en su totalidad por pymes que invierten muy poco en I+D y tímidamente en CT. A pesar de la disparidad del tamaño de los *clusters*, se observa con claridad una presencia altamente significativa de empresas extranjeras (cuyo capital es mayoritariamente extranjero) en el *cluster* de los establecimientos investigadores. Esto sugiere una asociación interesante entre inversión extranjera directa e inversión en actividades de I+D y CT en las grandes empresas, que es necesario continuar explorando.

En el Capítulo 8, Sierra, Malaver y Vargas indagan por la relación entre la financiación y la innovación en la industria de Bogotá y Cundinamarca y se concentran en las fuentes y los mecanismos de financiación a los que acuden las empresas cuando innovan, en la existencia de restricciones financieras y en la efectividad de los instrumentos de política pública de apoyo a la financiación de la innovación. Para el análisis micro de esta relación se utilizan los datos de la EIByC y se acude a la estadística descriptiva, complementada con análisis econométrico (modelos *probit* y *logit*) y de correspondencias. Para indagar por la existencia de restricciones financieras y del *pecking order* (jerarquía de preferencias financieras, a favor de los fondos propios, fondos de deuda bancaria y, por último, capital nuevo), se evalúan las razones por las cuales las empresas no innovaron en el período de observación, los obstáculos para innovar, las fuentes financieras de la innovación, los factores que obstruyeron el acceso a la financiación externa cuando esta no se dio y los problemas que las empresas encontraron cuando sí existió, las cuales se comparan entre los tipos de empresa, definidos en la EIByC y el tamaño de las empresas.

De acuerdo con la EIByC, en la industria de Bogotá y Cundinamarca existen restricciones financieras, como quiera que después de las barreras provenientes del mercado, la carencia de recursos financieros es la segun-

da razón por la cual las empresas no innovan. También constriñe a las empresas que innovan. Incluso, es el factor que más inhibe esta actividad. Más relevante aún es el hecho de que “la restricción financiera crece a medida que aumenta la capacidad de innovación de las empresas”, y dicha restricción es, además, estadísticamente significativa. Debe advertirse, sin embargo, que entre las empresas más pequeñas, que a la vez son las menos o no innovadoras, el análisis de correspondencias sugiere el predominio de razones de índole cultural (no es rentable, es innecesario) entre las micro, pequeñas y medianas empresas, mientras que las grandes empresas aducen razones fuertemente asociadas al tamaño inadecuado del mercado.

El análisis de las fuentes de financiación de la innovación confirma la existencia del *pecking order*; esto es, el marcado predominio de los recursos internos, incluso en el caso de las empresas grandes más innovadoras. Los resultados evidencian la insignificancia de las fuentes internacionales, así como el mínimo uso de los instrumentos de fomento financiero a la innovación, sobre todo por parte de las empresas de menor tamaño, en el papel, las más necesitadas, lo cual confirma que este tema sigue siendo un gran desafío para la política pública. De otra manera será muy difícil superar las fallas del mercado, que subyacen en la restricción financiera.

Al ahondar por la forma como opera la restricción financiera, y que se expresa en las principales razones para no utilizar financiación externa, el análisis señala la suficiencia de los recursos internos, así como el desconocimiento de las fuentes y los trámites exigidos por las entidades financieras. El desconocimiento de las fuentes aparece, de nuevo, como una restricción de acceso a la financiación externa, lo que también remite al problema de visibilidad de las fuentes financieras que se señaló en el caso de la financiación internacional.

Ahora bien, entre las empresas que accedieron a financiación externa, más del 75% señala no haber tenido problemas, parece más de acceso al financiamiento que de la relación misma. Las que reportan problemas, estos están asociados al bajo desarrollo del sector financiero tradicional para evaluar y financiar proyectos de innovación.

En síntesis, el trabajo muestra la existencia de restricciones financieras que inhiben el surgimiento y dificultan el desarrollo de la innovación; reve-

la, además, la insuficiencia de las políticas e instrumentos financieros para remediar las fallas del mercado y proveer acceso adecuado a las empresas, sobre todo a las de menor tamaño, que son también las que menores dinámicas innovadoras presentan. Quedan interrogantes relativos a las razones por las cuales las empresas tienden a financiar la innovación con recursos propios y la relación que existe entre esta tendencia y las restricciones de acceso a financiación externa.

El análisis de la relación entre fuentes de financiamiento y empresas en la perspectiva de la innovación también es abordado por Otálora, Hurtado y Quimbay en el Capítulo 9. En este caso, el análisis se plantea en el marco de los estudios de redes complejas y considera dos parámetros para capturar los rasgos generales del sistema empresas-fuentes de financiamiento: el *interés* de las empresas por conocer las fuentes y la *visibilidad* de las fuentes de financiamiento ante las empresas. La aproximación utilizada permite desacoplar ambos parámetros, de forma que el interés de las empresas por conocer las fuentes de financiamiento se independiza de la ocurrencia efectiva de la financiación; así mismo, la visibilidad de las fuentes incorpora su dinámica relacional de manera independiente al interés de las empresas.

El sistema empresas-fuentes de financiamiento está compuesto por 6.212 empresas y 39 fuentes, con datos provenientes de la EDIT II. Los resultados muestran que los empresarios tienen poco interés por conocer las fuentes de financiamiento, a pesar de la alta densidad de vínculos en la red. Esto se puede interpretar como carencia de motivación de las empresas por construir relaciones con las fuentes de financiamiento. La conclusión revela una nueva perspectiva de análisis de la relación innovación-financiamiento, que lleva la discusión al terreno de las estrategias empresariales. Desde esta perspectiva, se podría sugerir que estrategias empresariales que valoren la innovación conducirían a un mayor conocimiento y utilización de las fuentes externas de financiamiento, mientras que estrategias que desconozcan la innovación llevarían a pasividad, desconocimiento y distanciamiento de las fuentes de financiamiento.

Esto explicaría situaciones aparentemente contradictorias que, a partir de las conclusiones de los capítulos 7, 8 y 9, podrían sugerirse como un fenómeno generalizado en el país: el mínimo acceso de las empresas a las

fuentes de financiamiento respecto del conocimiento que tienen de estas, el hecho de que las restricciones financieras aumenten en importancia a medida que las empresas son más innovadoras y el alto porcentaje de empresas que manifiestan no haber tenido problemas con el financiamiento una vez han accedido a este. En otras palabras, la explicación al fenómeno de un alto conocimiento de ciertas fuentes, un muy bajo acceso a ellas, una mayor importancia de las restricciones financieras para las empresas más innovadoras y una alta satisfacción con el financiamiento una vez obtenido, también habría que buscarla en el ámbito de las estrategias empresariales.

RELACIONAMIENTO ENTRE EMPRESAS E INSTITUCIONES DE EDUCACIÓN SUPERIOR

Las empresas y las IES son dos agentes clave de los sistemas de innovación; el grado e intensidad de sus relaciones no sólo es signo de dinamismo del sistema, sino que revela un importante acoplamiento entre estrategias organizacionales fundamentales para fortalecer procesos de construcción de consensos sobre la identificación de áreas de importancia estratégica y temas de prioridad investigativa; acelerar los ritmos de la innovación; distribuir los altos costos y los beneficios de la I+D; propiciar la confluencia de capacidades especializadas y complementarias para controlar los riesgos de fracaso técnico de los proyectos, y generar oportunidades de gestión exitosa de recursos públicos de cofinanciación.

En el Proyecto DESCUBRIMIENTO, este tema fue abordado por Arbeláez, Gómez y Tamayo, y sus resultados se presentan detalladamente en el Capítulo 5. El trabajo, de carácter exploratorio como los demás del Proyecto, estudia las características de este relacionamiento y su asociación con el desempeño innovador de las empresas, a partir de los datos de la EDIT II. Metodológicamente, recurre a técnicas estadísticas y de minería de datos aplicadas a tres grupos de variables: cinco variables de relacionamiento, que miden directamente distintos aspectos de la relación empresa-IES; la variable de desempeño, representada por el número total de innovaciones de producto y proceso que la empresa logró obtener en los años 2003

y 2004, y ocho variables caracterizadoras, identificadas con apoyo en la literatura especializada por su conexión con el fenómeno bajo análisis.

Sobre las cinco variables de relacionamiento se aplicó la técnica de componentes principales y con ello se lograron reducir las variables a dos factores que explican casi el 60% de la varianza, denominados factor capacitación y factor de ADT, asociados al énfasis de la relación entre las empresas y las IES. Un análisis de *cluster* realizado sobre las puntuaciones de los dos factores logró identificar tres *clusters* de establecimientos empresariales que, por sus características predominantes, se denominan *aislados* (con muy poco relacionamiento), *capacitados* (relacionamiento basado en actividades de capacitación) y *relacionados* (relacionamiento basado en ADT, incluida también la capacitación).

Los resultados confirman la existencia de una importante desconexión de las relaciones empresa-IES en el Sistema Nacional de Innovación colombiano. En otras palabras, esta red de relaciones a la que se atribuye buena parte del dinamismo de un sistema de innovación, está prácticamente inactiva en Colombia. La gran mayoría de los establecimientos pertenecen al *cluster* de los aislados y un porcentaje muy alto son empresas no innovadoras. Por otra parte, el *cluster* que agrupa a las empresas más innovadoras y que muestran los mayores esfuerzos en I+D basan en la capacitación su relacionamiento con las IES. Finalmente, están los establecimientos relacionados, que son muy pocos y con un pobre desempeño innovador.

Se constata, así, el estado fracturado del Sistema Nacional de Innovación colombiano, en términos de las relaciones empresa-IES, hecho que concuerda con los hallazgos de Malaver y Vargas (2006) a partir de la EIByC. Aquí esto se explica en la medida en que la gran mayoría de empresas no establecen relaciones significativas con las IES, las empresas innovadoras no muestran relaciones importantes con la actividad investigativa de las IES y aquellas pocas empresas que exhiben relacionamientos significativos con las IES, a nivel de las ADT, no tienen un buen desempeño innovador.

Claramente, la innovación realizada en las empresas estudiadas no está vinculada a las capacidades investigativas de las IES y los pocos casos en los que esto ocurre son aislados y no alcanzan a configurar una realidad estadísticamente significativa. Estos análisis deben, sin embargo, pro-

fundizarse pues, en principio, son contradictorios, por ejemplo, con los hallazgos del Capítulo 8, en los cuales se encuentra que las empresas más innovadoras, que son grandes, son las que acuden en mayor medida a la financiación de Colciencias.

CARACTERIZACIÓN DE LAS EMPRESAS INNOVADORAS

En las ciencias económicas y administrativas, las clasificaciones de los objetos de estudio constituyen instrumentos poderosos de análisis que reciben gran atención de parte de investigadores, analistas y tomadores de decisiones, a pesar de su naturaleza artificial y, en cierto grado, arbitraria. Las clasificaciones utilizadas en los estudios tienen un efecto directo en los resultados, los cuales, a su vez, pueden incidir en importantes decisiones de política pública y de gestión empresarial. Debido a su importancia, y en aras de limitar el impacto no deseado de su artificialidad y arbitrariedad, las clasificaciones son en sí mismas tema de discusión y acuerdo. Lo que se busca es construir propuestas de clasificación que, partiendo de una base racional, sean adoptadas ampliamente como plataformas de análisis que posibiliten ejercicios de comparación o referenciación.

Esto, por supuesto, ocurre en el campo de los estudios de la innovación, en los cuales una de las propuestas centrales tiene que ver con la clasificación de las empresas innovadoras. Tal clasificación posibilita, por ejemplo, la caracterización de los grupos de empresas, los estudios demográficos de las empresas innovadoras —tanto de corte longitudinal en el tiempo (evolución de la población de empresas innovadoras) como transversal (distribución entre sectores industriales, por tamaño de empresas, etc.)— y la indagación sobre factores explicativos de las dinámicas de crecimiento o decrecimientos de estas poblaciones y sus agrupaciones.

El *Manual de Oslo* (OECD, 2005) tiene propuestas de clasificación de las empresas innovadoras, basadas fundamentalmente en dos dimensiones que conforman una matriz de clasificación: el tipo de innovación (de producto, proceso, *marketing* y organizacional) y el grado de novedad de la innovación (nueva para la empresa, nueva para el mercado y nueva para el

mundo). Según la propuesta del *Manual*, las empresas pueden ser innovadoras o no (si han introducido o no al menos un tipo de innovación en el período de análisis); en particular, las empresas innovadoras tecnológicas son aquellas que han introducido al menos una innovación de producto o de proceso. En su tercera edición (OECD, 2005), el *Manual* introduce la categoría de “empresa activa en innovación”, cuando esta realiza actividades de innovación en el período estudiado.

En Colombia, en los análisis de los resultados de las dos encuestas de innovación (la EDT y la EDIT II) se han propuesto dos clasificaciones diferentes, que no coinciden con las propuestas por el *Manual de Oslo* en sus distintas versiones, pues difieren en las dimensiones de análisis y en las definiciones de las categorías.⁶ Los cambios entre ambas propuestas son tan significativos que los resultados no son comparables.

En el Proyecto DESCUBRIMIENTO se caracterizaron las empresas innovadoras, utilizando los resultados de la EDT (véase el Capítulo 10) y partiendo de la recomendación del *Manual de Oslo* de clasificar las empresas innovadoras por tipo de innovación. Así, se consideraron las empresas como innovadoras tecnológicas (de producto y proceso) y como innovadoras organizacionales. La técnica de clasificación usada fue la de los árboles de clasificación y regresión. Además, se hizo un esfuerzo por identificar un indicador global de innovación construido vía análisis de componentes principales.

Los resultados muestran que en el 61,3% de las empresas innovadoras de producto la fuente usual de la innovación es interna y proviene de los directivos del establecimiento. Esto puede estar asociado al carácter informal y no sistemático de la innovación, como lo señalan Durán *et al.* (1998) y lo corroboran otros estudios (Vargas, Malaver y Zerda, 2003).

⁶ Para el caso de la EDT, Durán *et al.* (1998) clasifican las empresas como no innovadoras, potencialmente innovadoras, innovadoras en sentido amplio y en sentido estricto. Esta tipología fue acogida y ajustada en la EIByC (Malaver y Vargas, 2006), definiendo el grado de novedad de acuerdo con su alcance (la empresa, el mercado nacional o internacional) y su nivel de originalidad (una mejora significativa, nuevo sin patente y nuevo con patente). Para la EDIT II, Colciencias, el DNP y el DANE (2005) propusieron una clasificación distinta.

En pocas empresas, la fuente de la innovación es externa y proviene de los clientes. De esta manera, se configura una dinámica de la innovación inducida principalmente por las iniciativas de los directivos y en respuesta al mercado. Sin embargo, un grupo significativo de empresas innovadoras de producto tiene como fuente interna de la innovación la actividad permanente y organizada de I+D. Esto indica la existencia de un núcleo de empresas que desarrolla una gestión formal y especializada de la innovación con un importante respaldo en la I+D.

Las empresas innovadoras de proceso, por su parte, son más numerosas que las innovadoras de producto, se caracterizan por que su fuente de innovación son los directivos o, en su defecto, los clientes y muestran una clara similitud con las innovadoras de producto. Igual sucede con las empresas innovadoras organizacionales, que revelan unas características similares en las empresas innovadoras en cuanto a la fuente de sus innovaciones, independientemente del tipo de innovación.

Para caracterizar a las empresas innovadoras sin referencia a un tipo de innovación, se exploró la construcción de un indicador global de innovación utilizando la técnica del análisis de componentes principales. Los resultados obtenidos indican una asociación significativa entre desempeño exportador e innovación, pero en ciertos sectores industriales esto no se cumple. De este modo, queda abierta la discusión sobre la construcción de un indicador global de innovación que refleje de manera adecuada un desarrollo más equilibrado de los distintos tipos de innovación, más allá de la definición simple de empresa innovadora como aquella que implementó una innovación durante el período de referencia. Además, se confirma la necesidad de explorar clasificaciones alternativas a la CIU para establecer la relación entre innovación y desempeño empresarial.

A MODO DE COMENTARIO FINAL

Este capítulo inició como una propuesta de conclusiones, recomendaciones y trabajo futuro. Y de cierta manera lo es; sin embargo, peca por defecto y por exceso en el logro de este propósito. Por una parte, porque

ciertamente se queda corto frente a las riquezas de conclusiones específicas de cada uno de los capítulos que lo anteceden, a pesar del esfuerzo realizado para captar en él los principales hallazgos. Por la otra, porque fue inevitable matizar continuamente su contenido con reflexiones que llevan las ideas más allá de los límites fijados por los marcos conceptuales propuestos y los instrumentos de análisis aplicados en cada trabajo específico. No obstante, es necesaria una mirada en perspectiva y que contextualice los resultados de un trabajo exploratorio y multifacético como el presente.

Como en toda empresa del conocimiento, queda mucho camino por recorrer. Esta es quizás la principal conclusión de un trabajo de esta naturaleza, que se asoma curiosamente al descubrimiento: la exploración revela un amplio mundo por conquistar. Recorrido posiblemente por otros, en otras dimensiones del espacio y el tiempo, pero que es necesario emprender por todos los que aceptan el desafío del descubrimiento de conocimiento, en las dimensiones propias de un camino que, como el conocimiento mismo, es nuevo y diferente para quienes lo quieran recorrer.

REFERENCIAS

- Calderón, G.; Álvarez, C. J. y Naranjo, J. C. (2009). Orientación estratégica y recursos competitivos: un estudio en grandes empresas industriales de Colombia. *Cuadernos de Administración*, 22 (38), 49-72.
- Durán, X.; Ibáñez, R.; Salazar, M. y Vargas, M. (1998). *La innovación tecnológica en Colombia: características por tamaño y tipo de empresa*. Bogotá: Departamento Nacional de Planeación.
- Jiménez, C. y Amón, I. (2009, 23-25 de abril). *Hacia una metodología para la selección de técnicas de depuración de los datos*, documento presentado en el IV Congreso Colombiano de Computación. Auspiciado por la Sociedad Colombiana de Computación y organizado por la Universidad Autónoma de Bucaramanga. Bucaramanga, Colombia.
- Malaver, F. y Vargas, M. (2006). *Capacidades tecnológicas, innovación y competitividad de la industria de Bogotá y Cundinamarca: resultados de una encuesta de innovación*. Bogotá: Cámara de Comercio de Bogotá, Observatorio Colombiano de Ciencia y Tecnología, Agenda Regional de Ciencia y Tecnología y Consejo Regional de Competitividad.

- Organization for Economic Co-operation and Development (OECD), (2005). *Proposed guidelines for collecting and interpreting technological innovation data: Oslo Manual*. París: OCDE.
- Pavitt, K. (1984). Sectoral patterns of technological change: Towards a taxonomy and theory. *Research Policy*, 13, 343-373.
- Vargas, M.; Malaver, F. y Zerda, A. (Eds.), (2003). *La innovación tecnológica en la industria colombiana*. Bogotá: OCyT y Colciencias.

Autores

Manuel Santiago Arbeláez Cuartas

Estudiante de la Maestría en Gestión Tecnológica, Universidad Pontificia Bolivariana, Medellín, Colombia. Ingeniero de Telecomunicaciones, Universidad Santo Tomás, sede Medellín, Colombia. Ingeniero de proyectos en la Subdirección de Acceso de la Vicepresidencia de Tecnología de Información y Telecomunicaciones (TIC) de EPM Telecomunicaciones S. A. ESP, Medellín. Correo electrónico: manarbe@une.net.co.

Felipe Ardila Camelo

Estudiante de la Maestría en Hidrosistemas, Pontificia Universidad Javeriana, Bogotá, Colombia. Ingeniero de sistemas, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. Miembro del Grupo de Investigación del Instituto Geofísico, Pontificia Universidad Javeriana. Correo electrónico: felipeardilac@gmail.com.

Jorge Charum Díaz

DEA, Analyse numérique, Université de Bourgogne, Dijon, Francia. Ingeniero electrónico, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. Investigador asociado del Observatorio Colombiano de Ciencia y Tecnología, Bogotá, Colombia. Investigador de la Corporación Colombiana del Saber Científico (Sciocorp), Bogotá, Colombia. Correo electrónico: jcharum@ocyt.org.co.

Juan Carlos Correa Morales

PhD en Estadística, University of Kentucky, Lexington, Estados Unidos. Magíster en Estadística, University of Kentucky. Estadístico, Universidad de Medellín, Medellín, Colombia. Miembro del Grupo de Investigación en Estadística, Universidad Nacional de Colombia, sede Medellín. Profesor asociado de la Universidad Nacional de Colombia, sede Medellín. Correo electrónico: jccorrea@unal.edu.co.

Luis Felipe Duque Álvarez

Ingeniero electrónico, Universidad Pontificia Bolivariana, Medellín, Colombia. Miembro del Grupo de Política y Gestión Tecnológica, Universidad Pontificia Bolivariana. Correo electrónico: luisduque440@hotmail.com.

Fredy Alexander Gómez Jiménez

Magíster en Ingeniería Administrativa, Universidad Nacional de Colombia, sede Medellín, Colombia. Ingeniero industrial, Universidad Nacional de Colombia, sede Medellín. Miembro del Grupo de Investigación en Innovación y Gestión Tecnológica, Universidad Nacional de Colombia, sede Medellín; del Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia, y del Centro de Investigaciones del ITM. Correo electrónico: fagomez@unal.edu.co.

Rafael Germán Hurtado Heredia

Posdoctorado, Istituto Nazionale di Fisica Nucleare, Frascati, Italia. Doctorado, Università degli Studi di Perugia, Perugia, Italia. Físico, Universidad Nacional de Colombia, sede Bogotá, Colombia. Miembro del Grupo de Campos y Partículas, Universidad Nacional de Colombia. Profesor asociado de la Facultad de Ciencias, Universidad Nacional de Colombia. Correo electrónico: rghurtadoh@unal.edu.co.

Claudia Nelcy Jiménez Hernández

Magíster en Administración, Universidad Nacional de Colombia, sede Bogotá, Colombia. Ingeniera química, Universidad Nacional de Colombia, sede Bogotá. Miembro del Grupo Innovación y Gestión Tecnológica, Universidad Nacional de Colombia, sede Medellín, Colombia. Profesora auxiliar en dedicación exclusiva de la Universidad Nacional de Colombia, sede Medellín. Correo electrónico: cnjimenezh@unal.edu.co.

Claudia Jiménez Ramírez

PhD en Ingeniería-Sistemas, Universidad Nacional de Colombia, sede Medellín, Colombia. Magíster en Ingeniería de Sistemas, Universidad Nacional de Colombia, sede Medellín. Especialista en Programación y Bases de Datos, Universidad Nacional de Colombia, sede Medellín. Estadística, Universidad de Medellín, Medellín, Colombia. Miembro del Grupo de Investigación y Desarrollo en Inteligencia Artificial (GIDIA), Universidad Nacional de Colombia, sede Medellín. Profesora asociada en dedicación exclusiva en la Universidad Nacional de Colombia, sede Medellín. Correo electrónico: csjimene@unal.edu.co.

Florentino Malaver Rodríguez

Magíster en Economía, Universidad Nacional de Colombia, sede Bogotá, Colombia. Economista, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. Director del Grupo de Investigación *CINNCO* (Conocimiento, Innovación, Competitividad), Pontificia Universidad Javeriana, Bogotá, Colombia. Profesor asociado de la Facultad de Ciencias Eco-

nómicas y Administrativas, Pontificia Universidad Javeriana. Correo electrónico: fmalaver@javeriana.edu.co.

Jorge Manrique Henao

Magíster en Gestión Tecnológica, Universidad Pontificia Bolivariana, Medellín, Colombia. Ingeniero mecánico, Universidad Pontificia Bolivariana. Miembro del Grupo de Política y Gestión Tecnológica, Universidad Pontificia Bolivariana. Profesor de la Universidad Pontificia Bolivariana. Correo electrónico: jorge.manrique@upb.edu.co.

Diego Leonardo Otálora Montenegro

Estudiante de la Maestría en Ciencias Físicas, Universidad Nacional de Colombia, sede Bogotá, Colombia. Físico, Universidad Nacional de Colombia, sede Bogotá. Miembro del Grupo de Campos y Partículas, Universidad Nacional de Colombia. Correo electrónico: dlotaloram@unal.edu.co.

Carlos Quimbay Herrera

Doctor en Física Teórica, Universidad Autónoma de Madrid, Madrid, España. Magíster en Ciencias Físicas, Universidad Nacional de Colombia, sede Bogotá, Colombia. Físico, Universidad Nacional de Colombia, sede Bogotá. Director del Grupo de Campos y Partículas, Universidad Nacional de Colombia, sede Bogotá. Profesor asociado de la Facultad de Ciencias, Universidad Nacional de Colombia, sede Bogotá. Correo electrónico: cjquimbayh@unal.edu.co.

Juan Felipe Restrepo Arias

Estudiante de la Maestría en Ingeniería Administrativa, Universidad Nacional de Colombia, sede Medellín, Colombia. Ingeniero agrícola, Universidad Nacional de Colombia, sede Medellín. Miembro del Grupo de Investigación en Innovación y Gestión Tecnológica, Universidad Nacional de Colombia, sede Medellín. Correo electrónico: jfrestrepoa@unal.edu.co.

Martín Rico Herrera

Estudiante de la Maestría de Ingeniería de Sistemas, Universidad Nacional de Colombia, sede Medellín, Colombia. Ingeniero de sistemas e informática, Universidad Nacional de Colombia, sede Medellín. Miembro del Grupo de Investigación y Desarrollo en Inteligencia Artificial (GIDIA), Universidad Nacional de Colombia, sede Medellín. Correo electrónico: ericom@unal.edu.co.

Jorge Robledo Velásquez

PhD en Estudios de Política Científica y Tecnológica, Universidad de Sussex, Brighton, Inglaterra. Magíster en Sistemas de Generación de Energía Eléctrica, Universidad del Valle, Cali, Colombia. Ingeniero mecánico, Universidad Nacional de Colombia, sede Medellín, Colombia. Director del Grupo de Innovación y Gestión Tecnológica, Universidad Nacional de

Colombia, sede Medellín. Profesor asociado de dedicación exclusiva, Facultad de Minas, Universidad Nacional de Colombia, sede Medellín. Correo electrónico: jrobledev@unal.edu.co.

Sandra Carolina Serna Pineda

Magíster en Estadística, Universidad Nacional de Colombia, sede Medellín, Colombia. Ingeniera industrial, Universidad Nacional de Colombia, sede Medellín. Analista de riesgos de Bancolombia S. A., Medellín, Colombia. Correo electrónico: carolina.scsp2@gmail.com.

Jaime Humberto Sierra González

Master in Economics of the Internationalization of Business and Finance, Università di Roma "Tor Vergata", Italia. Master en Economía, Pontificia Universidad Javeriana, Bogotá, Colombia. Diploma en Relaciones Internacionales, Academia Diplomática de San Carlos, Bogotá, Colombia. Licenciado en Educación (Especialidad en Español e Inglés), Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. Miembro del Grupo de Investigación *CINNCO* (Conocimiento, Innovación, Competitividad), Pontificia Universidad Javeriana. Profesor asociado de la Facultad de Ciencias Económicas y Administrativas, Pontificia Universidad Javeriana. Correo electrónico: jhsierra@javeriana.edu.co.

Lida Isabel Tamayo Giraldo

Estudiante de la Maestría en Gestión Tecnológica, Universidad Pontificia Bolivariana, Medellín, Colombia. Ingeniera administradora, Universidad Nacional de Colombia, sede Medellín, Colombia. Miembro del Grupo de Política y Gestión Tecnológica, Universidad Pontificia Bolivariana. Docente interna de la Escuela de Ingenierías-Posgrados en Gestión Tecnológica, Universidad Pontificia Bolivariana. Correo electrónico: lida.tamayo@upb.edu.co.

Marisela Vargas Pérez

Economista, Universidad de los Andes, Bogotá, Colombia. Ingeniera industrial, Universidad de los Andes. Miembro del Grupo de Investigación *CINNCO* (Conocimiento, Innovación, Competitividad), Pontificia Universidad Javeriana, Bogotá, Colombia. Profesora asistente de la Facultad de Ciencias Económicas y Administrativas, Pontificia Universidad Javeriana. Observatorio Colombiano de Ciencia y Tecnología. Correo electrónico: marisela.vargas@javeriana.edu.co.

Fernán Villa Garzón

Estudiante de la Maestría en Ingeniería de Sistemas, Universidad Nacional de Colombia, sede Medellín, Colombia. Ingeniero de sistemas e informática, Universidad Nacional de Colombia, sede Medellín. Miembro del Grupo de Investigación Ingeniería y Software, Universidad de Antioquia, Medellín, Colombia, y del Grupo de Investigación en Computación Aplicada, Universidad Nacional de Colombia, sede Medellín. Profesor asistente del Departamento de Sistemas, Universidad de Antioquia. Correo electrónico: fernanvilla@gmail.com.co.