UNIVERSIDAD DE CANTABRIA

DEPARTAMENTO DE ADMINISTRACIÓN DE EMPRESAS

DOCTORADO EN ADMINISTRACIÓN DE EMPRESAS

ANÁLISIS DE LA ACTIVIDAD EN I+D+I EN EMPRESAS INDUSTRIALES: IDENTIFICACIÓN DE PATRONES Y RASGOS CARACTERÍSTICOS EN LA ALTA Y MEDIA ALTA TECNOLOGÍA

Doctorando: MARCO ANTONIO VILLAMIZAR ARAQUE

Directores: Dra. Rocío Rocha Blanco

Dr. Ángel Cobo Ortega

ABRIL, 2017

AGRADECIMIENTOS

A Dios por la vida y la oportunidad de realizar estudios de Doctorado.

A mis padres Magdalena y Luis por el amor y apoyo brindado desde siempre.

A mi esposa Victoria Helena por su presencia en mi vida.

A la Universidad de Cantabria por ser un espacio para el aprendizaje y a todos mis profesores por compartir sus experiencias y conocimiento.

A mis tutores la Doctora Rocío Rocha Blanco y el Doctor Ángel Cobo Ortega quienes con su conocimiento profundo, su ejemplo y amabilidad me han orientado en la investigación, ayudado ante la dificultad y motivado ante la adversidad.

TABLA DE CONTENIDO

1 CAPÍTULO 1. INTRODUCCIÓN, JUSTIFICACIÓN, ALCANCE Y OBJETIVOS 17

1.1 INTRODUCCIÓN 17

1.2 JUSTIFICACIÓN 19

1.2.1 ANTECEDENTES DE INVESTIGACIÓN EN SECTORES INDUSTRIALES DE ALTA Y MEDIA-ALTA TECNOLOGÍA 21

1.2.2 DIFERENTES ESTUDIOS DE EMPRESAS EN SECTORES INDUSTRIALES Y SUS ACTIVIDADES EN INNOVACIÓN E INVESTIGACIÓN Y DESARROLLO 23

1.2.3 ANTECEDENTES DE INVESTIGACIÓN EN LA UTILIZACIÓN DE MÉTODOS EMPÍRICOS PARA EL ESTUDIO DE LAS ACTIVIDADES PRESENTES EN INVESTIGACIÓN Y DESARROLLO (I+D) E INNOVACIÓN. 24

1.3 ALCANCE 25

1.4 OBJETIVOS 28

1.4.1 OBJETIVO GENERAL 28

1.4.2 OBJETIVOS ESPECÍFICOS 28

1.5 ESTRUCTURA DEL CONTENIDO 29

2 CAPÍTULO 2. MARCO TEÓRICO 31

2.1 INTRODUCCIÓN 31

2.2 INNOVACIÓN TECNOLÓGICA 32

2.3 INVESTIGACIÓN Y DESARROLLO (I+D) 35

2.4 ALTA Y MEDIA-ALTA TECNOLOGÍA 38

2.4.1 ALTA Y MEDIA-ALTA TECNOLOGÍA EN EL SECTOR INDUSTRIAL MANUFACTURERO 39

2.4.2 ORÍGENES DE LA CLASIFICACIÓN CNAE-2009 EN ESPAÑA. 40

2.5 INDICADORES PARA ALTA Y MEDIA-ALTA TECNOLOGÍA 42

2.6 ACTIVIDADES EN I+D+I EN LAS EMPRESAS 44

2.6.1 ACTIVIDADES DE COMERCIALIZACIÓN 46

2.6.2 ACTIVIDADES INNOVADORAS Y DE APOYO TECNOLÓGICO 46

2.6.3 INCENTIVOS Y FINANCIACIÓN DE LA I+D E INNOVACIÓN 48

2.6.4 ACTIVIDADES DE INNOVACIÓN 48

2.6.5 RECURSOS HUMANOS 49

3 CAPÍTULO 3. METODOLOGÍAS Y MÉTODOS DE ANÁLISIS 51

3.1 INTRODUCCIÓN 51
3.2 METODOLOGÍA DE INVESTIGACIÓN. 52
3.2.1 ORIGEN DE LOS DATOS 53
3.2.2 DESCRIPCIÓN DE LA MUESTRA DE DATOS DISPONIBLES 55
3.2.3 DEFINICIÓN DE VARIABLES 56
3.3 MÉTODOS DE ANÁLISIS 58
3.3.1 MINERÍA DE DATOS 58
3.3.2 TÉCNICAS ESTADÍSTICAS MULTIVARIANTES: ANÁLISIS FACTORIAL Y REGRESIÓN LOGÍSTICA BINARIA 65

4 CAPÍTULO 4. ANÁLISIS DE LOS ESTUDIOS REALIZADOS 75

4.1 ANÁLISIS EXPLORATORIO DESCRIPTIVO: PERFILES DE ACTIVIDADES EN I+D+I 76
4.1.1 CARACTERIZACIÓN DE LOS SECTORES MANUFACTUREROS DE ALTA Y MEDIA-ALTA TECNOLOGÍA FRENTE A OTROS SECTORES INDUSTRIALES 77
4.1.2 ANÁLISIS DESCRIPTIVO EN TORNO A VARIABLES CATEGÓRICAS 78
4.1.3 ACTIVIDADES INNOVADORAS Y DE APOYO TECNOLÓGICO 81
4.1.4 ACTIVIDADES DE INCENTIVOS Y FINANCIACIÓN EN I+D E INNOVACIÓN. 85
4.1.5 ACTIVIDADES DE INNOVACIÓN 86
4.1.6 RECURSOS HUMANOS 89
4.1.7 CONCLUSIONES DEL ANÁLISIS EXPLORATORIO DESCRIPTIVO 91
4.2 CARACTERIZACIÓN DE LAS EMPRESAS DE ALTA TECNOLOGÍA A PARTIR DE RASGOS Y PATRONES DE ACTIVIDAD DE I+D UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS 93
4.2.1 IDENTIFICACIÓN Y SELECCIÓN DE VARIABLES DISPONIBLES PARA LA REALIZACIÓN DE LA MINERÍA DE DATOS, Y TAMAÑO DE LA MUESTRA. 94
4.2.2 PREPROCESAMIENTO 95
4.2.3 APLICACIÓN DE TÉCNICAS DE EXPLOTACIÓN DE DATOS 96
4.2.4 ANÁLISIS Y EVALUACIÓN DE RESULTADOS. 107
4.2.5 CONCLUSIONES DEL ANÁLISIS DE APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS 108
4.3 ANÁLISIS FACTORIAL EXPLORATORIO 109
4.3.1 ESTADÍSTICAS DESCRIPTIVAS DE LA TOTALIDAD DE VARIABLES DISPONIBLES 110
4.3.2 ANÁLISIS FACTORIAL POR GRUPOS DE VARIABLES 114
4.3.3 ACTIVIDADES DE COMERCIALIZACIÓN 116
4.3.4 ACTIVIDADES INNOVADORAS Y DE APOYO TECNOLÓGICO 118
4.3.5 INCENTIVOS Y FINANCIACIÓN DE LA I+D E INNOVACIÓN 119
4.3.6 TALENTO HUMANO 120

4.3.7 CONCLUSIONES DEL ANÁLISIS FACTORIAL EXPLORATORIO UTILIZADO PARA IDENTIFICAR LA EXISTENCIA NO DE PATRONES OCULTOS 124
4.4 ANÁLISIS DE REGRESIÓN LOGÍSTICA BINARIA MULTIVARIANTE 124
4.4.1 SELECCIÓN DE VARIABLES PARA LA CLASIFICACIÓN DE EMPRESAS A PARTIR DE ACTIVIDADES DE I+D+I REQUERIDAS EN LA REGRESIÓN LOGÍSTICA BINARIA MULTIVARIADA. 125
4.4.2 MODELO DE CLASIFICACIÓN DE EMPRESAS DE ALTA Y MEDIA ALTA TECNOLOGÍA A PARTIR DE ACTIVIDADES EN I+D+I: MODELO DE REGRESIÓN LOGÍSTICA BINARIA MULTIVARIADA. 128
4.4.3 CONCLUSIONES DEL ANÁLISIS DE REGRESIÓN LOGÍSTICA BINARIA MULTIVARIADA..... 135
4.5 CONCLUSIONES GENERALES DEL CAPÍTULO 137

5 CAPÍTULO 5. DISCUSIÓN DE RESULTADOS, CONCLUSIONES, LIMITACIONES Y LÍNEAS DE TRABAJO FUTURAS 139

5.1 DISCUSIÓN DE RESULTADOS 140
5.2 CONCLUSIONES 148
5.3 LIMITACIONES DE LA INVESTIGACIÓN.....152
5.4 LÍNEAS DE INVESTIGACIÓN FUTURAS. 153

6 ANEXOS 157

6.1 DEFINICIÓN DE VARIABLES DISPONIBLES PARA EL ESTUDIO 157
6.2 RESUMEN DE LOS ESTADÍSTICOS DESCRIPTIVOS CON LAS VARIABLES TIPIFICADAS Y DISPONIBLES PARA EL ANÁLISIS FACTORIAL EXPLORATORIO. 189
6.3 RESUMEN GRÁFICOS Q-Q PARA EL ANÁLISIS VISUAL DE NORMALIDAD DE LAS VARIABLES.....191
6.4 RESUMEN DE DATOS PERDIDOS DISPONIBLE PARA EL ANÁLISIS DE REGRESIÓN LOGÍSTICA BINARIA MULTIVARIADA 200
6.5 REAL DECRETO 475/2007, POR EL CUAL SE APRUEBA LA CLASIFICACIÓN NACIONAL DE ACTIVIDADES ECONÓMICAS 2009 (CNAE-2009).....201
6.6 LISTADO DE SECTORES SEGÚN CNAE 2009: ALTA Y MEDIA-ALTA TECNOLOGÍA, SERVICIOS DE ALTA TECNOLOGÍA O DE PUNTA.....205
6.7 LISTA DE PRODUCTOS DE ALTA TECNOLOGÍA SEGÚN CÓDIGOS SITC.REV 4 Y CORRESPONDENCIAS CON CPA-2008 Y PRODCOM 2008 206

7 BIBLIOGRAFÍA 213

GLOSARIO DE SIGLAS.....239

LISTA DE FIGURAS

Figura 3.1. Representación de un modelo de factor común. Fuente: Elaboración propia a partir de Cea (2002). 68
Figura 3.2. Diagrama de decisión del análisis factorial, de la etapa 1 a la etapa 3. Fuente: Elaboración propia a partir de Hair et al, (1999). 69
Figura 3.3. Diagrama de decisión del análisis factorial, de la etapa 4 a la etapa 7. Fuente: Elaboración propia a partir de Hair et al, (1999). 70
Figura 3.4. Función Logística. Elaboración propia a partir de Kleinbaum, Klein, & Pryor (2010) 72
Figura 4.1. Sectores industriales presentes en la muestra. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) 77
Figura 4.2. Patrón en las actividades de comercialización. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 80
Figura 4.3. Actividades Innovadoras y de apoyo tecnológico. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 83
Figura 4.4. Patrón en las actividades tecnológicas. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) 85
Figura 4.5. Actividades de Innovación. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 88
Figura 4.6. Patrón en las actividades de innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 89
Figura 4.7. Personal contratado dedicado a actividades de I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 90
Figura 4.8. Árbol de decisión para la clasificación de empresas de alta tecnología a partir de los 9 atributos obtenidos en el ranking. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software Weka 98

Figura 4.9. Árbol de decisión para la clasificación de empresas de alta tecnología considerando ocho de los principales atributos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software Weka. 101
Figura 4.10. Variables a incluir en el análisis del modelo de regresión logística. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS. 132
Figura 4.11. Pruebas Omnibus. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS..... 132
Figura 4.12. Prueba de Hosmer y Lemeshow. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS. 133
Figura 4.13. Matriz de confusión. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS..... 133
Figura 4.14. Residuos normalizados. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS. 137

LISTA DE TABLAS

Tabla 2.1. Clasificación de la industria manufacturera en categorías de alta y media-alta tecnología. Fuente: INE (2016) 41
Tabla 2.2. Principales indicadores tecnológicos de alta tecnología. Fuente. Elaboración propia a partir del Observatorio Español de I+D+I (ICONO) (2013) e INE (2005). 44
Tabla 3.1. Encuesta Sobre Estrategias Empresariales (EESE). Elaboración propia a partir la Fundación SEPI (2017). 55
Tabla 3.2. Actividades de I+D+I, variables seleccionadas y autores que destacan su importancia. Fuente: Elaboración propia a partir de la revisión de la literatura científica. 58
Tabla 3.3. Etapas del proceso de extracción de conocimiento. Elaboración propia a partir de: Pérez & Santín (2007), Martín, Cabero, & De Paz (2008), Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004). 61
Tabla 3.4. Correspondencia entre las tareas identificadas y las técnicas que se aplican. Fuente: Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004) 64
Tabla 4.1. Actividades de comercialización. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 79
Tabla 4.2. Actividades Innovadoras y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 83
Tabla 4.3. Actividades de incentivos y financiación en I+D e Innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 86
Tabla 4.4. Actividades de innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 87
Tabla 4.5. Recursos Humanos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 90
Tabla 4.6. Variables seleccionadas para el análisis de minería de datos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 94
Tabla 4.7. Número de empresas de alta o media-alta tecnología en la muestra disponible para la minería de datos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 95

Tabla 4.8. Características de los dos grupos de empresas generados por el algoritmo de las k medias. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software Weka. 107
Tabla 4.9. Análisis Factorial Exploratorio, estadísticas descriptivas de las variables disponibles. Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS. 112
Tabla 4.10. Valores perdidos. Elaboración propia a partir de datos de la Fundación SEPI (2012). 113
Tabla 4.11. Estadísticos descriptivos de las actividades de comercialización. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 117
Tabla 4.12. Resumen de los modelos evaluados. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 117
Tabla 4.13. Estadísticos descriptivos de las actividades de innovación y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 119
Tabla 4.14. Resumen de los modelos elaborados en las actividades de innovación y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 119
Tabla 4.15. Estadísticos descriptivos de las actividades en incentivos y financiación en I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 120
Tabla 4.16. Resumen de los modelos en actividades de incentivos y financiación de la I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 120
Tabla 4.17. Estadísticos descriptivos de las actividades en relación al talento humano. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 121
Tabla 4.18. Resumen de los modelos elaborados para las actividades del talento humano. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 121
Tabla 4.19. Resumen de los modelos finales. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 123

Tabla 4.20. Definición de variables iniciales seleccionadas a incluir en el análisis del modelo de regresión logística binaria multivariada. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012). 127
Tabla 4.21. Resumen de los modelos de la primera fase de selección de variables. Elaboración propia. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012)..... 130
Tabla 4.22. Resumen de los criterios de decisión para la selección del modelo de regresión logística. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012)..... 131
Tabla 4.23. Variables y coeficientes del modelo de regresión logística de clasificación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS. 134
Tabla 6.1. Definición de variables disponibles para todo el estudio. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS..... 188
Tabla 6.2. Los estadísticos descriptivos con las variables tipificadas y disponibles para el análisis factorial exploratorio. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS..... 190
Tabla 6.3. Gráficos Q-Q de las variables tipificadas en el AFE. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS. 199
Tabla 6.4. Resumen de datos perdidos disponible para el Análisis de Regresión Logística Binaria Multivariada. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS. 200
Tabla 6.5. Listado de sectores según CNAE 2009: alta y media-alta tecnología, servicios de alta tecnología o de punta. Fuente: INE, en: www.ine.es/daco/daco42/daco4217/lstsectcnae.xls [Recuperado: 09-02-2017] 205
Tabla 6.6. Lista de productos de alta tecnología según códigos SITC.REV 4 y correspondencias con CPA-2008 y PRODCOM 2008. Fuente: Elaboración propia a partir de INE. http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=12547 36176979&menu=metodologia&idp=1254735576669 212

RESUMEN

Esta tesis está dirigida a analizar las actividades relacionadas con la I+D y la innovación en las empresas industriales de un alto o medio-alto nivel tecnológico. Se trata de identificar aquellos rasgos característicos que diferencian a las empresas manufactureras de alta y media-alta tecnología de las que pertenecen a otros sectores industriales. Para tal propósito se ha contado con datos de una muestra de 1540 empresas extraídos de la Encuesta de sobre Estrategias Empresariales de la Fundación SEPI. Haciendo uso de diferentes técnicas estadísticas y de minería de datos se han podido identificar las variables más importantes a la hora de poder realizar la clasificación de las empresas y descubrir las diferencias más notables con respecto a empresas de otros sectores manufactureros en España. La tesis incluye un pormenorizado análisis exploratorio descriptivo para posteriormente pasar a utilizar técnicas de clasificación supervisada y no supervisada, análisis factorial y regresión logística binaria. Con este trabajo se trata de ayudar a identificar los patrones y rasgos característicos que emergen en las actividades en I+D+I en las empresas industriales de alta y media-alta tecnología.

Palabras clave: I+D, innovación, sectores industriales, alta tecnología, minería de datos, modelos estadísticos.

ABSTRACT

This thesis is focused on the analysis of activities related to R&D and innovation in industrial companies of high and medium high technological level. The aim is to identify those characteristics or features that distinguish high and medium-high technology manufacturing companies from other industrial sectors. For this purpose data from a sample of 1540 companies is used. The data are extracted from the Survey on Business Strategies of the SEPI Foundation. The identification of the most important features in high and medium-high technology industries and the most notable differences with respect to companies from other manufacturing sectors in Spain had become possible thanks to the use of use of different statistical and data mining techniques. The thesis includes a detailed descriptive exploratory analysis to later use supervised and unsupervised classification techniques, factorial analysis and binary logistic regression. The aim of this work is to help identify the patterns that emerge from R&D and innovation activities in High and medium-high technology companies.

Keywords: R&D, innovation, industrial sectors, high technology, data mining, statistical models

1 CAPÍTULO 1. INTRODUCCIÓN, JUSTIFICACIÓN, ALCANCE Y OBJETIVOS

1.1 INTRODUCCIÓN

Los rápidos avances de la ciencia y la tecnología en la sociedad actual contribuyen a la cada vez mayor globalización de las economías y los mercados. Por otro lado, favorecen una economía basada en el conocimiento, en la que el cambio tecnológico y la innovación actúan de catalizadores e impulsores de la actividad económica. En este contexto, más que nunca, es esencial vigilar las tendencias y los cambios estructurales en la ciencia, la tecnología y la industria (OECD 1999).

Aunque es posible encontrar diferentes definiciones de tecnología, desde el punto de vista de este trabajo podríamos entender la tecnología como el conjunto de conocimientos necesarios para producir bienes y servicios (INE 2005). La alta y media-alta tecnología se caracteriza por la rápida renovación de conocimientos de base tecnológica, la complejidad de los mismos y el esfuerzo en investigación que requiere. A efectos estadísticos, las definiciones de alta y media-alta tecnología se suelen realizar mediante una enumeración exhaustiva de las ramas de actividad (enfoque por sectores) o de los productos (enfoque por productos) que son considerados de alto contenido tecnológico (INE 2005; 2011; 2013a; 2014; 2016).

Por una parte, la elaboración de una clasificación de actividades industriales basada en la intensidad tecnología presenta dificultades para establecer la definición de industria de alta tecnología, identificar un conjunto de criterios basados en el contenido tecnológico de una industria, y establecer los puntos de corte entre las clases de tecnologías (Hatzichronoglou 1997), lo que configura todo un reto de clasificación. No obstante, se cuenta con una serie de regulaciones y pautas para realizar la clasificación. En la actualidad, España dispone de la Clasificación Nacional de Actividades Económicas (CNAE-2009) con

el objeto de promover tal clasificación (Ministerio De Economía Y Hacienda 2007; INE 2009; 2012a).

De igual manera, resultado de las particularidades de la estructura de clasificación y el nivel máximo de desagregación permitido se presentan dificultades en torno a la disponibilidad de información, el secreto estadístico, la constante identificación y actualización de ramas productivas (o ramas de clasificación industrial), así como, la construcción de datos y series que permitan hacer comparaciones (Sociedad Asturiana de Estudios Económicos - SADEI 2008); también, en las regulaciones estadísticas nacionales se presentan dificultades (datos insuficientes) en aquellas categorías en las cuales existen muy pocas empresas (OECD 2011a).

Por otra parte, grandes esfuerzos para establecer temáticas, pautas y recomendaciones para la creación, medición y uso de información en temas de ciencia, tecnología e industria se han plasmado en importantes manuales, entre los que se podrían citar:

a) El Manual de Frascati 2002: Medición de las actividades científicas y tecnológicas: propuesta de una norma práctica para encuestas de investigación y desarrollo experimental (OCDE / FECYT 2003); y el Manual de Frascati 2015: Directrices para recopilar y reportar datos sobre investigación y desarrollo experimental.

b) El Manual de Oslo: Guía para la recogida e interpretación de datos sobre innovación (OECD/Eurostat 2007).

c) Manual de estadísticas de patentes (OCDE 2009).

A lo largo del periodo comprendido entre 1999 y 2015 múltiples estudios basados en indicadores de ciencia, tecnología e industria permiten confirmar el interés en el uso de información en relación a Investigación y Desarrollo (I+D), Innovación y la Clasificación de Actividades Económicas:

- Uso de estadísticas e indicadores de ciencia, tecnología e industria, y fuentes de datos que reflejen un mejor desempeño innovador (OECD 1999).

- Uso y evolución de clasificación mejorada de tecnologías. Interés en el uso y evolución de metodologías para el análisis e indicadores y fuentes de datos (OECD 2001).
- Necesidad y uso de nuevos indicadores en temas emergentes en torno a: Capital Humano (Movilidad internacional de los científicos), innovación medida por familia de patentes, biotecnología, nanotecnología, nuevos indicadores sobre la economía de la información, el papel de las empresas multinacionales, la productividad, la rotación de empresas y el desempeño industrial (OECD 2003).
- Análisis basados en estadísticas de clasificación industrial (OECD 2005c).
- Avances científicos y tecnológicos en los nuevos campos tecnológicos (OECD 2007a).
- Indicadores de ciencia, tecnología e industria que proporcionan información estadística en tiempos de recesión (OECD 2009a).
- Necesidad y uso de indicadores experimentales que proporcionan información sobre nuevas áreas de interés político (OECD 2011b).
- Nuevos indicadores basados en las nuevas tecnologías, características demográficas, comercio bilateral y la redefinición de indicadores tradicionales (OECD 2011c).
- Énfasis en que los indicadores y las estadísticas en ciencia, tecnología, innovación e industria ayuden a los gobiernos a diseñar políticas más eficaces y eficientes, y a supervisar los avances hacia las metas previstas (OECD 2013).

Por todo lo anterior, se observa un creciente interés en relación a la Clasificación Nacional de Actividades Económicas (CNAE), la Investigación y Desarrollo e Innovación (I+D+I), y la alta y media-alta tecnología.

1.2 JUSTIFICACIÓN

Estudios sobre la empresa en España destacan que el sistema empresarial español está compuesto en su mayoría por pequeñas y medianas empresas (OCDE/ FECYT 2007; Dirección General de Política de la Pyme 2010; 2010; 2014), y la estructura empresarial de base está configurada por las PYME de sectores de

tecnología media y media-baja. Por otro lado, las estadísticas expresan claramente la baja incidencia que tiene la innovación tecnológica (FECYT 2006).

Es de resaltar que a partir de estadísticas previas del INE se observa una tendencia, para el caso de España (2008-2012), de reducción en el número de empresas activas. Esta tendencia era un indicio del inicio de un periodo de crisis económica. Así, el número de empresas activas disminuyó un 1,6% durante el año 2011 y se situó en 3.199.617, siendo este el cuarto año consecutivo en el que el número de empresas activas se reducía (INE 2012b). De igual manera, el número de empresas activas disminuyó un 1,7% durante 2012 y se situó en 3.146.570, completando un quinto año consecutivo en el que el número de empresas activas se redujo (INE 2013b).

Analizando de nuevo la estructura del tejido empresarial español, el sector privado está compuesto por pequeñas y medianas empresas (PYMEs) en sectores tradicionales con baja intensidad investigadora respecto a los niveles internacionales, y es así que un escaso porcentaje de empresas españolas basan sus resultados en la innovación y la tecnología (OCDE/ FECYT 2007). Adicionalmente éstas PYMEs se caracterizan, en términos generales, por su escaso crecimiento en volumen de negocio y empleo, hecho que evidencia la necesidad de determinar las claves del éxito que configuran a aquellas empresas que efectivamente crecen, diversifican y crean empleo aportando un valor añadido (Dirección General de Política de La Pyme 2010). Por otra parte, los sectores con un elevado valor añadido, como son aquellos de alta y media-alta tecnología, requieren una preparación exigente, dado que no han experimentado el empuje deseado y todavía constituyen una minoría. El resultado de todo ello es que se es menos competitivo que otros países del entorno europeo (Instituto de Estudios Fiscales (IEF) 2014).

Investigaciones previas en diferentes países han establecido que las empresas de sectores de alta tecnología se caracterizan por operar bajo una serie de características internas determinadas por ambientes altamente cambiantes y competitivos (Duarte Castillo 2005; García-Manjón & Romero-Merino 2010; González Bañales 2007), por estar centradas en la innovación (Cordón et al. 2005) y con estrategias e inversiones en I+D (Wang & Wu 2012; Wang et al. 2013).

Dado el panorama previo en España, para la realización de la presente investigación se hace latente la necesidad de investigar sobre las características de las empresas de alta y media alta tecnología a partir de sectores industriales con un alto contenido tecnológico, siendo centrales los dos siguientes aspectos: 1) La clasificación de la actividad industrial a partir de la tecnología (Alta, media-alta, y otros sectores de manufactura) es relativa. 2) Los indicadores para realizar dicha clasificación atendiendo a aspectos de I+D e Innovación están en constante evolución. Tanto la clasificación de la actividad industrial como los indicadores para establecer sus características y niveles de desempeño, están sujetos a la dinámica de cambio y evolución tecnológica.

A continuación se procede a presentar los antecedentes de investigación que contemplan los sectores industriales de alta y media-alta tecnología, las actividades en I+D+I y las técnicas de análisis comunes para su estudio.

1.2.1 Antecedentes de investigación en sectores industriales de alta y media-alta tecnología

Al contemplar como objeto de estudio a los sectores industriales de alta y media-alta tecnología, se considera como fuente importante de información para realizar investigación, los esfuerzos científicos, conceptuales, metodológicos y estadísticos, realizados por entes nacionales y sectoriales para la recopilación de estadísticas y normalización de los datos.

A partir de un conjunto de estudios de las últimas décadas se puede observar un creciente interés por una línea de investigación enfocada en la alta tecnología y media-alta tecnología:

Hatzichronoglou (1997) plantea que en el contexto de globalización económica, la tecnología es un factor clave en la mejora del crecimiento y la competitividad en los negocios. Las empresas que son intensivas en tecnología innovan más, ganan nuevos mercados, utilizan los recursos disponibles de manera más productiva y por lo general ofrecen una mayor remuneración a las personas que emplean. Las industrias de alta tecnología son las que expanden con mayor fuerza en el comercio internacional y su dinamismo ayuda a mejorar el rendimiento en otros sectores.

Santos y Vázquez (1997) resaltan de la alta tecnología tres aspectos fundamentales:

1) La importancia de nuevos productos para garantizar una actividad rentable a largo plazo en las empresas de alta tecnología, teniendo en cuenta la rapidez con la que evolucionan las tecnologías, y el fenómeno de la obsolescencia.

2) La incertidumbre sobre las necesidades y requerimientos del mercado requiere el desarrollo continuado de nuevos productos que satisfagan las exigencias de los consumidores y evolucionen conforme lo hacen las preferencias de éstos.

3) La eficiencia en el desarrollo de nuevos productos en empresas de alta tecnología es sinónimo de su éxito.

Por otro lado, el acotamiento sectorial es recomendado en la literatura para la identificación y la medición de los recursos más críticos de las empresas (Alegre 2004); Grimpe & Sofka (2009) identifican en 13 países de Europa patrones en la actividades de innovación en empresas de los sectores alta tecnología, media y baja tecnología, destacando el acceso, fiabilidad y capacidad de transferencia de conocimientos.

En el caso particular de las empresas españolas, los resultados apoyan firmemente la opinión de que las actividades de I+D, como el diseño, el uso de maquinaria y la formación avanzada son cruciales para la comprensión del proceso de innovación de cualquier empresa. La evidencia empírica revela también la importancia de las fuentes externas, tales como el uso de consultores, la contratación de personal, acuerdos de colaboración y de I+D externa, con diferencias entre las empresas de baja y media tecnología y de alta tecnología (Santamaría et al. 2009).

En el caso de diversos ámbitos geográficos se observa cierta actividad que resalta el interés por la Innovación e Investigación y Desarrollo en empresas de alta tecnología: en un ámbito internacional (Becker & Hall 2013; Biscotti & D'Amico 2016; Chesbrough & Crowther 2006; De Jong & Freel 2010; Mishra 2010; Stam & Wennberg 2009; Wang & Wu 2012; Wöhrl et al. 2009; Zuniga & Crespi 2013; Wang et

al. 2013), en comparativos entre países en los cuales se incluye a España (Schmiele 2012; Grimpe & Sofka 2009; Sterlacchini & Venturini 2014; Cassou & Xavier de Oliveira 2011; Sáenz et al. 2009), e igualmente en el ámbito del contexto español (Santos & Vázquez 1997; Blanes & Busom 2004; Alegre 2004; Cordón et al. 2005; Santamaría et al. 2009; García-Manjón & Romero-Merino 2010; Delgado-Verde et al. 2011; PITEC 2011; FECYT 2012; Cruz-González et al. 2014; Hervas-Oliver et al. 2016).

Asimismo, se observa interés por investigar en relación a los patrones presentes en las actividades tecnológicas, de innovación y desarrollo tecnológico (Pavitt 1984; Grimpe & Sofka 2009; Heidenreich 2009; Dosi et al. 2008; FECYT/COTEC 2010; Lee 2010; Leydesdorff et al. 2013; Hirsch-Kreinsen 2015b; Zheng et al. 2016; Ryu & Lee 2016).

1.2.2 Diferentes estudios de empresas en sectores industriales y sus actividades en innovación e investigación y desarrollo

El estudio de fenómenos tecnológicos presentes en las empresas a partir de la observación por sectores industriales es común. A continuación se citan algunos de estos estudios sectoriales.

Feser (2001) estudia la función de producción en dos sectores manufactureros de Estados Unidos con diferentes intensidades de tecnología.

Vega, Gutiérrez & Fernández de Lucio (2010) realizan un estudio centrado en la industria manufacturera española, donde enfatizan en el riesgo de sobrestimar el papel de las fuentes externas de conocimiento y destacan que en muchos sectores industriales la mayor parte del esfuerzo innovador no solo es realizado por la propias empresas, sino que además se desarrolla en el interior de las mismas.

Cardamone (2012) reconoce la importancia del progreso tecnológico como fuente de crecimiento y los diferentes niveles de agregación de datos (empresas, industrias y países) con el fin de evaluar empíricamente el papel de los esfuerzos de innovación.

Guo, Gao & Chen (2013) presentan un análisis a nivel de la industria de empresas de fabricación de mediano y gran tamaño de China, identificando categorías de estrategia de tecnología.

Barbosa, Faria & Eiriz (2014) Investigan los factores subyacentes que podrían dar forma a las decisiones de la empresa con respecto a los grados de la novedad de la innovación, identificando que los factores históricos idiosincrásicos de la firma que dan lugar a las capacidades heterogéneas de I+D y el contexto de la industria tienen el poder para dar forma a las opciones de innovación de la empresa.

1.2.3 Antecedentes de investigación en la utilización de métodos empíricos para el estudio de las actividades presentes en Investigación y Desarrollo (I+D) e Innovación.

De igual manera, es usual la utilización de enfoques de análisis centrados en métodos empíricos para el estudio de las actividades en I+D+I y otras. Al respecto se observa que:

Schmiele (2012) elabora un análisis empírico de capacidades de absorción de conocimiento de las firmas y la capacidad de realizar investigación y desarrollo; Acosta y Coronado (2003) desarrollan un estudio en profundidad de los vínculos entre la ciencia y la tecnología en las Comunidades Autónomas de España; Gumbau-Albert & Maudos (2006) utilizan estudios empíricos para analizar la contribución del capital tecnológico y humano al desarrollo regional; Jamasb y Pollitt (2008) plantean un análisis de I+D en el contexto de sectores industriales mediante estudios empíricos; Lööf & Broström (2006) presentan un análisis econométrico del impacto de la colaboración con las universidades en la producción innovadora de las empresas; Aldieri & Cincera (2009) investigan la medida en que los de los efectos secundarios de I+D se intensifican por proximidades geográficas y tecnológicas.

Acosta, Coronado & Toribio (2011) plantean un modelo binomial negativo y modelo de *Poisson*, para cuantificar la intensidad en la generación de conocimiento tecnológico.

Fontana, Geuna & Matt (2006) establecen un modelo de regresión para analizar la variable dependiente (número de proyectos en I+D), basado en la distribución de *Poisson* y un modelo binomial negativo; Ciftci & Cready (2011) realizan un estudio empírico basados en un modelo de regresión transeccional para modelar las ganancias de la firma en función del tamaño, la actividad en I+D, entre otros; Yang et al. (2012) plantean en su análisis de la relación entre

estrategia e innovación, el uso del análisis de conglomerados y el modelo de regresión de mínimos cuadrados ordinarios; Block (2012) elabora un estudio empírico a partir de un modelo de regresión logarítmica del efecto de la inversión en I+D (medido a través de las ventas).

Wang & Wu (2012) emplean una función de producción para identificar el efecto del capital de I+D en las ganancias contables. Wang et al. (2013) exploran la relaciones entre los valores de mercado, la productividad, I+D y otros, mediante la técnica de análisis envolvente de datos.

Daunfeldt, Elert & Johansson (2016) examinan la relación entre alta tecnología, innovación y crecimiento, mediante un modelo de regresión *Logit* a partir de las empresas de alto crecimiento en industrias de la NACE[1] a cuatro dígitos.

En síntesis, se identifica la necesidad y pertinencia de investigar la alta y media-alta tecnología a partir de los análisis de las empresas presentes en sectores industriales intensivos en tecnología, atendiendo las particularidades de éstas en relación a actividades en Innovación e Investigación y Desarrollo. Así mismo, se identifica que el uso frecuente de técnicas estadísticas multivariadas para tal propósito y acorde a los objetivos de cada investigación.

1.3 ALCANCE

Esta tesis se centra en el análisis de las actividades de Innovación e Investigación y Desarrollo en el sector industrial de manufactura de alta y media-alta tecnología, contemplando la clasificación industrial de actividad CNAE 2009, a partir de un conjunto de 1540 empresas y datos de las mismas recolectados durante el 2012, en grupos de variables en torno a diferentes temáticas.

Así, este trabajo de investigación tiene como finalidad explorar, describir los patrones y rasgos presentes en empresas de manufactura consideradas de alta,

[1] NACE: Nomenclatura Estadística de Actividades Económicas de la Unión Económica Europea.

media-alta tecnología y asociados a la Innovación e Investigación y Desarrollo (I+D), para llegar finalmente a una alternativa complementaria de la clasificación de dichas empresas.

Debido a la evidencia empírica identificada a la fecha sobre la relación conjunta entre los temas centrales de la investigación (Innovación, I+D, Clasificación Nacional de Actividades Económicas -CNAE 2009, para el ámbito de España) y el objeto de estudio (empresas del sector de alta y media-alta tecnología), la investigación se considera de carácter exploratoria; en mayor grado descriptiva debido a la identificación y descripción de patrones y rasgos característicos en las actividades en I+D+I, y en menor grado correlacional puesto que se propone identificar un modelo para clasificar las empresas. Para todo ello se utilizan escalas cualitativas y cuantitativas presentes en el cuestionario que ha sido aplicado por parte de un ente especializado, de manera sistemática y con las prácticas de recolección de información que garantizan la calidad de los datos recolectados.

Para tal efecto, se identifica la importancia de analizar datos para años específicos, dado que no siempre es posible analizar las actividades industriales a partir de series o datos a lo largo del tiempo, producto principalmente de: i) Cambios en las clasificaciones industriales a lo largo del tiempo (Sociedad Asturiana de Estudios Económicos - SADEI 2008; INE 2009; 2012b; Ministerio De Economía Y Hacienda 2007) lo que dificulta preservar la consistencia de la serie de datos debido a que no necesariamente son comparables, ii) El desglose entre las industrias se hace, en principio, a nivel de la empresa. Las regulaciones estadísticas nacionales impiden la publicación de resultados donde hay muy pocas empresas en la categoría dada, de ahí las muchas lagunas en las series (OECD 2011a).

Se cuenta con la Encuesta Sobre Estrategias Empresariales (ESEE), la cual tiene su origen en un acuerdo suscrito entre el Ministerio de Industria y la Fundación SEPI. La encuesta abarca un grupo de empresas manufactureras a partir de un cuestionario, reportando de manera específica para esta tesis las observaciones correspondientes al año 2012.

Es de resaltar que, en la presente investigación y de acuerdo con Comrey (1985), la determinación de cuáles son las variables importantes es un paso

preliminar necesario antes de que sea posible comenzar las siguientes fases del trabajo científico.

En ésta tesis se plantea un enfoque que combina diferentes técnicas estadísticas y de minería de datos para abordar un conjunto de interrogantes centrales que orientarán la investigación, la identificación de patrones y la construcción de modelos multivariantes: ¿Cuáles son los patrones presentes en las actividades de Innovación e I+D (I+D+I) en las empresas manufactureras de sectores de alta y media-alta tecnología?, y ¿Cuáles son los rasgos relevantes para la clasificación de las empresas de manufactura en sectores de alta y media-alta tecnología frente a las empresas de otros sectores de manufactura en España?.

Hair et al. (1999) plantean desarrollar o construir modelos multivariantes, siguiendo de manera general los siguientes pasos: definir el problema de investigación, objetivos y técnica multivariante conveniente; desarrollo del proyecto de análisis; evaluación de la técnica multivariante requerida; estimación del modelo multivariante y valoración del ajuste del modelo; y finalmente, interpretación, análisis y/o validación de los resultados.

Otro de los aspectos importantes en la presente investigación, en este caso en relación con las técnicas multivariantes, es que no existe un método correcto único para el análisis de los datos. En lugar de eso y de acuerdo con Hair et al. (1999) se debe formular los objetivos de la investigación y aplicar la técnica seleccionada de manera adecuada para conseguir los objetivos de investigación. En algunos casos, las relaciones están especificadas estrictamente y el objetivo es una confirmación de la relación. Otras veces, las relaciones están vagamente reconocidas y el objetivo es el descubrimiento de las relaciones.

Debido a la necesidad de situar la innovación e investigación y desarrollo en un contexto de clasificación o identificación de sectores industriales, desde el punto de vista teórico, se han utilizado, las directrices adoptadas en España de las europeas correspondientes de la Clasificación Estadística de Actividades Económicas.

Se hace necesario resaltar que la innovación puede estar presente en cualquier sector de la economía (OECD/Eurostat 2007, p.23) e igualmente puede

ocurrir con la I+D. En esta tesis se plantea indagar en el sector empresarial, específicamente en la industria manufacturera.

Finalmente, se subraya en la presente investigación una especial atención a tres aspectos teóricos centrales: la innovación tecnológica, la I+D, la alta y media-alta tecnología a partir de la Clasificación CNAE 2009, destacando las actividades de I+D+I en empresas manufactureras de los sectores industriales de alta y media-alta tecnología.

1.4 OBJETIVOS

1.4.1 Objetivo general

El presente trabajo busca establecer, por medio de técnicas estadísticas multivariadas junto con métodos complementarios de minería de datos, patrones y rasgos a partir de actividades en la Innovación e I+D, que orienten la construcción de modelos para pronosticar la pertenencia o clasificación de empresas, en un sector industrial de alta y media-alta tecnología.

1.4.2 Objetivos específicos

- Explorar las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología a partir de un conjunto de variables de carácter tecnológico disponibles.
- Identificar y describir los patrones y rasgos presentes en las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología mediante la aplicación de diferentes técnicas de minería de datos como selección de atributos, árboles de decisión, reglas de asociación, clustering y/o análisis factorial exploratorio.
- Clasificar empresas industriales a partir de los rasgos característicos de la alta y media alta-tecnología y otros sectores de manufactura, mediante la obtención de un modelo de regresión logística binaria multivariada.

1.5 Estructura del contenido

El desarrollo de la presente tesis doctoral sigue una estructura por capítulos, y un conjunto de anexos respectivos.

En este capítulo uno se introduce el trabajo de investigación a partir de la justificación, el alcance y los objetivos propuestos. Posteriormente, en el capítulo dos se presentan los conceptos de Innovación, Investigación y Desarrollo, y alta y media-alta tecnología, la clasificación CNAE-2009, como las actividades en I+D+I de interés para la investigación, a partir de la revisión de la literatura científica disponible.

En el capítulo tres se aborda el enfoque de investigación propuesto, la fuente de los datos y descripción de la muestra disponible para el estudio. Se presentan igualmente los métodos de análisis seleccionados en el marco de la minería de datos.

En el capítulo cuatro se aborda el análisis de la actividad en I+D+I en las empresas disponibles a partir de la exploración, identificación, descripción de los patrones y posterior formulación de un modelo de clasificación para las empresas en la alta y media-alta tecnología. Se presentan tres secciones, una correspondiente al análisis exploratorio de datos, una segunda subsección dedicada al análisis de patrones a partir de algoritmos de minería de datos y técnicas estadísticas multivariadas, y finalmente, una tercera sección correspondiente a la elaboración de un modelo de clasificación de empresas a partir de los rasgos identificados mediante la técnica estadística de regresión logística binaria multivariada.

En el capítulo quinto se adelanta una discusión de resultados, conclusiones, identificación de limitaciones de la investigación, y líneas de investigación futuras de trabajo.

Se incluye una sección final de anexos que incluye la definición del conjunto total variables disponibles para la investigación, tablas de resumen complementarias para los análisis realizados, así como, un conjunto de anexos correspondientes a la normatividad por la cual se aprueba la Clasificación Nacional de Actividades Económicas, el listado de sectores de manufactura y

servicios identificados como de alta y media-alta tecnología, y un listado final de productos de alta tecnología.

2 CAPÍTULO 2. MARCO TEÓRICO

2.1 INTRODUCCIÓN

La transferencia de conocimiento y tecnología es uno de los procesos más susceptibles de mejora del Sistema de Innovación Español en al menos tres aspectos: el reducido número de empresas innovadoras, en comparación con las cifras medias de la Unión Europea; el perfil tecnológico medio y medio-bajo de la industria española (en comparación con los países centrales de la OCDE) y la reducida financiación privada de universidades y organismos públicos de investigación (OPIs) (FECYT/COTEC 2010). En este sentido, sectores orientados a la generación de alta tecnología requieren una preparación más exigente, dado que no han experimentado el empuje deseado y todavía constituyen una minoría, por ello, la necesidad de identificar los aspectos que configuran a aquellas empresas que crecen, diversifican y aportan valor (Dirección General de Política de La Pyme 2010). No basta pues, con saber si las empresas son innovadoras o no, es necesario saber cómo innovan y qué tipos de innovaciones introducen (OECD/Eurostat 2007).

Las actividades innovadoras de una empresa dependen en parte de la variedad y estructura de sus vínculos con las fuentes de información, del conocimiento, de las tecnologías, de las buenas prácticas, y de los recursos humanos y financieros. Cada vínculo conecta la empresa innovadora con otros agentes del sistema de innovación: laboratorios oficiales, universidades, autoridades reguladoras, competidores, proveedores y clientes. Los impactos de la innovación sobre los resultados de las empresas van desde los efectos sobre las ventas y la cuota de mercado a la mejora de la productividad y la eficiencia (OECD/Eurostat 2007).

La I+D desempeña un papel crucial en el proceso de innovación, aunque una gran parte de las actividades de innovación no se basan en ella. Un objetivo importante consiste en entender el papel de la I+D en el proceso de innovación y en determinar de qué manera se articula con los otros insumos de la innovación (OECD/Eurostat 2007).

En términos generales, la alta tecnología se caracteriza por una rápida renovación de conocimientos, muy superior a otras tecnologías (INE 2013a). Cuanto mayor sea la tasa de cobertura de alta tecnología, mayor será la capacidad de un país para comercializar internacionalmente los resultados de su investigación e innovación tecnológica en forma de productos de alto valor añadido (COTEC 2012). Por este motivo, el análisis de la alta tecnología proporciona una medida del impacto económico de las actividades de I+D (COTEC 2014) y es objeto de un tratamiento especial en las políticas públicas, especialmente las relacionadas con las PYMEs de alta tecnología (European Union 2016; Nunes et al. 2012).

Para el análisis de la alta tecnología, estudios previos han establecido que las empresas de sectores de alta tecnología se caracterizan por una serie de características internas determinadas por entornos cambiantes y competitivos (Cordón et al. 2005; Duarte Castillo 2005; González Bañales 2007; García-Manjón & Romero-Merino 2010).

A continuación se presenta una revisión de la literatura de la Innovación e Investigación y desarrollo (I+D), y las actividades presentes en ellas, así como, una síntesis en torno a la alta y media-alta tecnología, sus orígenes e indicadores.

2.2 INNOVACIÓN TECNOLÓGICA

La innovación es la introducción de un nuevo o significativamente mejorado producto y puede ser definida de manera más restrictiva (innovación tecnológica) como la introducción de un bien, servicio o proceso exitosamente en un mercado (OECD/Eurostat 2007). La innovación no es monolítica, y es crítico comprender el tipo de innovación en cuestión (Chesbrough & Crowther 2006).

Alrededor de la innovación en la empresa o grupos de empresas, se identifican los siguientes aspectos característicos de la misma:

- **Tipos de innovaciones**. Según la OECD (2005d) se distinguen cuatro tipos de innovaciones. En primer lugar, la innovación de producto es la introducción de un bien o de un servicio nuevo, o significativamente mejorado, en cuanto a sus características o en cuanto al uso al que se destina. La innovación de proceso es

la introducción de un nuevo o significativamente mejorado proceso de producción o de distribución. La innovación en mercadotecnia es la aplicación de un nuevo método de comercialización que implica cambios significativos del diseño o el envasado de un producto, su posicionamiento, su promoción o su tarificación. Y finalmente, la innovación de organización es la introducción de un nuevo método organizativo en las prácticas, la organización del lugar de trabajo o las relaciones exteriores de la empresa.

- **Grados de novedad e incorporación**. Pueden ser de índole incremental o radical (Barbosa et al. 2014; Zobel et al. 2017). Tal novedad está sujeta a un efecto moderador por el tipo de empresa e industria (Barbosa et al. 2014). El efecto de la novedad se puede observar en la incorporación de la tecnología en la empresa. Así, el grado en que las empresas incorporan las nuevas tecnologías en sus procesos y productos, se refleja en la relación entre innovación, tecnología y competitividad empresarial (FECYT 2006).

- **Tamaño de las empresas**. Se identifica como aspecto relevante el tamaño de las empresas. Las grandes empresas juegan un papel crucial en el proceso de innovación, mientras que las empresas pequeñas y medianas lo hacen en la aplicación de nuevos conocimientos en el mercado (Segarra-Blasco et al. 2008). El tamaño juega un papel importante en presencia de intensidad en I+D (Czarnitzki & Delanote 2012). Al respecto, las empresas jóvenes y pequeñas pueden llegar a ser innovadoras y con mayor expectativa de rendimientos anticipados a pesar de su gestión menos sistemática (Do 2014).

- **Actividades innovadoras.** Son las actividades diferentes a las de innovación, pero dirigidas al desarrollo o introducción de innovaciones tecnológicas. Comprenden las siguientes actividades: (1) I+D Interna, (2) I+D Externa, (3) Adquisición de maquinaria y equipo, (4) Adquisición de otros conocimientos para innovación, (5) Diseño y otros preparativos para producción y/o distribución, (6) Formación y (7) Introducción de innovaciones en el mercado (Observatorio Español de I+D+I (ICONO) 2013; 2014; 2015; 2016).

- **Actividades no incluidas en I+D**. La innovación abarca un cierto número de actividades, tales como las últimas fases del desarrollo de preproducción, la producción y distribución, las actividades de desarrollo con un bajo grado de

novedad, las actividades de preparación de formación y de comercialización, y actividades de desarrollo y de introducción para innovaciones que no son de producto o de proceso, tales como nuevos métodos de comercialización o nuevos métodos organizativos. Las actividades de innovación pueden también incluir las adquisiciones de conocimiento o bienes de capital que no son parte de la I+D (OECD/Eurostat 2007). Igualmente, la innovación requiere de actividades de aprendizaje colectivo, adquisición de conocimiento externo y tácito (Lawson & Lorenz 1999; Jeannerat & Crevoisier 2016), de la participación interorganizacional (Faems et al. 2005) o la participación en redes de alianzas tecnológicas (Gilsing et al. 2008; Chen 2014).

- **Las salidas y entradas de la innovación**. La salida más común de la innovación se mide por el número de patentes (Archibugi 1992; OECD 2005b; 2005c), el número de procesos y las innovaciones en los productos producidos. La medida más utilizada de la entrada a la innovación es el gasto en I+D (Gumbau-Albert & Maudos 2006). Por otra parte, los resultados o salidas de la innovación se pueden analizar por regiones o por sectores industriales (Jamasb & Pollitt 2008; Yang et al. 2012; OECD 2007a).

Finalmente, la innovación puede ser producto de la cooperación o sistemas de innovación. En el primer caso, las de tipo colaborativo, la innovación puede surgir a partir de actividades colaborativas subvencionadas (Marek et al. 2016), de una nueva concepción de innovación abierta (Berchicci 2013; Bonacci & Tamburis 2016; Huizingh 2011; Mortara & Minshall 2011), o de diferentes modalidades de relaciones con diferentes tipos de socios, y así, adquirir y explotar comercialmente tecnologías y conocimiento (Bianchi et al. 2011) a través de mecanismos formales e informales de apropiación (Zobel et al. 2017). En el segundo caso, las de tipo sistémico o de modelos, la innovación es producto de la creación y adopción de nuevas ideas y tecnologías, en el marco de subsistemas geográficos, socioeconómicos, políticos y legales (Zanello et al. 2016). Es importante entonces, la existencia de sistemas de innovación, dada la capacidad que tienen las organizaciones institucionales y empresariales, dentro de un área geográfica específica, de interactuar con el objetivo de asignar recursos a la realización de actividades orientadas a la generación y difusión de conocimiento, apoyando las innovaciones (Chen & Hung 2016; Buesa et al. 2006). Dichos sistemas de innovación

pueden ser sistemas nacionales, sistemas sectoriales, sistemas regionales, sistemas tecnológicos (Na-Allah & Muchie 2012), o de innovación local (Macneill & Jeannerat 2016) así como modelos de innovación centrados en territorios (Lagendijk 2006; Moulaert & Sekia 2003) o modelos de innovación de negocio centrados en cómo explotar la tecnología disruptiva (Chesbrough 2010), tales modelos median el vínculo entre la tecnología y los resultados de la empresa (Baden-Fuller & Haefliger 2013).

2.3 INVESTIGACIÓN Y DESARROLLO (I+D)

La investigación y el desarrollo experimental (I+D) comprenden el trabajo creativo y sistemático emprendido para aumentar el stock de conocimiento, incluyendo el conocimiento de la humanidad, la cultura y la sociedad, y para diseñar nuevas aplicaciones del conocimiento disponible (OECD 2015a). A su vez, la I+D es considerada como motor del crecimiento económico (Hansen & Winther 2011).

El término I+D cubre tres tipos de actividades: la investigación básica, la investigación aplicada y el desarrollo experimental. La investigación básica es un trabajo experimental o teórico emprendido principalmente para adquirir nuevos conocimientos de los fundamentos subyacentes de los fenómenos y hechos observables, sin ninguna aplicación o uso particular a la vista. La investigación aplicada, en cambio, es una investigación original emprendida para adquirir nuevos conocimientos que se dirigen principalmente hacia un objetivo específico y práctico. Finalmente, el desarrollo experimental es un trabajo sistemático, basado en el conocimiento adquirido de la investigación y la experiencia práctica y produciendo conocimiento adicional, dirigido a producir nuevos productos o procesos o a mejorar los productos o procesos existentes (OECD 2015a; OCDE / FECYT 2003). La investigación básica es abordada generalmente en las universidades y agencias de gobierno; la investigación aplicada se asocia comúnmente con la ingeniería; y el desarrollo se vincula con la posterior comercialización de los productos y los procesos productivos. Una vez realizada la investigación, la innovación se difunde a través de la industria vía licencias,

innovaciones patentadas, o adopción de innovaciones no patentadas (Tirole 1990).

La I+D apunta a nuevos hallazgos, basados en conceptos originales y su interpretación. Para que una actividad sea considerada realmente una actividad de I+D, debe satisfacer cinco criterios básicos: la actividad debe ser novedosa, creativa, incierta, sistemática, transferible y/o reproducible (OECD 2015a), sin embargo, rara vez se aborda la interacción al interior de la I+D, que suele considerarse como una entidad única, pero se puede considerar una clara distinción entre "I", de Investigación, responsable del desarrollo de nuevas tecnologías y "D", de Desarrollo, responsable de materializar en productos, procesos o servicios algunas de estas tecnologías (Maranzato et al. 2016).

Es importante destacar, que en el interior de la empresa se pueden realizar labores de investigación fundamental y aplicada para adquirir nuevos conocimientos y orientar su investigación hacia invenciones específicas o a la modificación de técnicas existentes (OCDE / FECYT 2003). En la mayoría de los casos, las actividades de I+D pueden agruparse para formar proyectos de I+D, los cuales se organizan y gestionan con un propósito específico, y tiene sus propios objetivos y resultados esperados (OECD 2015a). De esta manera, la empresa puede poner a punto nuevos conceptos de producto o proceso u otros nuevos métodos con el fin de evaluar su factibilidad y viabilidad, fase que puede incluir el desarrollo y los ensayos, y posteriores investigaciones para modificar los diseños o las funcionalidades técnicas (OECD/Eurostat 2007). También se encarga de identificar, asimilar y absorber el conocimiento externo disponible, mediante un área o departamento de la empresa (Aldieri & Cincera 2009).

En relación a la I+D se identifican los siguientes aspectos relevantes:

- **Los gastos en I+D**[2]. La OCDE (2003) establece que los gastos en I+D están compuestos por gastos internos y gastos externos. La medida básica la constituyen los gastos internos, que comprenden los gastos correspondientes a las actividades de I+D en una empresa o en un sector de la economía durante un período

[2] Para mayor detalle en relación a los gastos de I+D, se puede consultar el Manual de Frascati en la versión 2002 (Capítulo 6. Medición de los gastos dedicados a I+D), y en la versión 2015 ("Chapter 4. Measurement of R&D expenditures: Performance and sources found").

determinado. Cualquiera que sea el origen de los fondos, se miden los gastos corrientes y los gastos de capital. Los gastos corrientes comprenden los costes salariales de personal de I+D, compra de materiales, suministros y equipos de apoyo a la I+D, gastos corrientes cubiertos por financiación indirecta, alquiler de instalaciones, y costes de seguridad social y de pensiones para el personal de I+D, e impuesto sobre el valor añadido. Los gastos de capital son los gastos brutos anuales correspondientes a los elementos del capital fijo utilizados en los programas de I+D, y comprenden terrenos y edificios, equipos e instrumentos y software. Otra medida de los gastos en I+D, son los gastos externos que cubren los pagos de la I+D realizados fuera de la empresa o del sector.

- La gestión y otras actividades de apoyo a la I+D. La realización efectiva de las actividades de I+D exige la provisión de fondos, la gestión del proyecto y de su financiación. En el caso de una gestión interna de los proyectos de I+D y de su financiación se distinguen las actividades de apoyo directo a la I+D, de aquellas otras realizadas, por ejemplo, por los directores financieros, que no suministran más que un apoyo indirecto o auxiliar (OCDE / FECYT 2003). En este mismo sentido, las actividades indirectas de apoyo abarcan una serie de labores que no son I+D pero que proporcionan apoyo a I+D. Ejemplos típicos son transporte, almacenamiento, limpieza, reparación, mantenimiento y seguridad (OECD 2015a).

- **La I+D requiere de incentivos públicos y financiación (ayudas directas).** Los incentivos fiscales contribuyen a aumentar la probabilidad de que una empresa realice actividades de I+D+I, especialmente en las PYMEs. Las ayudas directas contribuyen a incrementar la probabilidad de que las empresas cooperen con organismos públicos de investigación y, en algunos casos, con proveedores o clientes (Busom et al. 2010). La distribución y el efecto de las subvenciones nacionales a la I+D difieren entre regiones con distinto nivel de concentración de las actividades de I+D (Herrera & Nieto 2010), y se ofrece apoyo público desde programas de distintos niveles gubernamentales (Fernández Ribas & Catalán 2010). El apoyo gubernamental a la I+D empresarial va en aumento, pero su demanda también es importante (OECD 2015d); de igual manera, se requiere de estímulos alternativos para que las empresas compartan el conocimiento cuando participan en procesos de I+D en un entorno incierto (Roelofs et al. 2016).

Un último aspecto observado es que la I+D puede ser resultado de la colaboración. Junto con la innovación, la I+D es usada como una medida o indicador del progreso de un sector industrial (Jamasb & Pollitt 2008) y puede ser resultado de colaboraciones (Marek et al. 2016) o redes colaborativas entre la industria y la universidad (Bychkova 2016). Más aún, es un mecanismo de cooperación para desarrollar innovación en producto y complementario a la innovación en proceso. Tal cooperación se analiza teniendo en cuenta el tipo de socio: industrial y no Industrial. Industrial como otras empresas, clientes, competidores y proveedores y no industrial como universidades, centros tecnológicos, consultores y laboratorios (Vega Jurado et al. 2010). La cooperación requiere de alianzas y mecanismos contractuales entre múltiples socios (Barbic et al. 2016). Obviamente, la I+D también puede ser resultado de colaboración internacional extraterritorial (Einola et al. 2016).

2.4 ALTA Y MEDIA-ALTA TECNOLOGÍA

Se considera la tecnología como el stock de conocimientos necesarios para producir nuevos productos y procesos. La alta tecnología se caracteriza especialmente por: i) una rápida renovación de conocimientos, muy superior a otras tecnologías, y ii) por su grado de complejidad, que exige un continuo esfuerzo en investigación y una sólida base tecnológica (INE 2005; 2013a; INE 2016; Observatorio Español de I+D+I (ICONO) 2013).

INE (2005) señala que, por la naturaleza cambiante de la alta tecnología, su definición igualmente lo es: "*la alta tecnología del momento será, si se mantiene, tecnología tradicional en el futuro. La velocidad con la que se produce el cambio tecnológico representa una dificultad para la medición de este fenómeno, ya que el ámbito de estudio puede variar de un momento a otro, lo cual dificulta la obtención de series históricas estables*".

A efectos estadísticos, la definición de alta tecnología se realiza mediante una enumeración exhaustiva de las ramas de actividad (enfoque por sectores) y de los productos (enfoque por productos) que son considerados, en un momento determinado, de alto contenido tecnológico (INE 2005; 2013a; 2014; 2016). Alternativamente, la alta tecnología y media-alta tecnología es descrita por la

OCDE (2011a) como industrias intensivas en I+D (aeroespacial, electrónica, maquinaria de oficina y computadoras, productos farmacéuticos e instrumentos).

A partir de lo expuesto por el INE (2005), la Clasificación Nacional de Actividades Económicas (CNAE) utilizada en España, es acorde a las clasificaciones que para tal efecto tiene la OECD, y EUROSTAT. Para la elaboración de la lista de sectores en función del grado de importancia de su tecnología, la OCDE estudia, a partir de la base de datos ANDERD (*Analytical Business Entrerprises Research and Development Database*), las intensidades en I+D de los distintos sectores industriales. Eurostat utiliza la clasificación de la OCDE, estableciendo la correspondencia con una agrupación de sectores de la NACE (Nomenclatura de Actividades Económicas de la Comunidad Europea), siguiendo la recomendación de alcanzar una segmentación de los sectores a tres dígitos de la NACE.

2.4.1 Alta y media-alta tecnología en el sector industrial manufacturero

Los sectores y productos considerados de "alta tecnología" (sectores industriales o de servicios) se seleccionan de acuerdo con la metodología propuesta por la OCDE y engloban aquellos que pueden definirse como de alto contenido tecnológico (Observatorio Español de I+D+I (ICONO) 2013).

Las industrias de manufactura son agrupadas en cuatro categorías acorde a su intensidad en I+D: alta, media-alta, media-baja y baja tecnología (OECD 2015c), a efectos estadísticos, las definiciones de alta tecnología y media-alta tecnología se suelen realizar mediante una enumeración exhaustiva de las ramas de actividad [3](enfoque por sectores) y de los productos[4] (enfoque por productos) que son considerados, en un momento determinado, de alto contenido tecnológico. Es de resaltar que al establecer definiciones para la alta y media-alta tecnología, se observa cierta ambigüedad, dado que, la elección de umbrales para separar los segmentos es arbitraria (INE, 2005). Por otro lado, según

[3] Dado que las unidades con una actividad económica común pueden ser clasificadas dentro de diferentes sectores institucionales, la clasificación por actividad económica puede proveer una fuente adicional de información sobre la estructura y dinámica de la I+D a través de la economía, y no solamente en el sector empresarial de negocios donde la principal clasificación de la actividad económica es más sistemáticamente aplicada. (OECD 2015a, p.92) (Traducción propia). Un listado de sectores de manufactura y servicios de alta tecnología se puede consultar en el Anexo 6.6.

[4] Una lista detallada de los productos de alta tecnología se puede consultar en el Anexo 6.7 de la presente tesis.

Hatzichronoglou (1997) la construcción de una clasificación completa de las industrias basada en su intensidad tecnológica presenta una serie de dificultades:

- Dificultad para establecer los criterios para identificar el contenido tecnológico de una industria, dada la dinámica de cambio en las tecnologías.
- Dificultad para establecer el concepto de una industria de alta tecnología: ¿Es la que produce tecnología, o la que usa la tecnología de una forma intensiva?
- Un tercer aspecto o problema es que siempre hay algún grado de arbitrariedad en la elección de los puntos de corte entre las clases de tecnología.

Cabe señalar que, por la propia naturaleza de la alta tecnología, su definición es cambiante en el tiempo (INE, 2005).

Eurostat (2007) hace énfasis en que los subsectores de manufactura de alta tecnología y media-alta tecnología son de gran importancia para la ciencia y la tecnología. En concordancia, durante los últimos años el INE realiza *análisis conjuntos para los sectores de alta y media-alta tecnología*, uniendo estos dos sectores y asociando datos estadísticos entre una y otra categoría (INE 2016; 2013a; 2011), igual fenómeno se observa en los análisis de la OECD (OECD 2005a; 2007a) y las cifras presentadas por el Observatorio Español de I+D+I (2013; 2014; 2015; 2016). Siguiendo esta vía de análisis de los sectores en conjunto, en el presente trabajo igualmente se asume la alta y media-alta tecnología como un único segmento industrial de estudio. Además, estudios previos (Delgado-Verde et al. 2011; Grimpe & Sofka 2009; Robertson et al. 2009; Santamaría et al. 2009; Santos & Vázquez 1997) resaltan los vínculos y particularidades al interior de cada uno de esos sectores y entre ellos.

2.4.2 Orígenes de la clasificación CNAE-2009 en España.

El MINISTERIO DE ECONOMIA Y HACIENDA (2007), a través del RD 475/2007[5], aprueba la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009),

[5] El texto correspondiente al RD 475/2007 se puede consultar en el Anexo 6.5 de la presente tesis.

con el objeto de reflejar los cambios estructurales de la economía, y en especial el desarrollo tecnológico. Este decreto se desarrolla a partir de la última revisión de la clasificación de actividades económicas (RD 1560/1992), denominada CNAE-93, que sufriría una actualización de orden menor en el año 2003. De acuerdo a esa clasificación CNAE-2009, la industria manufacturera de alta y media-alta tecnología, de acuerdo a la intensidad en I+D, estaría conformada por los sectores que se muestran en la Tabla 2.1.

CNAE 2009	**SECTORES**
	Sectores manufactureros de tecnología alta
21	Fabricación de productos farmacéuticos
26	Fabricación de productos informáticos, electrónicos y ópticos
30.3	Construcción aeronáutica y espacial y su maquinaria
	Sectores manufactureros de tecnología media-alta
20	Industria química
25.4	Fabricación de armas y municiones
27 a 29	Fabricación de material y equipo eléctrico; Fabricación de maquinaria y equipo n.c.o.p.; Fabricación de vehículos de motor, remolques y semirremolques
30- 30.1 - 30.3	Fabricación de otro material de transporte excepto: construcción naval; construcción aeronáutica y espacial y su maquinaria.
32.5	Fabricación de instrumentos y suministros médicos y odontológicos

Tabla 2.1. Clasificación de la industria manufacturera en categorías de alta y media-alta tecnología. Fuente: INE (2016)

Se desea enfatizar que, la clasificación CNAE garantiza que los sectores de actividad utilizados tengan la misma consideración en el marco de cualquier país miembro de la Unión Europea (Cordón et al. 2005). Para todas las fuentes de información utilizadas en el enfoque por sectores, los sectores han sido clasificados según la actividad principal de la empresa, entendiendo por tal aquella que genera el mayor valor añadido, o en su defecto la que proporcione el mayor valor de producción, o emplee un mayor número de personas ocupadas (INE 2005).

Al respeto, Eurostat[6] (2008) hace énfasis en que las clasificaciones dividen el universo de las observaciones estadísticas de acuerdo a conjuntos tan homogéneos como sea posible con respecto a las características del objeto de la encuesta estadística. Las clasificaciones estadísticas se caracterizan por la

[6] En: NACE Rev. 2 – Statistical classification of economic activities in the European Community.

cobertura exhaustiva del universo observado; las categorías mutuamente exclusivas; y los principios metodológicos que permitan una asignación coherente de los elementos a las distintas categorías de la clasificación.

Específicamente, las clasificaciones jerárquicas se caracterizan por una división más fina de las categorías, lo que permite recopilar y presentar la información a diversos niveles de agregación.

2.5 Indicadores para Alta y Media-Alta Tecnología

Los indicadores de alta tecnología se conciben precisamente como una medida de los resultados y del impacto de la I+D(INE 2011). Con este propósito, la OCDE (2011a) ha estado recogiendo datos de I+D de los países miembros de forma regular desde 1960. Durante la década de los noventa, recogió igualmente datos de I+D para una selección de economías de países no miembros.

De acuerdo con el INE (INE 2005) existen múltiples fuentes nacionales de estadísticas que presentan conjuntos de indicadores de I+D para el sector de alta y media-alta tecnología.

i) Estadísticas sobre actividades de I+D:
   - Proporción de empresas que realizan I+D interna.
   - Personal empleado en actividades de I+D, investigadores, técnicos, auxiliares.
   - Gasto en actividades en I+D interna.
   - Intensidad en I+D.

ii) Encuestas sobre innovación tecnológica en las empresas:
   - Innovaciones de productos, de procesos.
   - Actividades innovadoras.
   - Gastos en actividades innovadoras, en I+D interna y externa, en adquisición de maquinaria y equipo relacionado con productos y procesos tecnológicamente nuevos o mejorados, en adquisición de tecnología inmaterial, en formación relacionada con productos y

procesos tecnológicamente nuevos o mejorados, en comercialización de productos tecnológicamente nuevos o mejorados.

iii) Encuesta industrial de empresas:

- Cifra de negocios.
- Venta de productos.
- Valor añadido.

De acuerdo con el INE (2005) y el Observatorio Español I+D+I (2013), los principales indicadores que estas entidades usan para presentar la situación del sector de alta tecnología son los que se muestran en la Tabla 2.2.

Principales indicadores	**Descripción**
Gastos de I+D [en millones de euros]	Gastos internos en I+D se consideran los gastos internos en I+D a todas las cantidades destinadas a actividades de investigación científica y desarrollo tecnológico, realizadas dentro de la unidad o centro investigador cualquiera que sea el origen de los fondos. Los gastos internos comprenden tanto los gastos corrientes como los de capital.
Personal I+D [EJC]	Comprende el personal investigador, técnico y auxiliar. Los investigadores son profesionales que trabajan en la concepción o creación de nuevos conocimientos, productos, procesos, métodos y sistemas y en la gestión de los respectivos proyectos. Se incluye a los estudiantes postgraduados que desarrollan actividades de I+D. Los técnicos y el personal asimilado son personas cuyas tareas principales requieren conocimientos y experiencia de naturaleza técnica en uno o varios campos de la ingeniería, de las ciencias físicas y de la vida o de las ciencias sociales y humanidades. Participan en la I+D ejecutando tareas científicas y técnicas que requieren la aplicación de métodos y principios operativos, generalmente bajo la supervisión de investigadores. Los auxiliares (resto de personal) incluyen los trabajadores, cualificados o no, y el personal de secretariado y oficina, que participan en la ejecución de proyectos de I+D o que están directamente relacionados con la ejecución de tales proyectos.

Principales indicadores	Descripción
N° Empresas Innovadoras	Las empresas innovadoras son aquellas que han introducido en los últimos años productos tecnológicamente nuevos o mejorados en el mercado o procesos tecnológicamente mejorados en sus métodos de producción de bienes o de prestación de servicios.

Tabla 2.2. Principales indicadores tecnológicos de alta tecnología. Fuente. Elaboración propia a partir del Observatorio Español de I+D+I (ICONO) (2013) e INE (2005).

Adicionalmente, un conjunto particular de indicadores de comercio internacional en industrias intensivas en I+D es utilizado por la OECD (2011a). En este conjunto se incluyen indicadores tales como el total de importaciones, total de exportaciones y cuota de mercado de exportación.

No obstante, dada la existencia de indicadores para la actividad tecnológica en relación a la I+D, otros aspectos que pueden desempeñar un papel importante a la hora de identificar el desempeño de las empresas de alta y media-alta tecnología son:

- La tecnología incorporada en patentes, licencias y know-how, estrategias de cooperación técnica entre las empresas, la rápida obsolescencia de los conocimientos disponibles, la rotación rápida de los equipos (Hatzichronoglou 1997),
- Fuentes externas, tales como el uso de consultores, la contratación de personal, acuerdos de colaboración y de I+D externa (Santamaría et al. 2009),
- Y la adopción del conocimiento (OECD/Eurostat, 2007).

2.6 ACTIVIDADES EN I+D+I EN LAS EMPRESAS

Se consideran actividades en I+D+I todas aquellas labores presentes en la innovación, en investigación y desarrollo, así como, las actividades tecnológicas y de apoyo que conducen tanto a la investigación como a la innovación.

En principio, la OECD (2007) define un conjunto amplio de actividades que pueden reforzar las capacidades para desarrollar innovaciones o brindar apoyo a la I+D, incluso para adoptar con éxito innovaciones desarrolladas por otras

empresas. En primer lugar, por medio de su servicio de mercadotecnia y de sus relaciones con los clientes, una empresa puede desarrollar nuevos métodos de comercialización (venta de sus bienes y servicios), reorganizar sus sistemas de gestión, identificar posibilidades de comercialización que resultan de las investigaciones fundamentales o estratégicas, propias o ajenas, o monitorear a sus competidores, y acceder a consultores. En segundo lugar, la empresa puede comprar información técnica, pagar derechos o cánones por invenciones patentadas, o comprar conocimientos tecnológicos y experiencia recurriendo a los servicios de ingeniería, diseño o cualquier otro servicio de consultoría. Otro aspecto a tener en cuenta es que se puede invertir en equipos, programas informáticos, o insumos intermedios que incorporen el trabajo de innovación realizado por otros. Y finalmente, puede señalarse también el incremento de experiencia profesional, a través de la formación interna o la contratación de personal. En definitiva, el objetivo principal de todas estas actividades es mejorar los resultados de la empresa.

Con el propósito de dar un ordenamiento inicial a lo anteriormente expuesto, se complementa la identificación de las actividades en I+D+I por la OECD, con otros aspectos relacionados y presentes en la revisión de la literatura al respecto. Producto de ello, se identifican cinco temáticas en torno a las cuales, se seleccionan y asignan variables, que serán utilizadas para el análisis de las actividades en I+D+I.

Un primer aspecto que se incluye en este apartado, es el aspecto interno de la empresa asociado a su tamaño. El tamaño juega un papel determinante, en empresas grandes y pequeñas en relación a los resultados de innovación y desempeño en los mercados, especialmente en presencia de intensidad en I+D (Segarra-Blasco et al. 2008; Czarnitzki & Delanote 2012; Do 2014).

Las actividades de obtención de innovación (en producto, proceso, marketing y en gestión de la organización), al igual que las actividades de formación, se tratarán en un grupo específico, dada su relevancia. Las innovaciones como tal y el talento humano son la fuente del conocimiento central en la empresa.

El agrupamiento por temáticas y los referentes de la literatura considerados, se exponen a continuación.

2.6.1 Actividades de Comercialización

Una característica común a todos los tipos de innovación es que deben haber sido introducidos en el mercado (OECD/Eurostat 2007).

La actividad comercial de la innovación está relacionada con: la superioridad del producto (Cooper 1987), las características del proceso de comercialización, que a la vez es influenciado por el tamaño de la empresa, el nivel de I+D, el tipo de financiamiento, la novedad de la innovación y la protección de las propiedades intelectuales (Do 2014); asimismo con la actividad en exportación (Na-Allah & Muchie 2012; Chen 2012) e importación de tecnología (García Vega & Huergo 2010), y al desarrollo y transferencia de conocimiento a unidades de I+D fuera del país de origen (Schmiele 2012).

2.6.2 Actividades Innovadoras y de apoyo tecnológico

Se consideran como actividades innovadoras:

a) El gasto en I+D (tanto interno como externo) (OCDE / FECYT 2003; Observatorio Español de I+D+I (ICONO) 2012; 2013; 2014; 2015; 2016). Se destaca que la I+D hace parte de las actividades de innovación y posibilita una mayor introducción de la misma (OECD 2015d). Este gasto se dedica a crear tecnología propia mediante un proceso de I+D interna, o externa a la empresa mediante la compra de equipos, patentes u otros títulos de propiedad industrial (FECYT 2006).

b) La adquisición de maquinaria y equipo (Observatorio Español de I+D+I (ICONO) 2013; 2014; 2015; 2016), entre otras, en presencia de obsolescencia o rotación rápida de los equipos (Hatzichronoglou 1997). Los bienes de equipo (maquinaria de producción, equipos industriales y de oficina, equipos de medida, etc.) incorporan tecnologías innovadoras (COTEC 2012). La relación entre innovación tecnológica y la competitividad empresarial, depende del grado en que las empresas incorporan las tecnologías en sus procesos y productos (FECYT 2006). La

innovación implica inversión y puede incluir adquisiciones de activos materiales e inmateriales (OECD/Eurostat 2007).

c) Adquisición de tecnología y evaluación del cambio tecnológico. La selección de una tecnología de fabricación tiene una implicación importante en el desempeño del negocio y en toda la cadena de suministro. En particular, para el sector innovador que no se ocupa de materiales y tecnologías normalizados, la selección de tecnología es un tema central (Mastrocinque et al. 2016).

d) Adquisición de otros conocimientos para la innovación (Observatorio Español de I+D+I (ICONO) 2013; 2014; 2015; 2016), de adopción de conocimiento (OECD/Eurostat, 2007), en presencia de la rápida obsolescencia de los conocimientos disponibles. Estos pueden ser licencias y *know-how* (Hatzichronoglou 1997). Igualmente se puede considerar la adopción de la innovación abierta, la cual requiere diferentes modalidades organizativas, es decir, acuerdos de licencia, alianzas sin capital, compra y suministro de servicios técnicos y científicos. A través de éstas se pueden establecer relaciones con diferentes tipos de socios o adquirir y explotar comercialmente tecnologías y conocimientos (Bianchi et al. 2011). Se pueden considerar también labores de apropiación, requeridas para la consecución de las patentes, marcas registradas, derechos de autor y derechos de diseño (Zobel et al. 2017).

Se consideran como actividades de apoyo a la I+D:

a) Las actividades para la consecución de fondos, la gestión de proyectos y las encaminadas a la obtención de financiación (OCDE / FECYT 2003) o aquellas actividades de apoyo como el diseño y otros preparativos para producción y/o distribución (OECD/Eurostat 2007; OECD 2015a).

b) Las actividades de colaboración en I+D+I, generación y difusión de conocimiento. El desempeño en I+D está influenciado por las interacciones entre los actores en un sistema de innovación (Chen &

Hung 2016). La colaboración entre empresas competidoras se utiliza en varias industrias para lograr beneficios relacionados con la innovación (Ritala & Sainio 2014), para ello se ha propuesto, el acceso a consultores, el uso de acuerdos de colaboración de I+D externa (Santamaría et al. 2009) o la colaboración extraterritorial (Einola et al. 2016). Otras modalidades de colaboración son la que se produce con otros participantes del sistema de ciencia y tecnología, basada en los sistemas para la innovación (Na-Allah & Muchie 2012; Zanello et al. 2016; Macneill & Jeannerat 2016) y la capacidad de éstos para interactuar y asignar recursos (Chen & Hung 2016; Buesa et al. 2006). En esta categoría se pueden considerar la colaboración con las instituciones de investigación o universidades (Vega Jurado et al. 2010; Paunov 2013) o las redes entre la industria y la universidad (Bychkova 2016). Esta colaboración debe ser consecuencia de alianzas y mecanismos contractuales (Barbic et al. 2016); asimismo, se requiere de actividades de aprendizaje colectivo, adquisición de conocimiento externo y tácito (Lawson & Lorenz 1999; Jeannerat & Crevoisier 2016), de la participación entre organizaciones (Faems et al. 2005) o la participación en redes de alianzas tecnológicas (Gilsing et al. 2008; Chen 2014).

2.6.3 Incentivos y financiación de la I+D e Innovación

Como fuente de recursos son importantes las actividades de financiación de I+D de organismos públicos (OCDE / FECYT 2003) y el apoyo desde programas de distintos niveles gubernamentales (Fernández Ribas & Catalán 2010; OECD 2015b). Las subvenciones nacionales a la I+D surten efectos diferentes por regiones (Herrera & Nieto 2010). Los incentivos fiscales propician la realización de actividades de I+D+I, las ayudas directas contribuyen a la cooperación con organismos públicos de investigación, con clientes y proveedores (Busom et al. 2010), en presencia de estímulos alternativos en entornos con incertidumbre (Roelofs et al. 2016).

2.6.4 Actividades de Innovación

De acuerdo con la OECD (2005d), en este apartado, se identifica a las actividades de innovación inherentes a la Innovación de producto, la innovación de proceso, la innovación en mercadotecnia y de organización.

La innovación de producto es vital para la competitividad de la empresa (Cordón et al. 2005), se ve influenciada por la cooperación con proveedores (Fossas-Olalla et al. 2015) y presenta un impacto en la marca (Nørskov et al. 2015), pero se afecta por los escasos recursos de las pequeñas empresas (LÖFQVIST 2017). Por otro lado, se evidencia complementariedad entre producto y la innovación de procesos en presencia de los proyectos de desarrollo de productos (Hullova et al. 2016).

La innovación de procesos se ve afectada por las características de la firma (tamaño y características del propietario) y del entorno externo (Plotnikova et al. 2016). En las industrias tradicionalmente de procesos de transformación , este tipo de innovación, representa una gran parte de los recursos totales de desarrollo de la compañía (Bergfors & Lager 2011).

La Innovación en marketing refleja una relación favorable entre marca y distribuidores (Gupta et al. 2016). El volumen de negocios, las actividades de exportación y la I+D interna parecen ser las características que hacen que los minoristas tengan más probabilidades de desarrollar la innovación de marketing (Sáez & Pascual 2013).

La Innovación organizacional propende por nuevos procesos (Hecker & Ganter 2016) y la colaboración entre empresas la facilita (Mas-Verdú et al. 2016).

La innovación puede también mejorar los resultados reforzando la aptitud de la empresa a innovar. Es decir, la mejora de las capacidades de los métodos de producción puede permitir poner a punto una nueva gama de productos, y la adopción de nuevos métodos organizativos puede mejorar la aptitud de la empresa a apropiarse y a crear un nuevo conocimiento que podría utilizar para poner a punto otras innovaciones (OECD/Eurostat 2007). El impacto de las estrategias de innovación difiere entre la innovación de productos y procesos. La innovación de productos está motivada principalmente por inversiones en tecnología propia, mientras que la innovación de procesos se basa básicamente en estrategias de "compra" (Zuniga & Crespi 2013).

2.6.5 Recursos Humanos

Las actividades de innovación dependen del conocimiento disponible en todos los ámbitos y para ello se requiere de actividades de apoyo tales como

actividades de formación (OECD/Eurostat 2007), que favorezcan la introducción de dichas innovaciones (Observatorio Español de I+D+I (ICONO) 2013; 2014; 2015; 2016).

Niveles más altos de innovación están asociados con culturas que enfatizan el aprendizaje (Hurley & Hult 1998), para ello, se requiere de actividades de aprendizaje colectivo, adquisición de conocimiento externo y conocimiento interno (Lawson & Lorenz 1999; Jeannerat & Crevoisier 2016). La aplicación de la gestión del conocimiento apoya a una empresa en el desarrollo de productos innovadores (Su et al. 2006) y éste es considerado como un condicionante, o barrera para la innovación cuando es escaso o no está disponible (Segarra-Blasco et al. 2008). En consecuencia, para implementar las adaptaciones requeridas se requiere un conjunto de competencias humanas (Dos Santos & Neto 2007). La innovación implica la utilización de un nuevo conocimiento, un nuevo uso o una combinación de conocimientos existentes (OECD/Eurostat 2007), y así, elevar los niveles educativos contribuye a favorecer los tipos más complejos de innovaciones (Paunov 2013). Finalmente, las empresas exitosas invierten en mayor medida en las competencias de los trabajadores y la formación específica que permita hacer frente al cambio (OECD 2015d).

3 CAPÍTULO 3. METODOLOGÍAS Y MÉTODOS DE ANÁLISIS

3.1 INTRODUCCIÓN

Este tercer capítulo se dedica a presentar el enfoque de metodología de investigación seleccionado junto con la fuente de los datos y la descripción de la muestra, asimismo los métodos de análisis empleados para alcanzar los objetivos planteados.

Con el propósito de explorar, identificar, describir y clasificar (patrones y rasgos), a partir de la estructura de los datos, se contempla el enfoque de investigación exploratorio descriptivo[7], el cual se desarrollará a partir de un análisis de datos (centrado en la extracción de conocimiento) mediante un análisis exploratorio de datos, un análisis de minería de datos, un análisis factorial exploratorio y un análisis de regresión logística binaria.

Los estudios exploratorios sirven fundamentalmente para descubrir y prefigurar. Los estudios descriptivos son útiles para mostrar con precisión las dimensiones de un fenómeno, suceso, comunidad, contexto o situación (Hernández Sampieri et al. 2006). La minería de datos, por otro lado, permite clasificar y agrupar (Hernández Orallo et al. 2004) lo que es útil para identificar y describir de patrones, el análisis factorial exploratorio para identificar relaciones ocultas en los datos (Cea 2002) y el análisis de regresión logística para clasificar (Hernández Orallo et al. 2004; Jovell 1995; Pérez Hoyos 1996).

[7] Se destaca que estudios recientes aplicados al campo de la I+D e innovación, proveen evidencia de la aplicación del enfoque exploratorio descriptivo. Es así que, Schreiber, Ermel, Figueiredo, & Zeni (2016) plantean un estudio exploratorio descriptivo cuyo objetivo es analizar las innovaciones implementadas por empresas del sector industrial químico; Rua & Correia (2015) analizan la innovación en las organizaciones del sector sanitario portugués, con el fin de identificar sus factores críticos de éxito. En este estudio, los autores siguen un enfoque cuantitativo, combinando un estudio exploratorio, descriptivo y transversal.

3.2 METODOLOGÍA DE INVESTIGACIÓN.

Para el establecimiento de la metodología de investigación, se considera un referente a Hernández, Fernández y Baptista(2006), quienes realizan los siguientes planteamientos en relación a los estudios exploratorios descriptivos:

1. Los estudios exploratorios se realizan cuando el objetivo es examinar un tema poco estudiado o no abordado antes. Los estudios exploratorios sirven para familiarizarse con fenómenos relativamente desconocidos, obtener información sobre la posibilidad de llevar a cabo una investigación más completa respecto de un contexto particular, investigar nuevos problemas, identificar conceptos o variables promisorias, establecer prioridades para investigaciones futuras, o sugerir afirmaciones y postulados.

2. Por otro lado, los estudios descriptivos describen fenómenos, situaciones, contextos y eventos, detallan cómo son y se manifiestan. Los estudios descriptivos buscan especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis. Es decir, miden, evalúan o recolectan datos sobre diversos conceptos, aspectos, dimensiones o componentes del fenómeno a investigar; su objetivo no es indicar cómo se relacionan las variables medidas, más bien sirven para analizar cómo es y se manifiesta un fenómeno y sus componentes.

3. La utilidad y el propósito principal de los estudios correlacionales es conocer cómo se puede comportar un concepto o variable conociendo el comportamiento de otras variables relacionadas; es decir, intentar predecir. Los estudios correlacionales se distinguen de los descriptivos en que, mientras estos últimos se centran en medir con precisión las variables individuales, los estudios de correlación evalúan el grado de relación entre variables. Es decir, pretenden ver cómo se relacionan o vinculan diversos fenómenos entre sí, o si no se relacionan.

4. Ningún tipo de estudio es superior a los demás, todos son significativos y valiosos. La diferencia para elegir uno u otro tipo de investigación estriba en el grado de desarrollo del conocimiento respecto al tema a estudiar y a los objetivos planteados.

5. La investigación no experimental es la que se realiza sin manipular variables, es decir, no se hace variar intencionalmente las variables independientes, ésta procura observar fenómenos tal y como se dan en su contexto natural, para después analizarlos. Un tipo de investigación no experimental es la investigación transeccional o transversal, las cuales recolectan datos en un único momento, en un único tiempo y cuyo propósito es describir variables y analizarlas para un momento dado.

Atendiendo a las pautas anteriores, se considera que la investigación que se realiza en la presente tesis es de índole explotaría, debido a que en la realización de la revisión de la literatura disponible a la fecha, se detecta una escasa evidencia empírica sobre los antecedentes de investigación que contemplen las actividades en I+D+I en empresas en los sectores manufactureros de alta y media alta tecnología en el ámbito de España, y que dicha clasificación de alta y media-alta tecnología sea realizada a partir de la reciente actualización de la Clasificación Nacional de Actividades Económicas (CNAE-2009).

Así mismo, se considera la investigación de carácter descriptiva en mayor grado, ya que busca identificar, explorar y describir los patrones presentes en las actividades de I+D+I en las empresas manufactureras de alta y media alta tecnología, y en menor grado correlacional puesto que propone identificar un modelo para clasificar; no experimental dado que no hay manipulación a partir de variables independientes durante la recolección de la información y transversal pues los datos se recolectan en un único momento para un año específico.

3.2.1 Origen de los datos

Para los propósitos de esta investigación ha sido necesario disponer de un conjunto de datos e indicadores sobre la actividad de empresas de diferentes sectores manufactureros. Es importante disponer de datos que permitan realizar valoraciones en función del nivel tecnológico del sector al que pertenece cada empresa. En este caso, el origen de los datos es la fundación SEPI y las variables

utilizadas para el análisis de la actividad en I+D+I son en total ciento uno (101), con observaciones correspondientes al año 2012.

La Fundación SEPI tiene la condición de fundación del sector público estatal, por haberse constituido con la aportación exclusiva de la Sociedad Estatal de Participaciones Industriales (SEPI) y desde febrero de 2017, en cumplimiento de lo previsto en el artículo 128. 2 de la ley 40/2015 de 1 de octubre, se añadieron a su denominación las siglas F.S.P. (Fundación del Sector Público) quedando por tanto establecido como Fundación SEPI, F. S. P. (Fundación SEPI 2017).

3.2.1.1 La encuesta sobre estrategias empresariales (ESEE)

La Fundación SEPI realiza anualmente una encuesta de panel dirigida a empresas industriales manufactureras radicadas en España, denominada Encuesta Sobre Estrategias Empresariales.

En la Tabla 3.1 se muestra información sobre las características principales de esta encuesta.

ESEE	Descripción
Misión	La ESEE recoge un abanico amplio de decisiones de las empresas. Realiza un esfuerzo específico de depuración y validación de la información suministrada por la empresa para asegurar su calidad y consistencia temporal. Toda la información contenida en la ESEE está sometida a controles de validación y de consistencia lógica, sin que en ningún caso se proceda a la imputación de valores cuando hay falta de respuesta de la empresa.
Cobertura	La población de referencia de la ESEE son las empresas con 10 o más trabajadores de lo que se conoce habitualmente como industria manufacturera. El ámbito geográfico de referencia es el conjunto del territorio nacional y las variables tienen dimensión temporal anual. Una de las características más destacadas de la ESEE es su representatividad. La selección inicial de empresas se realizó combinando criterios de exhaustividad y de muestreo aleatorio.
Información que contiene	La ESEE está orientada a captar información acerca de las estrategias de las empresas, es decir, relativas a aquellas decisiones que adoptan sobre los instrumentos de competencia a su alcance. Resumen de contenidos del cuestionario de la ESEE: (a) actividad, productos y procesos de fabricación; (b) clientes y proveedores; (c) costes y

ESEE	Descripción
	precios; (d) mercados servidos; (e) actividades tecnológicas; (f) comercio exterior; (g) empleo; (h) datos contables.
Trascendencia	El valor de esta encuesta radica en la investigación aplicada que sustenta su base de datos, así como la realizada internamente por los servicios del Ministerio de Industria. Al cabo de los años, la Fundación ha logrado atesorar una muy importante base estadística de la industria española, con 17 millones de micro-datos, distribuidos entre 721 variables.

Tabla 3.1. Encuesta Sobre Estrategias Empresariales (EESE). Elaboración propia a partir la Fundación SEPI (2017).

En las siguientes subsecciones se describe el conjunto de datos y las variables concretas que se han analizado, todos ellas suministradas por la Fundación SEPI.

3.2.2 Descripción de la muestra de datos disponibles

Para el desarrollo de los objetivos de la investigación propuesta, se parte de un conjunto inicial de 1540 empresas manufactureras provenientes de 17 subsectores industriales de manufactura. Los 17 sectores, y por tanto las empresas que en ellos se encuentran, se distribuirán en dos grandes grupos: el sector compuesto por empresas de alta y media-alta tecnología y el compuesto por otros sectores de manufactura.

Se destaca que, para organizar los dos grupos de empresas se parte de la clasificación nacional de actividad económica que realiza el INE, basados en los códigos CNAE-2009 a tres dígitos. Esta clasificación permite identificar la actividad económica de cada subsector industrial de manufactura y, a la vez, identificar a qué sector pertenece cada empresa debido a su actividad económica principal.

Producto de la clasificación CNAE-2009 y previa especificación por parte del INE de aquellos sectores, que según su código CNAE-2009, son considerados de alta y media-alta tecnología, se establece qué empresas pertenecen a dicho grupo; en el contexto de la presente tesis, los sectores restantes se identificarán como otros sectores industriales de manufactura.

De acuerdo a estos criterios de clasificación, en la muestra se dispone para su uso en los diferentes análisis de 439 empresas identificadas en el sector

manufacturero de alta y media-alta tecnología. Estas empresas provienen de los sectores de: la industria química y productos farmacéuticos (137 empresas), la industria de productos informáticos, electrónicos y ópticos (31 empresas), la industria de maquinaria y material eléctrico, la industria de máquinas agrícolas e industriales (110 empresas) y la industria de vehículos de motor (87 empresas).

Igualmente se dispone para su uso, de 1101 empresas que se identifican en el grupo de otros sectores de manufactura: artes gráficas (64 empresas), bebidas (40 empresas), cuero y calzado (60 empresas), industria cárnica (82 empresas), industria de la madera (64 empresas), industria del mueble (85 empresas), industrial del papel (86 empresas), metales férreos y no férreos (59 empresas), productos alimenticios y de tabaco (216), productos de caucho y plástico (105 empresas), productos minerales no metálicos (124 empresas), textiles y confección (116 empresas).

Como se ha mencionado anteriormente, para cada una de las 1540 empresas se dispone de un conjunto variables relacionas con las actividades de I+D+I, información que ha recogido para cada una la Fundación SEPI, y en consistencia, con los códigos a tres dígitos de la clasificación de actividades económicas CNAE-2009.

Establecido el tamaño de muestra y la composición de los dos grupos de empresas: alta y media-alta tecnología, y otros sectores de manufactura, se procede en el siguiente apartado a introducir las variables seleccionadas y en secciones posteriores una descripción pormenorizada de las mismas.

3.2.3 Definición de variables

Se dispone para el análisis de la actividad en I+D+I de variables que previamente se han agrupado en cinco tipos de actividad (comercialización, actividades innovadoras y de apoyo tecnológico, incentivos y financiación de la I+D e Innovación, actividades de innovación y talento humano).

El conjunto total de variables por su extensión se presenta en el Anexo 6.1 en el que se identifican para cada una de las variables disponibles, su identificador (compuesto por siglas), nombre, definición, tipo, unidades y estado de la variable.

Es importante destacar que la selección de las variables se ha realizado tomando como referencia diferentes autores que plantean la importancia del análisis del correspondiente factor o característica a la hora de realizar una valoración global de las actividades de I+D+I de las empresas. La selección de las variables, su relación con las actividades en I+D+I y los autores que resaltan su importancia se presentan en la Tabla 3.2.

Actividad	**Variables seleccionadas a partir de la EESE de la Fundación SEPI para el estudio de las actividades en I+D+I**	**Autores que resaltan su importancia**
Actividades de Comercialización	COMPE, COMPE1, COMPN, COMPN1, COMSV, DMERPN, EXPORT, EXPTEC, GPIDSV1, IDMERP1N, IDMERPN, IGM, IMPORT, IMPTEC, NTES, NUMER, PM, PX, RXM, STEREL, VENTAS, VEXPOR, VIMPOR, WEBB2B, WEBB2C, WEBVEN.	(OECD/Eurostat 2007), (Cooper 1987), (Do 2014), (Na-Allah & Muchie 2012), (Chen 2012), (García Vega & Huergo 2010), (Schmiele 2012).
Actividades Innovadoras y de apoyo tecnológico	ACT, ADBEM, AID, AUTEC, AUTEC1, CTCL, CTCO, CTPR, CUCT, DCT, EPCT, ETAE, GEID, GEIDE, GIID, GPIDSV, GTID, IBESV, IBESV1, IDV, IDV1, IRI, MODUT, PAI PATESP, PATEXT, PEIT, PIUE, SPR, UAIT.	(OCDE / FECYT 2003), (Observatorio Español de I+D+I (ICONO) 2012; 2013; 2014; 2015; 2016), (OECD 2015d), (Hatzichronoglou 1997), (COTEC 2012), (FECYT 2006), (OECD/Eurostat 2007), (Mastrocinque et al. 2016), (OECD/Eurostat, 2007), (Zobel et al. 2017), (OECD/Eurostat 2007), (OECD 2015a), (Chen & Hung 2016), (Ritala & Sainio 2014), (Santamaría et al. 2009), (Einola et al. 2016), (Na-Allah & Muchie 2012), (Zanello et al. 2016), (Macneill & Jeannerat 2016), (Chen & Hung 2016), (Vega Jurado et al. 2010), (Paunov 2013), (Bychkova 2016), (Barbic et al. 2016), (Lawson & Lorenz 1999), (Jeannerat & Crevoisier 2016), (Faems et al. 2005), (Gilsing et al. 2008), (Chen 2014), (Marek et al. 2016), (Lagendijk 2006), (Moulaert & Sekia 2003), (Berchicci 2013), (Bonacci & Tamburis 2016), (Huizingh 2011), (Mortara & Minshall 2011), (Bianchi et al. 2011).
Incentivos y financiación de la I+D e Innovación	BSFE, DEDID, DEDIDG, DEDIT, DEDITG, DEDTOGN, DEDTON, FICS, FPIDCA, FPIDES, FPIDOT, FPIDTO.	(OCDE / FECYT 2003), (Fernández Ribas & Catalán 2010), (OECD 2015b), (Herrera & Nieto 2010), (Busom et al. 2010), (Roelofs et al. 2016).
Actividades de Innovación	ICOCAN, ICODIS, ICOFIJ, ICOPRO, IMO, IMOGE, IMOPE, IP, IPNC, IPNF,	(OECD 2005d), (Cordón et al. 2005), (Fossas-Olalla et al. 2015), (Nørskov et al. 2015), (LÖFQVIST 2017), (Hullova et

Actividad	Variables seleccionadas a partir de la EESE de la Fundación SEPI para el estudio de las actividades en I+D+I	Autores que resaltan su importancia
	IPNM, IPR, IPRME, IPRPI, IPRTM, NIP.	al. 2016), (Plotnikova et al. 2016), (Bergfors & Lager 2011), (Gupta et al. 2016), (Sáez & Pascual 2013), (Hecker & Ganter 2016), (Mas-Verdú et al. 2016), (OECD/Eurostat 2007), (Zuniga & Crespi 2013).
Talento Humano	GEFI, GEFIF, GEFIFN, GEFIN, GEFIT, GEFITC, GEFITN, GEFT, GEFTC, GEFTN, GEFVM, GEFVMN, IILR, REEID, REPID	(Observatorio Español de I+D+I (ICONO) 2012; 2013; 2014; 2015; 2016), (OECD/Eurostat 2007), (Hurley & Hult 1998), (Lawson & Lorenz 1999), (Jeannerat & Crevoisier 2016), (Su et al. 2006), (Segarra-Blasco et al. 2008), (Dos Santos & Neto 2007), (Paunov 2013, p.12), (OECD 2015d).

Tabla 3.2. Actividades de I+D+I, variables seleccionadas y autores que destacan su importancia. Fuente: Elaboración propia a partir de la revisión de la literatura científica.

3.3 MÉTODOS DE ANÁLISIS

Teniendo en cuenta los objetivos específicos de explorar, identificar y describir patrones y rasgos, y clasificar empresas de alta y media-alta tecnología a partir de los datos disponibles, se propone la utilización de técnicas de extracción de conocimiento desde la perspectiva de la minería de datos y otras técnicas estadísticas de análisis multivariados.

3.3.1 Minería de datos

La minería de datos puede definirse como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar los datos. Está compuesta por técnicas de análisis útiles en la obtención de información y análisis, dado que permiten la evaluación automática o semiautomática de volúmenes de datos, la identificación de patrones comportamentales, tendencias y relaciones, proporcionando así una mejor visión del fenómeno de interés. Como resultado de un extenso proceso de investigación y desarrollo, dentro de esta disciplina se incluyen técnicas avanzadas de análisis de datos (árboles de decisión, regresión lineal, redes neuronales artificiales, técnicas bayesianas, máquinas de soporte vectorial, etc.), provenientes de diversos campos (aprendizaje automático

e inteligencia artificial, estadística, bases de datos, etc.) y aborda una tipología variada de problemas (clasificación, categorización, estimación/regresión, agrupamiento) (Witten & Frank 2005; Pérez López & Santín González 2007; Lakshmi & Raghunandhan 2011; Seng & Chen 2010).

La minería de datos es una de las disciplinas de la matemática aplicada de mayor aplicabilidad práctica en una enorme variedad de campos. Por su afinidad con la temática de ésta tesis, podrían citarse estudios recientes donde se utiliza la minería de datos como soporte de información para las actividades de innovación tecnológica de las empresas (Lucheng et al. 2010), o para adelantar investigación sobre innovación (Leydesdorff et al. 2013), en áreas tales como la gestión estratégica de tecnologías (Limonge et al. 2007), la Innovación abierta (Di Benedetto 2010) y redes de colaboración (Rio et al. 2011), análisis de patentes (Lucheng et al. 2010; Cordeiro et al. 2014), la aparición de tecnologías emergentes y su financiación (Hopkins & Siepel 2013). También pueden encontrarse aplicaciones en el análisis de la importancia del conocimiento del cliente en el desarrollo de productos innovadores (Su et al. 2006) o el aprendizaje innovador en el desarrollo de la estrategia de marca de servicio (Wang et al. 2014).

La minería de datos se desarrolla en un contexto donde la información está disponible y crece de manera exponencial; más que un producto de la era de la información, también es una nueva forma de superar el volumen de datos en un mundo en progreso (Wang & Yuan 2011).

Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004) resumen el fin último de la minería de datos como la capacidad para analizar los datos para extraer conocimiento, el cual puede adoptar la forma de:

a) Relaciones, patrones o reglas inferidos de los datos y previamente desconocidos, o bien en forma de una descripción más concisa, es decir, un resumen de los mismos.

b) Modelos, que pueden ser de dos tipos: predictivos y descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés (variables objetivo o dependientes), usando otras variables independientes o predictivas. Los modelos descriptivos, en cambio, identifican patrones que explican o resumen los datos, es decir,

sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos.

Finalmente, la minería de datos puede verse como una etapa de un proceso completo denominado de extracción de conocimiento a partir de datos (*Knowledge Discovery in Databases - KDD*).

Todo proceso de minería de datos debe pasar por cuatro etapas: i) preparación de los datos o pre-procesamiento (selección, limpieza y transformación), ii) exploración, iii) aplicación de técnicas de explotación de datos, iv) análisis y difusión de resultados (Witten & Frank 2005; Pérez López & Santín González 2007).

3.3.1.1 El proceso de extracción de conocimiento (KDD)

En la Tabla 3.3 se presenta la síntesis del proceso a seguir para llegar a extraer conocimiento a partir de un conjunto de datos que pueden provenir de diferentes fuentes.

Etapa	Descripción de la etapa
Preparación de los datos. Pre-procesamiento	Selección: se integran y recopilan los datos, se identifican y se seleccionan las variables relevantes de los datos y de ser requerido se aplican técnicas de muestreo (Pérez López & Santín González 2007).
	Limpieza de los datos: Esta actividad es necesaria dado que los datos pueden contener valores atípicos, valores faltantes y valores erróneos. Se analiza la influencia de los valores atípicos, se imputan los valores faltantes y se eliminan o corrigen los datos incorrectos (Pérez López & Santín González 2007); la presencia de datos atípicos y valores desaparecidos pueden llevarnos, en etapas posteriores según sea el caso, a usar algoritmos robustos (p.ej. árboles de decisión), a filtrar información, a reemplazar valores mediante técnicas de imputación y a transformar datos continuos en discretos mediante técnicas de discretización (Pérez López & Santín González 2007).
	Transformación de los datos: de ser necesaria, se lleva a cabo la transformación de los datos, generalmente mediante técnicas de reducción o aumento de la dimensión y escalado simple y multidimensional. Entre las técnicas avanzadas de transformación

Etapa	Descripción de la etapa
	se tiene las de reducción (por ejemplo: Análisis Factorial Exploratorio o Análisis de componentes principales (Pérez López & Santín González 2007).
Exploración	En la exploración se usan técnicas de análisis exploratorio de datos como histogramas, diagramas, etc., se realizan el análisis de medias, varianzas, frecuencias o correlaciones, según sea el caso dependiendo de la naturaleza de la variable (cuantitativa o cualitativa) y los objetivos planteados.
Aplicación de las técnicas de explotación de Minería de Datos	En esta etapa se decide el objetivo a alcanzar (clasificar, agrupar, etc.,) y se elige la técnica descriptiva o predictiva que se va a utilizar (Pérez López & Santín González 2007). Es en esta etapa donde se generan los modelos, los cuales son una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicción, para entender mejor los datos o para explicar situaciones pasadas (Martín et al. 2008). En la construcción del modelo es donde se ve mejor el carácter iterativo del proceso de extracción de conocimiento, ya que será necesario explorar modelos alternativos hasta encontrar aquel que resulte más útil para resolver el problema (Hernández Orallo et al. 2004).
Análisis y difusión	En la fase de análisis e interpretación se evalúan los patrones y, si es necesario se vuelve a las fases anteriores para una nueva iteración. Idealmente, los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles (es decir, inteligibles) e interesantes (útiles y novedosos) (Hernández Orallo et al. 2004).
	En la etapa de difusión se realiza la divulgación de resultados, para ello, se cuenta con herramientas de visualización o presentación de datos. En la difusión se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios (Pérez López & Santín González 2007).

Tabla 3.3. Etapas del proceso de extracción de conocimiento. Elaboración propia a partir de: Pérez & Santín (2007), Martín, Cabero, & De Paz (2008), Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004).

Establecido el proceso general para la aplicación de la minería de datos, se procede a exponer las tareas que se han de implementar en el análisis de los datos, los cuales se presentarán más adelante en el Capítulo 4.

3.3.1.2 Tareas de la minería de datos

En esta sección se procede a presentar una síntesis de las tareas o problemas a ser resueltos por los algoritmos de minería de datos utilizados en el análisis de las actividades en I+D+I en las empresas manufactureras de alta y media-alta tecnología. Un estudio detallado de las tareas seleccionadas y otras, se encuentra en el libro de referencia de Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004).

Para los objetivos específicos planteados de explorar, identificar, describir y clasificar, y de acuerdo con los planteamientos de los autores de referencia, se selecciona como tareas a realizar, la clasificación mediante técnicas de selección de atributos y árboles de decisión, reglas de asociación y clustering.

Según Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004) dentro de la minería de datos se ha de distinguir diferentes tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Esto significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra. A continuación las tareas definidas por Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004):

1. **Clasificación**: también conocida como clasificación supervisada, es quizá la tarea más utilizada. En ella, cada instancia (o registro de la base de datos) pertenece a una clase, la cual se indica mediante el valor de un atributo que llamamos la clase de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno corresponde a una clase. El resto de atributos de la instancia se utilizan para predecir la clase. El objetivo es predecir la clase de nuevas instancias de las que se desconoce la clase. Más concretamente, el objetivo del algoritmo es maximizar la razón de precisión de la clasificación de las nuevas instancias, la cual se calcula como el cociente entre las predicciones correctas y el número total de predicciones (correctas e incorrectas). Una de las opciones más habituales para la clasificación es la utilización de árboles de decisión, formados por un conjunto de condiciones organizadas en una estructura jerárquica y que permiten comprender de manera gráfica un modelo expresado como un conjunto de normas de decisión.

2. **Reglas de asociación**: son tareas descriptivas muy similares a las correlaciones, que tienen como objetivo identificar relaciones no explícitas entre atributos categóricos. Pueden ser de muchas formas, aunque la formulación más común es del estilo "*si el atributo X toma el valor A entonces el atributo Y toma el valor de B*". Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados. Un caso especial de reglas de asociación, que recibe el nombre de reglas de asociación secuenciales, se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo.

3. **Clustering**: también conocida como agrupamiento o clasificación no supervisada, es la tarea descriptiva por excelencia y consiste en obtener grupos "naturales" a partir de los datos. Se habla de grupos y no de clases, porque, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, se analizan precisamente para generar esta etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando al mismo tiempo la similitud entre los distintos grupos. Es decir, se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí, al mismo tiempo, son muy diferentes a los objetos del otro grupo. Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos.

Obviamente, a la hora de realizar estas tareas es posible la utilización de una gran variedad de algoritmos o técnicas. En la Tabla 3.4 se presentan algunas de estas técnicas con su vinculación con las tareas expuestas y su carácter predictivo o descriptivo.

Técnica	**Predictivo**		**Descriptivo**		
	Clasificación	**Regresión**	**Agrupamiento**	**Reglas de asociación**	**Correlaciones**
					Factorizaciones
Redes neuronales	✓	✓	✓		

Técnica	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Reglas de asociación	Correlaciones Factorizaciones
Árboles de decisión ID3.C4.5 C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓		✓	
Regresión logística	✓			✓	
Kmeas			✓		
Apriori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial					✓
Twostep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores de soporte	✓	✓	✓		
CN2 rules (cobertura)	✓				
Análisis discriminante multivariante	✓				

Tabla 3.4. Correspondencia entre las tareas identificadas y las técnicas que se aplican. Fuente: Hernández Orallo, Ramírez Quintana, & Ferri Ramírez (2004).

A partir de lo anteriormente expuesto, se procede a plantear la selección de tareas y métodos que permitan alcanzar los objetivos específicos de identificar, explorar, describir y clasificar. En concreto, se aplicarán tareas de clasificación y las técnicas seleccionadas serán los árboles de decisión y la regresión logística. En cuanto a las tareas de clustering se utilizarán las técnicas *kmeans*; el algoritmo "*Apriori*" para reglas de asociación y para la identificación de factores ocultos, las técnicas de análisis factorial.

Después de identificar las tareas y técnicas que se van a aplicar, se continua con las técnicas estadísticas multivariantes que se han seleccionado para las tareas de clasificación e identificación de patrones ocultos.

3.3.1.3 Fase previa a la aplicación de tareas y técnicas de minería de datos. Análisis exploratorio de datos

Siguiendo la propuesta de (Martín et al. 2008) se propone iniciar la extracción de conocimiento a partir de las actividades en I+D+I con un Análisis Exploratorio y Descriptivo de los de Datos (AED), cuyo propósito es examinar los datos previamente a la aplicación de cualquier técnica estadística. De esta forma se consigue un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas. El AED proporciona métodos sencillos para organizar y comprender los datos. En la aplicación de cada técnica estadística multivariante se procede a detectar datos ausentes, identificación de casos atípicos y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes. El análisis exploratorio de datos postula que previamente a cualquier análisis de datos es necesario un análisis visual de esto. Este análisis previo tiene los mismos objetivos que la estadística descriptiva, pero pretende además detectar en los datos patrones o modelos.

Una vez que se haya realizado este análisis exploratorio de datos, y la identificación de patrones a través de tareas y técnicas de minería, se requiere según los objetivos planteados, ahondar en la comprensión de los patrones para detectar así los rasgos característicos u ocultos.

3.3.2 Técnicas estadísticas multivariantes: análisis factorial y regresión logística binaria

En las secciones anteriores se ha introducido el uso de la minería de datos, el proceso de aplicación para la extracción de conocimiento, se han identificado las tareas o problemas que resuelve y la correspondencia de éstas con las diferentes técnicas. En este apartado se presenta con mayor detalle el análisis factorial exploratorio y la regresión logística binaria, dado que su aplicación se abordará desde una perspectiva clásica de análisis estadístico multivariante[8].

[8] El análisis multivariante cumple también con los objetivos descriptivo e inductivo, pero trabajando simultáneamente con varias variables en lugar de una sola. Es una metodología estadística más complicada, pero también más potente, que utiliza extensamente los métodos del álgebra lineal, cálculo numérico, geometría lineal y otras clases de geometría (ultramétrica y no euclídea) (Cuadras, 1991).

Se entiende por Análisis Multivariante a la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de n>1 variables, que pueden ser de tipo cuantitativo, cualitativo o una mezcla de ambos. La información en análisis multivariante es, por lo tanto, de carácter multidimensional. Quizás la propiedad más determinante del análisis multivariante es que las n variables son dependientes, de naturaleza similar y que ninguna de ella tiene importancia superior a las demás (Cuadras 1991).

Cuando se considera la aplicación de tales técnicas, la primera cuestión que uno se debe hacer es: ¿Qué tipo de relaciones están siendo examinadas?. En este caso se distinguen las relaciones de dependencia e interdependencia.

i) *De dependencia*: donde una variable o grupo de variables es identificado como la variable dependiente y las demás son identificadas como independientes (Hair et al. 1999). Si se desea predecir una variable dependiente en una relación única y las variables son cuantitativas y cualitativas la técnica recomendada es del tipo de regresión logística (Pardo Merino & Ruíz Díaz 2002).

ii) *De interdependencia*: donde ninguna variable o grupo de variables es definida como dependiente o independiente, en vez de esto, todas las variables son analizadas simultáneamente en un esfuerzo por encontrar una estructura subyacente para el conjunto total de variables o sujetos. Si se está analizando la estructura de las variables, se recomienda la técnica de análisis factorial (Hair et al. 1999).

3.3.2.1 Análisis factorial

El análisis factorial[9] (AF) es un método de análisis multivariante que intenta explicar, según un modelo lineal, un conjunto extenso de variables observables mediante un número reducido de variables hipotéticas llamadas factores. Es decir, intenta identificar variables subyacentes, o factores, que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas.

[9] "Los antecedentes inmediatos del análisis factorial se remontan a 1904 cuando Karl Spearman publicó un artículo con el título "*General intelligence objectively determined and mesuared*" (En la revista *American Journal of Psycology*, 15:201-293) que trata sobre la covariación entre variables. Pero fue en 1947 cuando Thurstone (en *Múltiple Factor Analysis, The university of Chicago Press*) incorpora al análisis factorial el álgebra matricial para el análisis de las correlaciones. A partir de entonces se suceden las publicaciones sobre el análisis factorial en cualquiera de sus variantes (Cea 2002).

El análisis factorial se suele utilizar en la reducción de los datos para identificar factores que expliquen la mayoría de la varianza observada en un número mayor de variables manifiestas. También puede utilizarse para generar hipótesis relacionadas con los mecanismos causales o para inspeccionar las variables para análisis subsiguientes (Martín et al. 2008). El análisis factorial puede ser utilizado cuando se requiera cumplir cualquiera de estos dos objetivos: la identificación de la estructura mediante el resumen de datos, o bien la reducción de datos (Hair et al. 1999).

El modelo analítico de AF se asemeja a una ecuación de regresión múltiple. Cada variable observada X_i se expresa mediante una combinación lineal de un número pequeño de factores comunes latentes y un factor único, también latente. Estos últimos representan la parte de la varianza de la variable observada que "no" es explicada por los factores comunes. Se asume que son independientes, lo que quiere decir que no se hallan correlacionados entre sí (unos factores únicos con otros factores únicos) ni tampoco con los factores comunes. Algebraicamente, el modelo AF puede representarse mediante las ecuaciones siguientes:

$$X_{1=}\,\lambda_{11}\;F_1+\;\lambda_{12}\;F_2+\cdots+\lambda_{1K}\,F_K+e_1$$
$$X_{2=}\,\lambda_{21}\;F_1+\;\lambda_{22}\;F_2+\cdots+\lambda_{2K}\,F_K+e_2$$

.....

$$X_{P=}\,\lambda_{P1}\;F_1+\;\lambda_{P2}\;F_2+\cdots+\lambda_{PK}\;F_P+e_p$$

Para cada variable X_i, siendo i = (1,2,…, p), el modelo se resume en:

$$X_i=\sum_{j=1}^{k}\lambda_{ij}\,F_j+\;e_i$$

Donde "k" representa el número de factores comunes y "λ_{ij}" las saturaciones o coeficientes factoriales. Estos últimos cuantifican la relación de los factores no observables con cada indicador. En el desarrollo del modelo se intenta que estos coeficientes sean o muy elevados o muy pequeños, en los distintos factores comunes. Lo que lleva a la rotación del modelo factorial. La finalidad es alcanzar

un modelo factorial que sea lo más sencillo e interpretable posible, en el que cada indicador se relacione con un número mínimo de factores.

Un ejemplo de este modelo se muestra en la Figura 3.1, que corresponde a un modelo de factor común, compuesto por tres factores comunes y 5 indicadores.

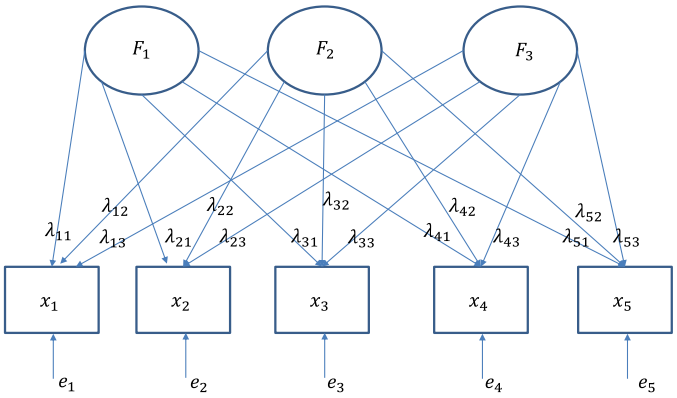


Figura 3.1. Representación de un modelo de factor común. Fuente: Elaboración propia a partir de Cea (2002).

En el análisis factorial, los valores teóricos (factores) se forman para maximizar su explicación de la serie de variables entera, y no para predecir una(s) variable(s) dependiente(s)(Hair et al. 1999).

Los factores representan las variables originales con una pérdida mínima de información. En este conjunto de factores es donde se estudian las relaciones entre ellas. Se puede distinguir entre dos objetivos: extraer la varianza máxima o reproducir del mejor modo posible las correlaciones entre todas las variables de estudio. Para que el análisis factorial tenga sentido deberían cumplirse dos condiciones básicas: parsimonia e interpretabilidad. Según el principio de parsimonia, los fenómenos se deben explicar con el menor número de elementos posibles. Por tanto, respecto al análisis factorial, el número de factores deber ser lo más reducido posible y éstos deben ser susceptibles de interpretación sustantiva. Una buena solución factorial es aquella que es sencilla e interpretable (Martín et al. 2008).

En relación a la cantidad de variables a considerar, el incremento del número de las variables aumenta la posibilidad de que las variables estén no correlacionadas y no sean representativas de unos conceptos distintos. Es necesario saber cómo se relacionan las variables para interpretar mejor los resultados. Si el número de variables es demasiado grande, o existe la necesidad

de dar una mejor representación a un número de conceptos más pequeño en vez de facetas múltiples, el análisis factorial puede asistir en la selección de un subgrupo de variables representativo o incluso crear nuevas variables como sustitutas para las variables originales mientras mantengan su carácter original (Hair et al. 1999). Así mismo, las características de medida de las variables de datos y también los sujetos de los que se obtienen los datos pueden tener una gran influencia sobre los resultados obtenidos (Comrey 1985).

En las Figura 3.2 y Figura 3.3 se muestra una propuesta de un proceso con siete etapas para realizar un análisis factorial: las tres primeras etapas asociadas a la metodología de investigación (identificación del problema, diseño de investigación, y verificación de supuestos), y posteriormente las etapas específicas para el análisis factorial exploratorio: selección del modelo de factores, métodos de rotación, especificación del modelo.

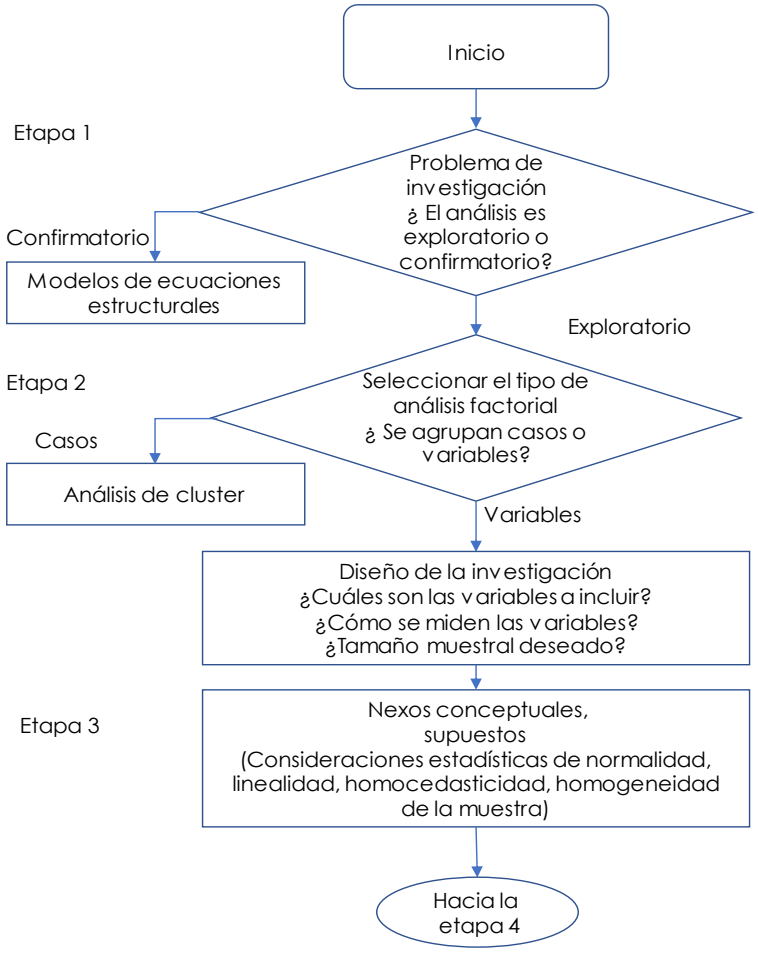


Figura 3.2. Diagrama de decisión del análisis factorial, de la etapa 1 a la etapa 3. Fuente: Elaboración propia a partir de Hair et al, (1999).

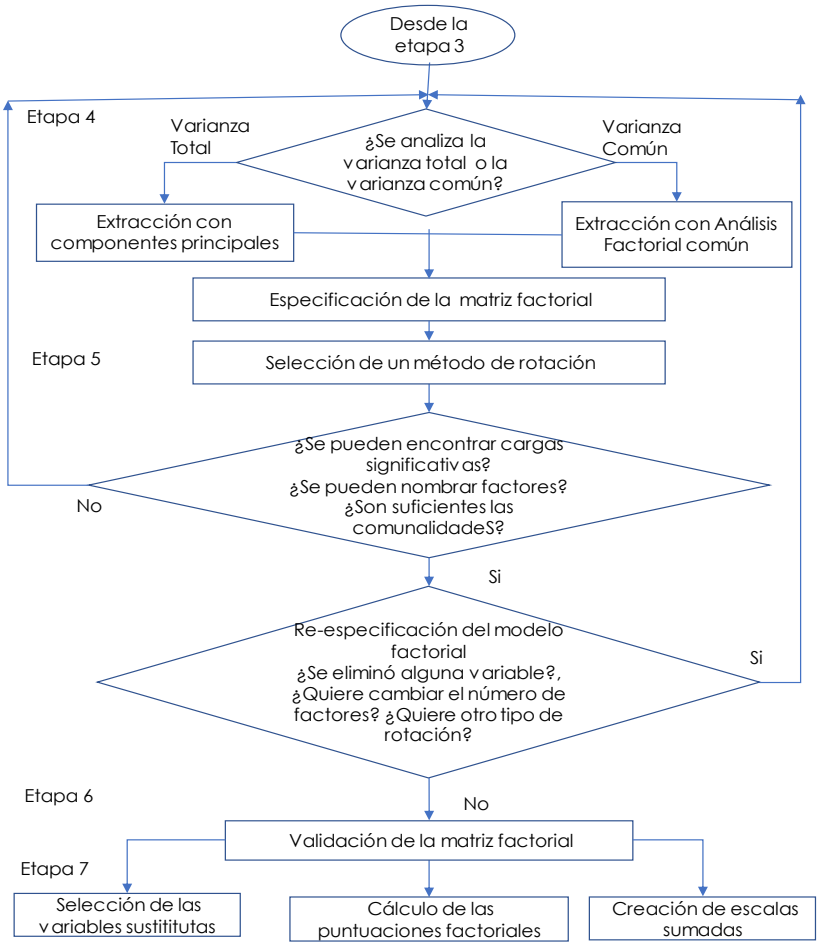


Figura 3.3. Diagrama de decisión del análisis factorial, de la etapa 4 a la etapa 7. Fuente: Elaboración propia a partir de Hair et al, (1999).

Las técnicas analíticas de factores pueden lograr sus propósitos desde una perspectiva exploratoria o confirmatoria. Muchos investigadores lo consideran meramente exploratorio, útil para la búsqueda de una estructura entre una serie de variables o como un método de reducción de datos. Desde esta perspectiva, las técnicas de análisis factorial "extraen lo que proporcionan los datos". No obstante, en otras situaciones se tiene unos pensamientos preconcebidos sobre la restructura real de los datos, que se basan en un apoyo teórico o investigaciones previas. Es posible que se desee probar las hipótesis que implican como qué variables deberían ser agrupadas en un factor o el número exacto de factores. En estos casos, se requiere un una análisis factorial que adopte un enfoque confirmatorio, es decir, valorar hasta qué punto los datos se ajustan a la a la estructura esperada (Hair et al. 1999).

El análisis factorial exploratorio se aplica en las primeras fases de análisis, en el estadio "exploratorio", cuando se quiere agrupar variables muy correlacionadas, como un fin en sí mismo, o como un paso previo y necesario a la aplicación de otras técnicas analíticas multivariables. De su práctica resulta la obtención de un modelo hipotético de la estructura común latente en los datos analizados. Hasta la fecha ésta ha sido la aplicación más frecuente del análisis factorial: "explorar" las dimensiones latentes en los datos más que confirmarlas (Cea 2002).

3.3.2.2 Regresión logística binaria multivariante

La regresión logística es un enfoque de modelado matemático que se puede utilizar para describir la relación de varios variables independientes a una variable dependiente dicotómica (Kleinbaum 1998; Jovell 1995).

El análisis de regresión logística tiene como finalidad principal pronosticar la pertenencia a un grupo a partir de una serie de variables independientes (Pérez Hoyos 1996), permite identificar las características que diferencian dos grupos y genera pronósticos que ayudan a diferenciar entre los miembros de uno u otro grupo y permite utilizar variables independientes categóricas(Pardo Merino & Ruíz Díaz 2002).

Para llevar a cabo un análisis de regresión logística binaria es necesario disponer de una variable categórica que defina dos grupos. Esta variable dicotómica es la variable dependiente del análisis, es decir, la variable cuyos valores se desea pronosticar. Y para averiguar en qué se diferencian los grupos es necesario disponer de la información (cuantificada en una serie de variables) en la que se supone que se diferencian. Estas variables en las que se supone se diferencian los grupos se utilizan como variables independientes o predictoras o covariables. El análisis de regresión logística genera una serie de pesos o coeficientes que: (1) informan sobre la capacidad individual de cada variable independiente para diferenciar entre los grupos y (2) permiten obtener pronósticos que sirven para clasificar sujetos (Pardo Merino & Ruíz Díaz 2002).

3.3.2.3 El modelo de regresión logística

El modelo de regresión logística se utiliza para predecir la probabilidad estimada "*P(Y)*" de que una variable dependiente "*Y*" presente uno de los dos

valores posibles (1=sí o 0=no) en función de los diferentes valores que adoptan el conjunto de variables independientes "*Xi*", normalmente se intenta predecir la probabilidad de que se produzca el acontecimiento o suceso definido como Y=1. En otras palabras, el modelo de regresión logística permite relacionar una variable dependiente dicotómica con una o más variables independientes cuantitativas y/o cualitativas. Las variables dicotómicas son aquellas que definen, mediante los indicadores (1,0), dos características mutuamente excluyentes y opuestas, como son la presencia "1" o ausencia "0" de un acontecimiento o suceso (Jovell 1995).

Considérese la siguiente ecuación logística:

$$P(Y = 1) = \frac{1}{1 + e^{-(B_0 + B_1X_1 + B_2X_2 + \cdots + B_iX_i)}}$$

La curva que se genera es similar a la ofrecida en la siguiente Figura 3.4

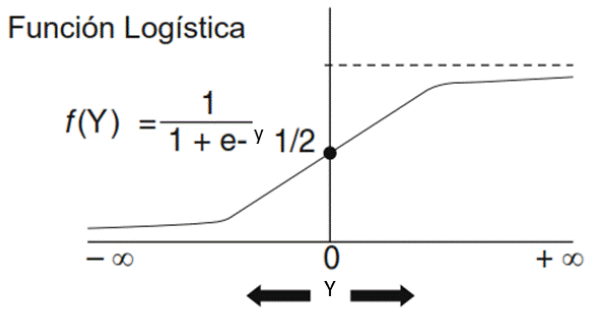


Figura 3.4. Función Logística. Elaboración propia a partir de Kleinbaum, Klein, & Pryor (2010).

Los objetivos según Jovell (1995) del modelo de regresión logística, al estudiar la relación entre una variable dependiente dicotómica "Y" y una o más variables independientes "*Xi*", son:

1. Determinar la existencia o ausencia de relación entre una o más variables independientes "*Xi*" y la variable dependiente "*Y*"
2. Medir la magnitud de dicha relación.
3. Estimar o predecir la probabilidad de que se produzca un suceso o acontecimiento definido como "*Y=1*" en función de los valores que adoptan las variables independientes (*Xi*).

El tipo de diseño del estudio condiciona la interpretación de los resultados del modelo de regresión logística. En aquellos estudios en que los valores definidos

por las variables independientes preceden en el tiempo al suceso o acontecimiento señalado por la variable dependiente, la relación entre ambos tipos de variables se explica en términos de predicción o determinación de la variable dependiente por una o más variables independientes. En cambio, en los estudios en que las características definidas por ambos tipos de variables -independiente y dependientes- se miden en el mismo momento en el tiempo, su relación se interpreta en términos de correlación o asociación (Jovell 1995, p.15).

3.3.2.4 Fases al aplicar la técnica de regresión logística

Según Pardo Merino (2002) un análisis de regresión logística consta de cuatro fases fundamentales:

1. La selección de las variables del análisis: La selección de variables puede realizarse a partir de criterios teóricos (en cuyo caso se utilizará una estrategia de inclusión forzosa de variables) o puede obedecer a criterios estadísticos (en cuyo caso se utilizará algún método de selección por pasos).
2. La estimación de los pesos o coeficientes de las variables seleccionadas: La estimación de los pesos o coeficientes asociados a cada variable se realiza mediante un algoritmo iterativo de máxima verosimilitud.
3. La clasificación de los casos: La clasificación de los casos se realiza a partir del pronóstico del modelo estimado.
4. El análisis de los residuos: El análisis de los residuos permite detectar posibles casos atípicos o predicciones anómalas.

Para poder utilizar una variable dicotómica como variable dependiente en un análisis de regresión logística binaria, es necesario filtrar previamente las dos categorías con las que se desea trabajar, o alternativamente, recodificar la variable original haciéndole tomar sólo dos valores, cuando esto tenga sentido. Una ecuación de regresión logística raramente genera una curva en forma de escalón. Lo habitual es encontrar que las probabilidades pronosticadas adoptan valores comprendidos entre 0 y 1. Este es el motivo que obliga a tener que establecer un punto de corte para poder tomar la decisión de clasificar a los sujetos en uno u otro grupo a partir de las probabilidades pronosticas, para ello, se

opta ir variando el punto hasta optimizar el porcentaje de casos correctamente clasificados[10] (Pardo Merino & Ruíz Díaz 2002).

En este capítulo se ha realizado la introducción y presentación de las técnicas de análisis seleccionadas, en el siguiente capítulo se realiza la aplicación de las técnicas y en el análisis de las actividades en I+D+I de las empresas de alta y media alta tecnología.

[10] Un método alternativo consiste en utilizar la *Curva COR* (Pardo Merino & Ruíz Díaz 2002, p.663)

4 CAPÍTULO 4. ANÁLISIS DE LOS ESTUDIOS REALIZADOS

Con el propósito de cumplir los objetivos específicos de la tesis de explorar las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología, de identificar y describir los patrones y rasgos presentes en las actividades en I+D+I de estas empresas y obtener un modelo para clasificarlas a partir de los rasgos característicos de la alta y media alta-tecnología, se contempla la aplicación de diferentes técnicas de análisis:

1. ***Análisis exploratorio descriptivo***, que tiene el propósito de *describir y explorar* las *diferencias existentes entre los sectores* de Alta y Media-Alta tecnología (AMAT) y otros sectores industriales de manufactura (OSMA). El análisis exploratorio se centra en aquellas variables agrupadas en categorías previamente establecidas en relación a las actividades en I+D+I. Las variables seleccionadas se analizan a la luz de la realización o no de las actividades de interés, para ello se analizan tales actividades a partir de la proporción de empresas que realizan o no cada una de ellas.

2. ***Análisis de minería de datos***, el cual se desarrolla con el propósito de *detectar patrones* en las actividades de I+D+I e identificar rasgos característicos de las empresas de alta y media-alta tecnología. Para la aplicación de las técnicas de minería de datos se recurre al uso del software *open source Weka*, ampliamente utilizado en tareas de minería de datos y que tiene implementado un gran número de algoritmos para la aplicación de las diferentes técnicas de minería de datos.

3. ***Análisis factorial exploratorio (AFE)***, tiene como objetivo *identificar la existencia de factores latentes* presentes en las actividades en I+D+I de las empresas de alta y media-alta tecnología. Para la aplicación de la técnica se recurre al software estadístico SPSS, de uso comercial y ampliamente utilizado en el tratamiento y análisis de datos multivariante.

4. ***Análisis de regresión logística***, se utilizará para *establecer criterios complementarios de Clasificación* Nacional de Actividades Económicas, basados en el desempeño de las empresas pertenecientes a cada subgrupo de clasificación industrial. Para ello se hará uso de un *modelo de regresión logística binaria multivariada*, partiendo de un grupo amplio de variables hasta obtener un conjunto relevante y estadísticamente significativo de variables. Para ello se considera la probabilidad de pertenecer a un subgrupo u otro de la clasificación industrial (AMAT u OSMA). De igual manera, para la aplicación de la técnica se recurre al software estadístico SPSS de uso comercial y que es ampliamente utilizado en el tratamiento y análisis de datos multivariante.

Es de resaltar, y se hace especial énfasis, en que no todas las técnicas permiten el análisis multivariado a partir de variables de índole numérica y categórica simultáneamente. En cada uno los casos o etapas se seleccionan las variables a utilizar de acuerdo a la naturaleza de la técnica de análisis de datos, y considerando los factores teóricos de agrupación de las variables.

4.1 ANÁLISIS EXPLORATORIO DESCRIPTIVO: PERFILES DE ACTIVIDADES EN I+D+I

Utilizando un conjunto de datos (44 variables categóricas sobre 1540 empresas) se pretende dar respuesta al objetivo que se marca en la tesis de identificar las principales características de las actividades centradas en la I+D e Innovación (I+D+I) dentro del sector industrial de alta y media-alta tecnología. Se trata igualmente de buscar la identificación de las principales diferencias con respecto a otros sectores no tecnológicos. Para ello se parte de un conjunto de datos asociados a 1540 empresas de manufactura en España que están organizadas en dos subgrupos a partir de la Clasificación Nacional de Actividades Económicas–CNAE 2009. Las 44 variables analizadas han sido organizadas en 5 cinco bloques de actividades de I+D e innovación. En esta primera etapa de la investigación se realiza un análisis exploratorio-descriptivo con el propósito de aportar datos, información y evidencias de las características destacadas de los

sectores de mayor componente tecnológica frente a otros sectores industriales de manufactura en España.

4.1.1 Caracterización de los sectores manufactureros de alta y media-alta tecnología frente a otros sectores industriales

Un primer análisis realizado se centra en descubrir los sectores con mayor presencia de empresas industriales sobre la base de los datos de la encuesta de la Fundación SEPI. En concreto, como se muestra en la Figura 4.1 la muestra de empresas está formada por un total de 1540 empresas, de las que 439 pertenecen a segmentos de alta y media-alta tecnología (un 28,50 % del total de empresas de la muestra), frente a 1101 empresas de otros sectores industriales de manufactura. Dentro de las empresas de segmentos tecnológicos, están representadas en la muestra la industria química y farmacéutica (137 empresas), de productos informáticos y electrónicos (31 empresas), maquinaria y material eléctrico (74 empresas), máquinas agrícolas e industriales (110 empresas) y vehículos de motor (87 empresas).

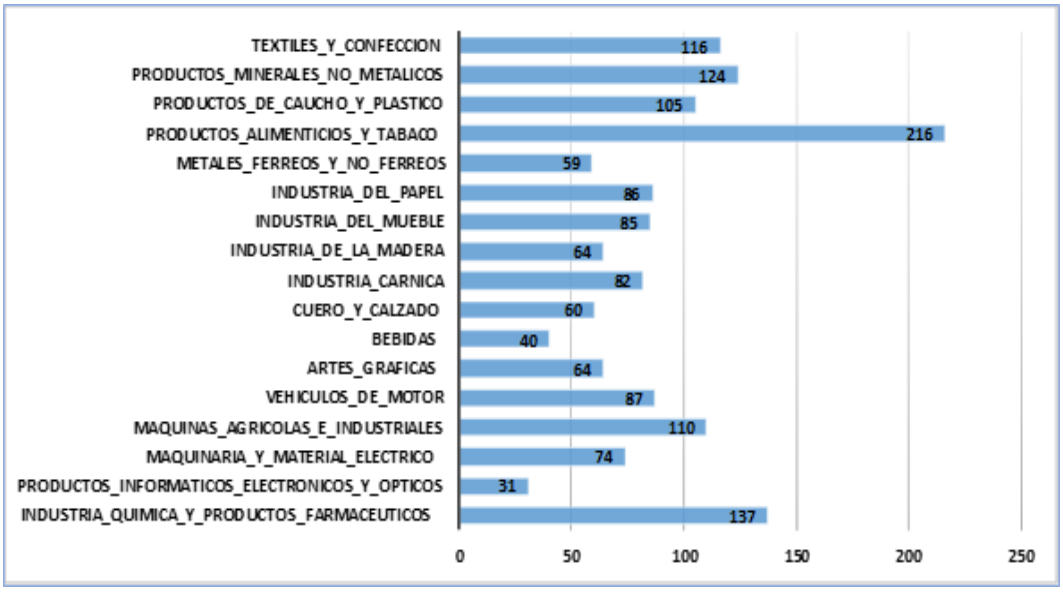


Figura 4.1. Sectores industriales presentes en la muestra. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En la muestra aparecen también empresas de diferentes tamaños. Si se atiende a una clasificación en función del tamaño, distinguiendo entre empresas de 200 y menos trabajadores, o empresas de más de 200 trabajadores, el grupo de empresas de menor tamaño es mayoritario en los dos sectores industriales. En las empresas de alta y media-alta tecnología las empresas de 200 o menos

trabajadores representan un 69,25% de la muestra, frente al 84,74% en las restantes empresas. Visto de otra manera, el grupo de empresas con tamaño de más de 200 trabajadores alcanza la mayor proporción, con el 30,75%, en el sector de alta tecnología, en comparación con el 15,26% en el segmento de otros sectores de manufactura.

4.1.2 Análisis descriptivo en torno a variables categóricas

Como se ha comentado previamente, para la identificación de características relevantes que permitan describir las actividades de I+D e innovación en las empresas de alta y media-alta tecnología se considera un conjunto de 44 variables* de tipo nominal dicotómicas (existencia o no). El nombre original de cada variable va acompañado del símbolo "*" para señalar que el estado de la variable que se considera para los respectivos cálculos es el de actividad realizada. Las variables fueron agrupadas, en función al tipo de actividad que miden, en los siguientes 5 bloques o grupos:

1) Actividades de comercialización (7 variables).
2) Actividades tecnológicas (16 variables).
3) Actividades de incentivos y financiación en I+D e innovación (2 variables).
4) Actividades de innovación (16 variables).
5) Tipo de personal contratado dedicado a actividades de I+D (3 variables).

En las próximas secciones se indicarán las variables correspondientes a cada grupo de actividades al mismo tiempo que se presentarán los resultados del análisis de sus valores en las empresas de la muestra. Se prestará especial atención a las diferencias más prominentes cuando se comparan los resultados sobre el subconjunto de las 439 empresas de alta y media-alta tecnología frente a las 1101 empresas manufactureras de otros sectores.

4.1.2.1 Actividades de comercialización

Dentro de las variables relacionadas con actividades de comercialización se incluyen 7 que vienen a medir si se comercializan productos no fabricados por la empresa, si se realizan acciones de importación y exportación, ventas por internet y la incidencia de estas últimas. La Tabla 4.1 muestra las variables vinculadas a las actividades de comercialización y se incluye para cada una su identificador de variable junto a su nombre, una breve descripción, y los

porcentajes de empresas dentro de los dos subconjuntos de la muestra (alta y media-alta tecnología, por un lado, y el resto por otro) que realizan la correspondiente actividad. La última columna de la tabla muestra el ratio entre el número de empresas de alta y media-alta tecnología y las empresas de otros sectores de manufactura que realizan la actividad.

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Comercialización de productos nacionales no fabricados por la empresa (COMPN1*)	% de empresas que reportan comercialización en España de productos no fabricados por ellas	27,79	35,60	0,78
Comercialización de productos extranjeros no fabricados por la empresa (COMPE*)	% de empresas que comercializan productos extranjeros	35,31	18,17	1,94
Importación (IMPORT*)	% de empresas que importan	89,98	65,67	1,37
Exportación (EXPORT*)	% de empresas que exportan	89,98	61,31	1,47
Ventas a empresas por Internet (WEBB2B*)	% de empresas que cuentan con un sistema de ventas a otras empresas por Internet	10,71	9,45	1,13
Ventas a consumidores finales por Internet (WEBB2C*)	% de empresas que disponen de un sistema de ventas a consumidores finales por Internet	6,15	10,17	0,60
Incidencia de Internet sobre las ventas (WEBVEN*)	% Incidencia (fuerte o ligera) que la presencia de Internet ha tenido sobre las ventas de la empresa	37,81	36,97	1,02

Tabla 4.1. Actividades de comercialización. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

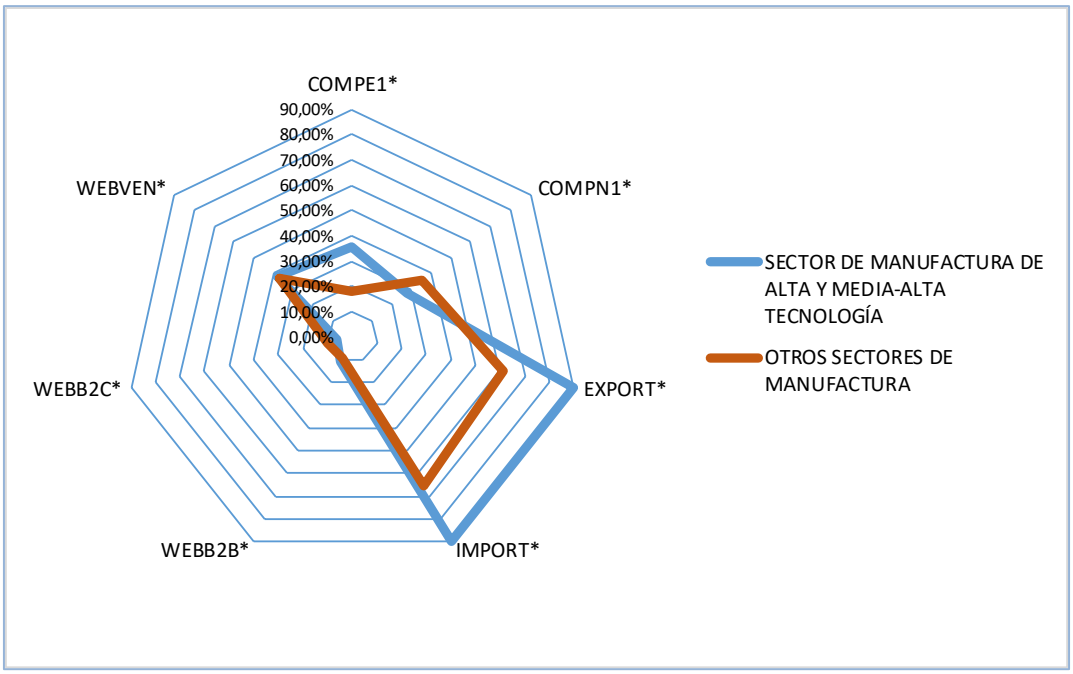


Figura 4.2. Patrón en las actividades de comercialización. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Como se observa tanto en la Tabla 4.1 como en la Figura 4.2 en el sector industrial de manufactura de alta y media-alta tecnología, el porcentaje de empresas que realizan la actividad de comercialización en el extranjero de productos no fabricados por ellas es mayor que el porcentaje que reportan comercialización en el propio país de dichos productos. Caso opuesto en lo referente a otros sectores, donde el porcentaje de empresas que reportan comercialización de productos no fabricados por ellas y comercializados en el propio país es mayor que la comercialización de tales productos en el extranjero.

Es de resaltar también que en las empresas de alta y media-alta tecnología el 89,98 % realizan operaciones de importación, y el mismo porcentaje del 89,98% realizan operaciones de exportación. De igual manera, todas estas actividades son de mayor proporción en el segmento de alta y media-alta en relación a otros sectores de manufactura en proporciones de 1,37:1 (en importación), 1,47:1 (en exportación), respectivamente.

En cuanto a la actividad en comercio electrónico, el porcentaje de empresas que realizan actividades de ventas por Internet a otras empresas (modelos B2B) está en el 9,45% para las empresas pertenecientes al segmento de

otros sectores industriales de manufactura y el 10,71% para las empresas del segmento de alta y media-alta tecnología, valores muy parecidos. Sin embargo, en las ventas de comercio electrónico a consumidores finales (modelos B2C) las empresas de alta tecnología desarrollan una menor actividad. Por otra parte, al analizar la variable en la que las empresas declaran la incidencia que el comercio electrónico tiene en las ventas de la empresa; el 36,97% de las empresas manufactureras de otros sectores declaran una incidencia entre ligera y fuerte, mientras que en el sector de mayor tecnología ese porcentaje es muy similar, un 37,81%, en una proporción 1: 1,02 con una ligera diferencia a favor de las empresas de alta y media-alta tecnología.

Se observa, especialmente en la Figura 4.2, como los perfiles de actividades de las empresas pertenecientes al segmento industrial de alta y media-alta tecnología y la de las restantes empresas manufactureras siguen un patrón similar, donde la intensidad de las actividades es mayor en el segmento de empresas de alta y media-alta tecnología. Una leve diferencia respecto a este patrón general se produce al analizar la comercialización de productos nacionales no fabricados por la empresa y las ventas a consumidores finales por internet.

4.1.3 Actividades Innovadoras y de apoyo tecnológico

El segundo de los factores a analizar viene determinado por un total de diecisiete (17) variables que hacen referencia a aspectos como la existencia de acuerdos de cooperación tecnológica, adquisición de bienes de equipo para mejora de productos, actividades de I+D, existencia de direcciones o comités de tecnología, plan de actividades de innovación, subcontratación de productos o componentes, utilización de asesores para informarse sobre tecnología, indicadores de resultados de la innovación, evaluación de perspectivas de cambio tecnológico, colaboración tecnológica con clientes, competidores, proveedores, con universidades, participación en programas de investigación de la U.E., o el registro de patentes en el extranjero y en España. Todas estas variables, junto con los valores medidos en la muestra se incluyen en la Tabla 4.2, y la Figura 4.3 trata de ilustrar las diferencias observadas al analizar dichas variables dentro de los segmentos industriales de alta y media-alta tecnología y otros sectores de manufactura. Los valores mostrados indican los porcentajes de empresas de cada segmento que realizan la correspondiente actividad tecnológica.

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Acuerdos de cooperación tecnológica (ACT*)	% de empresas que poseen acuerdos de cooperación (*joint venture*)	5,92	2,45	2,42
Adquisición de bienes de equipo para mejora de productos (ADBEM*)	% de empresas que adquieren bienes de equipos para mejora de productos	22,10	14,71	1,50
Actividades de I+D (AID*)	% de empresas con actividades de I+D contratadas y/o realizadas	61,50	27,70	2,22
Dirección o comité de tecnología (DCT*)	% de empresas con dirección o comité de tecnología	45,10	15,53	2,90
Plan de actividades de innovación (PAI*)	% de empresas con planes de actividades de innovación	41,46	17,71	2,34
Subcontratación productos o componentes (SPR*)	% de empresas que subcontratan productos o componentes	44,87	31,15	1,44
Utilización de asesores para informarse sobre tecnología (UAIT*)	% de utilización de asesores para informarse sobre tecnología	27,79	17,08	1,63
Indicadores de resultados de la innovación (IRI*)	% de empresas que elaboran indicadores de resultados de innovación	31,66	14,08	2,25
Evaluación perspectivas cambio tecnológico (EPCT*)	% de empresas que evalúan perspectivas de cambio tecnológico	37,59	19,16	1,96
Colaboración tecnológica con clientes (CTCL*)	% de empresas que colaboran tecnológicamente con clientes	33,49	9,81	3,41
Colaboración tecnológica con competidores (CTCO*)	% de empresas con colaboración tecnológica con competidores	4,56	1,73	2,64
Colaboración tecnológica con proveedores (CTPR*)	% de empresas con colaboración tecnológica con proveedores	37,13	15,71	2,36
Colaboración con universidades y/o centros tecnológicos (CUCT*)	% de empresas con colaboración tecnológica con universidades y/o centros tecnológicos	40,32	18,71	2,15
Participación en programas de	% de empresas que participan en	1,82	0,91	2,01

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
investigación de UE (PIUE*)	programas de la Unión Europea			
Patentes registradas en el extranjero (PATEXT*)	% de empresas que durante el ejercicio registran patentes en el extranjero	6,83	1,63	4,18
Patentes registradas en España (PATESP*)	% de empresas que durante el ejercicio registran patentes en España	7,74	2,63	2,94

Tabla 4.2. Actividades Innovadoras y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

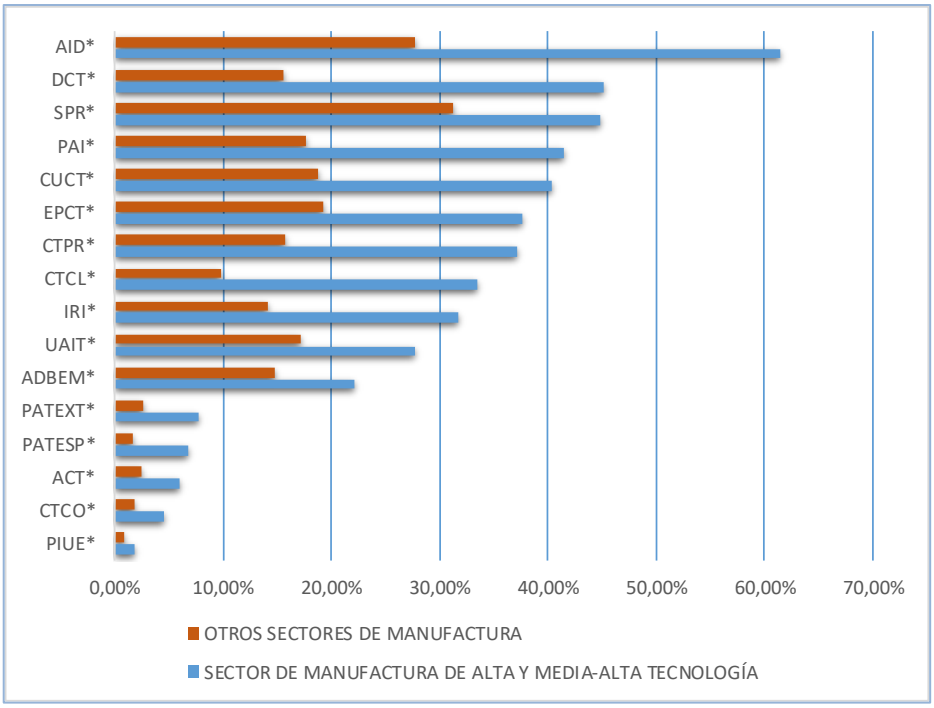


Figura 4.3. Actividades Innovadoras y de apoyo tecnológico. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

- En el primer subgrupo se encuentran aquellas actividades que se realizan por parte de más del 20% de las empresas del segmento de alta y media alta tecnología. En este grupo se sitúan actividades como: la contratación y/o realización de I+D, la existencia de direcciones o comités de tecnología, la subcontratación de productos o componentes, la elaboración de planes de actividades de innovación, la colaboración tecnológica con universidades y/o centros tecnológicos, y la evaluación de perspectivas de

cambio tecnológico, la colaboración tecnológica con proveedores, y la colaboración tecnológica con clientes, indicadores de resultados de innovación, la utilización de asesores para informarse sobre tecnología y la adquisición de bienes de equipos para mejora de productos.

- Finalmente, aparecen un grupo de actividades que están presentes en pequeños porcentajes de empresas (entre el 1,82 y el 7,74%): registro de patentes en el extranjero, en España, acuerdos de cooperación tecnológica (*joint venture*), acuerdos de colaboración tecnológica con competidores, o participación en programas de investigación de la Unión Europea.

En el caso del subgrupo de otros sectores industriales de manufactura se encuentran actividades que se realizan entre un 14,71% y un 31,15% de las empresas, como son: actividades de I+D contratadas y/o realizadas, dirección o comité de tecnología, subcontratación productos o componentes, planes de actividades de innovación, colaboración tecnológica con universidades y/o centros tecnológicos, evaluación de perspectivas de cambio tecnológico, colaboración tecnológica con clientes, realización de indicadores de resultados de innovación, utilización de asesores para informarse sobre tecnología y adquisición de bienes de equipo para mejora de productos.

En menor grado se observan en éste mismo sector, un porcentaje menor empresas que realizan (0,91% y un 9,81%) las siguientes actividades: registro de patentes en el extranjero y en España, acuerdos de cooperación tecnológica (*joint venture*), colaboración tecnológica con clientes, con competidores y participación en programas de la Unión Europea.

Es de resaltar que, en todos los casos, el porcentaje de empresas que realizan las actividades es mayor para el segmento de alta y media-alta tecnología. Precisamente cuando se comparan los resultados obtenidos en el segmento de empresas de alta y media-alta tecnología con el resto de empresas, se observa un patrón similar en relación a las actividades realizadas (ver Figura 4.4). La diferencia más notable se encuentra en la contratación y/o realización de I+D donde por una de cada cien empresas que realizan esta actividad en otros sectores de manufactura, existen 2.22 empresas por cada cien en el sector de alta

y media-alta tecnología. La menor diferencia se observa en la participación en programas de investigación de la Unión Europea donde por una de cada cien empresas que realizan la actividad en otros sectores, existe 1.44 por cada cien en el sector de alta y media-alta tecnología.

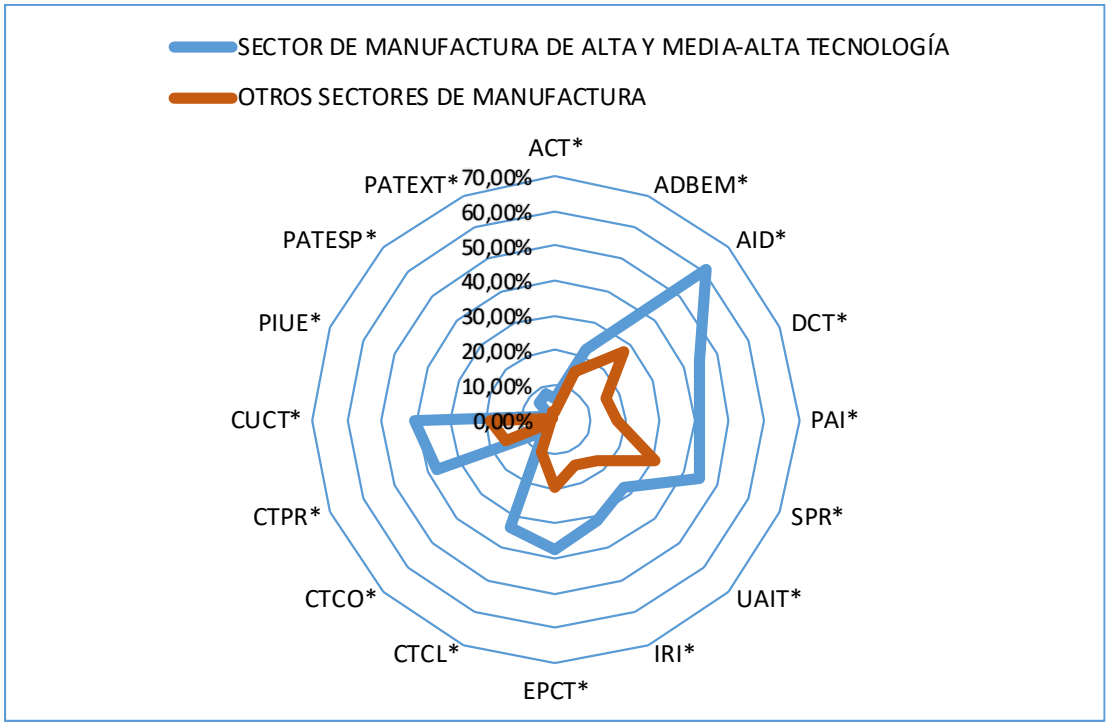


Figura 4.4. Patrón en las actividades tecnológicas. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

4.1.4 Actividades de incentivos y financiación en I+D e Innovación.

Dentro del factor de incentivos y financiación en I+D e innovación se han considerado únicamente dos variables, por un lado, identificar si las empresas aplican incentivos fiscales para I+D e innovación tecnológica y, por otro, si optan a la financiación de la innovación con créditos subvencionados. La Tabla 4.3 muestra los porcentajes de empresas en los dos segmentos industriales que realizan estas dos actividades. Como se observa, ambas actividades tienen una mayor presencia en las empresas del segmento industrial de alta y media-alta tecnología. Por cada una de cien empresas que se acogen a incentivos fiscales en otros sectores de manufactura, 2.83 empresas por cada cien lo hacen en el sector de alta y media-alta tecnología. De igual manera sucede en relación a la

financiación de innovación con créditos subvencionados, donde la relación es 1 a 2,76 empresas.

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Aplicar incentivos fiscales I+D (APLIFN)	% De empresas que aplican incentivos fiscales en I+D	32,35	11,44	2,83
Financiación de la innovación con créditos subvencionados (FICS)	% De empresas que financian la innovación con créditos subvencionados	20,05	7,27	2,76

Tabla 4.3. Actividades de incentivos y financiación en I+D e Innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

4.1.5 Actividades de Innovación

Las actividades de innovación podrían agruparse en cuatro categorías: innovaciones en comercialización, en métodos organizativos, en producto y en proceso, en total se han considerado las 16 variables. Dentro de la primera categoría se encuentran innovaciones en comercialización por diseño, por promoción, en canales de venta o en fijación de precios. Las innovaciones en métodos organizativos incluyen también las innovaciones en organización del trabajo o en las relaciones externas. A la hora de analizar innovaciones en productos se diferencian las innovaciones en nuevos componentes, nuevas funciones o nuevos materiales. Finalmente, las innovaciones de procesos se pueden originar por el uso de nuevos equipos, programas informáticos, o nuevas técnicas. En la Tabla 4.4 se incluyen los porcentajes de empresas en los dos segmentos considerados que realizan las diferentes actividades de innovación. Al igual que en los casos anteriores, se observa una mayor intensidad de las acciones de innovación en el segmento de empresas de alta y media-alta tecnología, excepto en el porcentaje de empresas que realizan innovaciones en canales de venta, donde es ligeramente mayor en el segmento de otros sectores de manufactura.

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Innovaciones de comercialización (ICO*)	% de empresas que realizan innovaciones de comercialización	22,55	19,53	1,15
Innovaciones en canales de venta (ICOCAN*)	% de empresas que realizan innovaciones en canales de venta	7,29	8,08	0,90

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Innovaciones de comercialización por diseño (ICODIS*)	% de empresas que realizan innovaciones en comercialización por diseño	12,76	11,35	1,12
Innovaciones en la fijación de los precios (ICOFIJ*)	% de empresas que realizan innovaciones en la fijación de precios	7,52	5,00	1,50
Innovaciones de comercialización por promoción (ICOPRO*)	% de empresas que realizan innovaciones en comercialización por promoción	11,16	9,81	1,14
Innovaciones de métodos organizativos (IMO*)	% de empresas que realizan innovaciones de métodos organizativos	31,21	18,98	1,64
Innovaciones en la gestión de las relaciones externas (IMOGE*)	% de empresas que realizan innovaciones en la gestión de las relaciones externas	14,35	9,54	1,50
Innovaciones en la organización del trabajo (IMOPE*)	% de empresas que realizan innovaciones en la organización del trabajo	28,93	17,98	1,61
Obtención innovaciones de producto (IP*)	% de empresas con innovación en producto	30,75	14,08	2,18
Innovación producto por nuevos componentes (IPNC*)	% de innovación de producto por nuevos componentes	22,78	8,27	2,76
Innovación producto por nuevas funciones (IPNF*)	% de innovación de producto por nuevas funciones	23,01	9,08	2,53
Innovación producto por nuevos materiales (IPNM*)	% de innovación de producto por nuevos materiales	20,05	9,99	2,01
Obtención innovaciones de proceso (IPR*)	% de innovación de proceso	43,28	28,52	1,52
Innovaciones de proceso por nuevos equipos (IPRME*)	% de innovación de proceso por nuevos equipos	33,03	21,71	1,52
Innovaciones de proceso por programas informáticos(IPRPI*)	% de innovaciones de proceso por programas informáticos	20,96	10,90	1,92
Innovaciones de proceso por nuevas técnicas (IPRTM*)	% de innovaciones de proceso por nuevas técnicas	28,25	14,17	1,99

Tabla 4.4. Actividades de innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

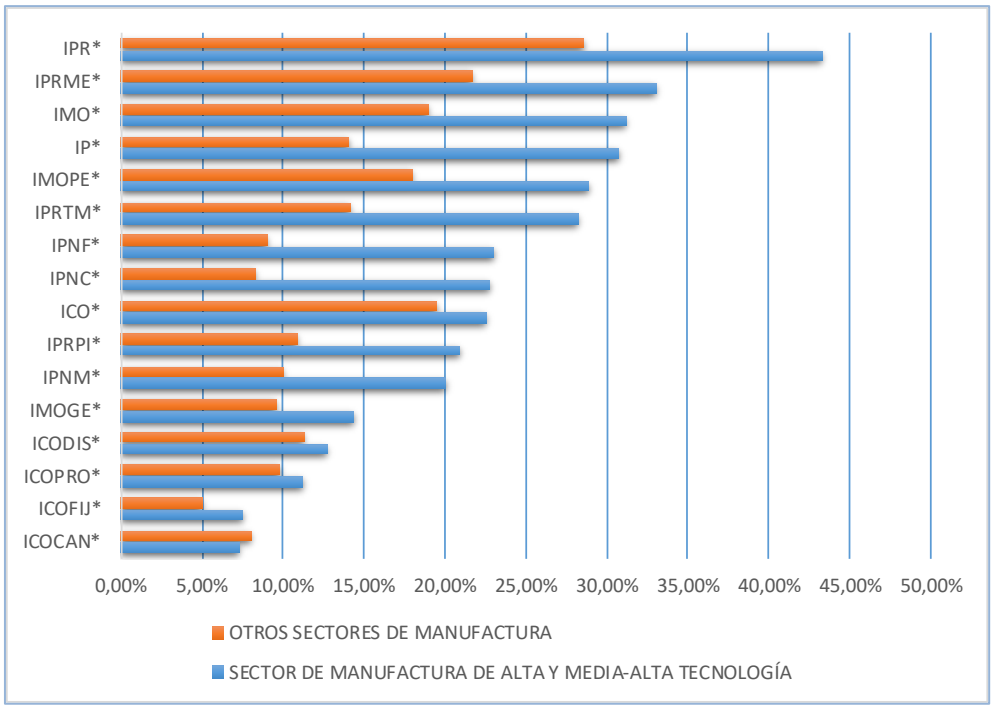


Figura 4.5. Actividades de Innovación. Comparación porcentual por tipo de actividad y de sector industrial de manufactura. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En la Figura 4.5 se observa que aquellas actividades de innovación con más presencia en las empresas de alta y media-alta tecnología, con porcentajes de mayor a menor valor entre el 43,28 al 20,05%, son: innovaciones de proceso, innovaciones de proceso por nuevos equipos, innovación de métodos organizativos, innovaciones de producto, innovaciones en la organización del trabajo, innovaciones de proceso por nuevas técnicas, innovación de producto por nuevas funciones, innovación de producto por nuevos componentes, innovaciones de comercialización, innovaciones de proceso por programas informáticos e innovación de producto por nuevos materiales.

En un segundo nivel, con porcentajes entre el 7,29 y el 14,35 %, de las empresas, se encuentran, en orden descendente, las innovaciones en gestión de las relaciones externas, innovaciones de comercialización por diseño, innovaciones de comercialización por promoción, innovación en la fijación de precio e innovaciones en canales de venta.

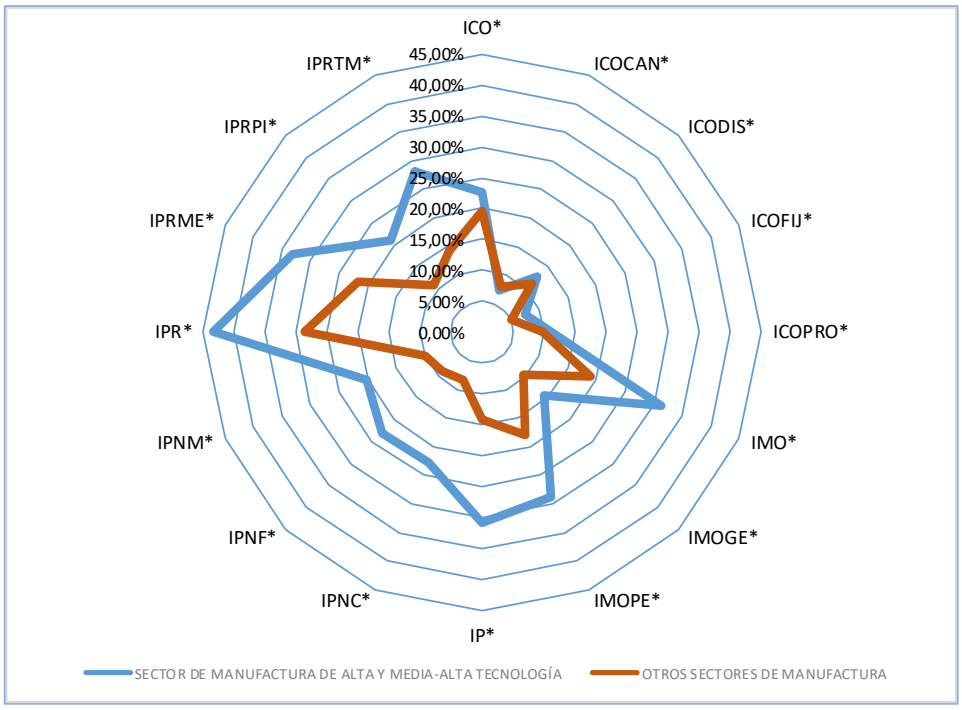


Figura 4.6. Patrón en las actividades de innovación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

La Figura 4.6 muestra de una manera global cómo las empresas de alta y media-alta tecnología y otros sectores de manufactura, siguen un patrón similar en relación a las actividades realizadas. Sin embargo, en términos generales las actividades de innovación son realizadas en todos los casos en mayor proporción de empresas en sector más tecnológico. En mayor media, por una de cada cien empresas de otros sectores industriales que realicen innovaciones de proceso existen en promedio 1,52 empresas por cada cien en el segmento industrial de alta y media-alta tecnología. En menor media, por una de cada cien empresas de otros sectores de manufactura que realizan actividades innovaciones en canales de venta existe en promedio 0,90 empresas por cada cien que realizan tales actividades en el segmento más tecnológico.

4.1.6 Recursos Humanos

Finalmente, el último factor que se consideró importante analizar fue el asociado con los recursos humanos que las empresas dedican a actividades de I+D. En este caso se consideraron únicamente 3 variables: si las empresas reclutaron en el año analizado personal con experiencia empresarial en I+D, personal con experiencia profesional en el sistema público de I+D, e ingenieros y/o licenciados de graduación reciente. La Tabla 4.5 muestra los porcentajes de las

empresas de los dos segmentos; esa misma información se muestra de manera gráfica en la Figura 4.7.

Variable	Descripción	Alta y media-alta tecnología	Otros sectores	Ratio
Reclutar personas con experiencia empresarial en I+D (REEID*)	% de empresas que reclutan personas con experiencia empresarial en I+D	10,71	3,09	3,47
Reclutar personas con experiencia profesional en I+D en el sector público (REPID*)	% de empresas que reclutan personas con experiencia en I+D en el sector público	4,10	0,82	5,02
Incorporación de ingenieros y/o licenciados recientes (IILR*)	% de empresas que incorporan ingenieros y/o licenciados recientes	32,57	10,17	3,20

Tabla 4.5. Recursos Humanos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

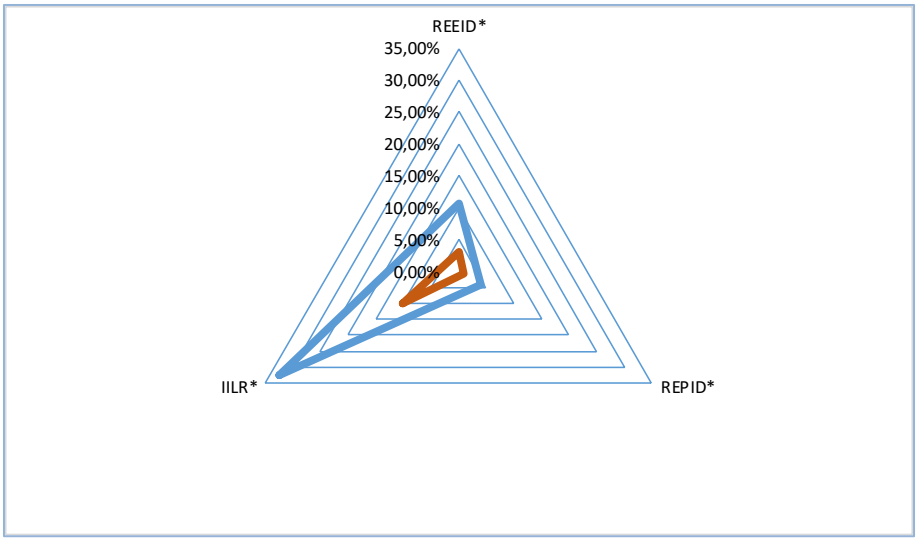


Figura 4.7. Personal contratado dedicado a actividades de I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Tanto en la Tabla 4.5 como en la correspondiente Figura 4.7 se observa de nuevo el mismo patrón en cuanto a la contratación de personal de I+D en los dos segmentos empresariales, pero con proporciones sensiblemente mayores en las empresas de alta y media-alta tecnología. El esfuerzo en la contratación de

personal con experiencia empresarial en I+D es más del triple en estas últimas empresas. Especialmente llamativa es la diferencia que se observa en la proporción de empresas con contratación de personal con experiencia en I+D en el sector público, siendo inferior en los dos segmentos, con un 4,1 % de las empresas en el sector de alta y media-alta tecnología, y aún más baja, en el segmento de otros sectores de manufactura con tan solo 0, 82 % de las empresas. También se destaca que una tercera parte de las empresas del sector tecnológico contrataron titulados superiores recién graduados.

4.1.7 Conclusiones del análisis exploratorio descriptivo

En síntesis, el presente estudio exploratorio descriptivo ha servido para recolectar evidencias preliminares que permiten identificar rasgos característicos de las empresas manufactureras de alta y media-alta tecnología, y que diferencien a estas empresas de las pertenecientes a otros sectores industriales de manufactura en España. Debe enfatizarse que el estudio se ha realizado en forma de consolidado para cada uno de los dos segmentos establecidos, sin particularizarse por tipo de subsector industrial en el interior de cada uno de ellos. A continuación, se destacan algunos de los aspectos más importantes observados en el estudio.

Las pequeñas y medianas empresas, con 200 o menos trabajadores, son predominantes en todos los sectores industriales, sin embargo, el predominio es menor en las empresas del segmento de alta y media-alta tecnología. En concreto, un 69,25% de las empresas de este segmento responden a ese perfil de empresas de tamaño pequeño o mediano, frente a un porcentaje del 84,74% en otros sectores.

Se observa que el perfil de actividades de las empresas pertenecientes al segmento industrial de alta y media-alta tecnología y a otros sectores industriales de manufactura sigue un patrón similar; no obstante, las proporciones de empresas que realizan las actividades es sensiblemente mayor en el segmento de empresas de alta y media-alta tecnología. Como excepciones, las empresas de otros sectores de manufactura reportan mayor proporción de empresas que realizan comercialización en España de productos no fabricadas por ellas, venta a consumidores finales por internet. Asimismo, la proporción de empresas que

realizan innovaciones en canales de venta es mayor en los sectores menos tecnológicos.

Se ha podido identificar un conjunto de actividades tecnológicas que son bastante comunes en las empresas de alta y media-alta tecnología, estando presentes en más del 40% de las empresas de la muestra. Entre estas acciones están: Importación (89,98 %), exportación (89,98%), contratación y realización de actividades de I+D (61,50 %).

Un segundo grupo está formado por actividades que son realizadas aproximadamente por una de cada dos empresas; en concreto: mantienen una dirección o comité de tecnología (45,10%), disponen de planes de actividades de innovación (41,46%), subcontratan productos o componentes (44,87%), realizan innovación de proceso (43,28%), colaboran tecnológicamente con universidades y/o centros tecnológicos (40, 32%).

En un tercer grupo, donde aproximadamente 1 de cada 3 o 4 empresas lo realizan, se encuentran las siguientes actividades: evalúan perspectivas de cambio tecnológico (37,59%), colaboran con proveedores (37,13%), colaboran tecnológicamente con clientes (33, 49%), reportan innovación de proceso por nuevos equipos (33,03%), incorporan ingenieros o licenciados recientes (32,57%), aplican incentivos fiscales en I+D (32,35%), utilizan asesores para informarse sobre tecnología (27,79%), elaboran indicadores de resultados de innovación (31,66 %), presentan innovaciones en métodos organizativos (31, 21%), realizan innovaciones en la organización del trabajo (28,93), introducen innovación en producto (30, 75%), y/o realizan innovación de proceso por nuevas técnicas (28,25%).

Se observa en menor grado, un conjunto de actividades donde aproximadamente 1 de cada 5 las realizan: financian la innovación con créditos subvencionados (20,05%), realizan innovación de comercialización (22,55%), innovación de producto por nuevos componentes (22,78%), innovación por producto por nuevas funciones (23,01%), innovación de producto por nuevos materiales (20,05%), innovación de proceso por programas informáticos (20,96 %).

Puede ser importante analizar aquellas acciones en las que la presencia porcentual en las empresas tecnológicas es notablemente mayor que en las de otros sectores. En general presentan las ratios con los valores más diferenciados a

favor de aquellas empresas con mayor componente tecnológico, son, por un lado, el reclutamiento de personal con experiencia profesional en el sector público (5,02), donde a pesar de que en ambos grupos el porcentaje de empresas que la realiza es inferior al 5%, se verifica que por cada una de cien empresas de otros sectores, existen 5 por cada cien empresas en el sector de alta y media-alta tecnología. Un segundo aspecto con diferencias importantes es el registro de patentes el extranjero, donde la proporción es 1 a 4,18 empresas por cada cien.

Finalmente, esta parte del estudio contribuye al objetivo de identificar unos primeros rasgos que caractericen a las empresas de sectores industriales manufactureros vinculados con alta y media-alta tecnología, sentando las bases para la identificación posterior de patrones que puedan estar más ocultos en los datos. A este respecto, se plantea la utilización de técnicas de minería de datos para explorar los datos de la encuesta analizada y descubrir esas relaciones ocultas en los datos.

4.2 CARACTERIZACIÓN DE LAS EMPRESAS DE ALTA TECNOLOGÍA A PARTIR DE RASGOS Y PATRONES DE ACTIVIDAD DE I+D UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS

Esta sección tiene como objetivo utilizar técnicas de minería de datos que permitirán identificar y explorar aquellos patrones y rasgos presentes en las actividades de I+D+I en las empresas del sector de alta o media-alta tecnología. Se parte de los datos de la encuesta utilizada se cuenta con 53 variables categóricas asociadas a la actividad de I+D+I, y se consideran las 1540 empresas (439 pertenecientes al sector de alta y media-alta tecnología y 1101 a otros sectores de manufactura).

La primera fase de cualquier proceso de descubrimiento de conocimiento con minería de datos comprende el pre-procesamiento de los datos, para posteriormente aplicar las técnicas más adecuadas de acuerdo a los objetivos planteados. En este caso se empezará por la elaboración de un ranking para obtener un conjunto de atributos relevantes sobre las actividades de I+D+I. Posteriormente se aplicarán técnicas de clasificación para la identificación de los patrones característicos en las actividades de I+D+I en las empresas de alta y

media-alta tecnología (AMAT). En concreto, mediante árboles de decisión se facilitará la visualización de los patrones característicos, permitiendo predecir nuevos casos. También se hace uso de técnicas de generación de reglas de asociación que permitirán identificar las correlaciones o interdependencias que existen entre las variables. Finalmente, en un quinto paso se identifican los patrones característicos mediante técnicas de *clustering* que permitirán identificar grupos de empresas con rasgos similares.

4.2.1 Identificación y selección de variables disponibles para la realización de la minería de datos, y tamaño de la muestra.

A partir de la revisión de la literatura sobre las actividades de I+D e Innovación en la empresa se han identificado a priori las variables que pueden ser agrupadas bajo las categorías que se muestran en la Tabla 4.6.

Muchas de las variables requeridas aparecen en el estudio previo, por ello se presenta solo su nombre, se añaden nuevas variables que refuerzan los datos sobre los bloques de actividades innovadoras y de apoyo tecnológico, así como las actividades relacionadas con los incentivos y financiación de la I+D e Innovación. Explícitamente para estas nuevas variables se presenta el nombre y la descripción.

ASPECTO	SELECCIÓN DE VARIABLES
Actividades de Comercialización	COMPE, COMPN1, EXPORT, IMPORT, WEBB2B, WEBB2C, WEBVEN.
Actividades Innovadoras y de apoyo tecnológico	ACT, ADBEM, AID, CTCL, CTCO, CTPR, CUCT, DCT, EPCT, IRI, PAI, PATESP, PATEXT, PIUE, SPR, UAIT. ETAE, EVALUACIÓN DE TECNOL. ALTERNATIVAS. GTID, GASTOS TOTALES I+D. PEIT, PARTICIPACIÓN EN EMPRESAS CON INNOVACIÓN TECNOL.
Incentivos y financiación en I+D e Innovación	APLIFN, FICS. BSFE, BUSCAR SIN ÉXITO FINANCIAC. EXTERNA DE LA INNOVAC. FPIDCA, FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS. FPIDES, FINANCIACIÓN PÚBLICA I+D ESTADO. FPIDOT, FINANCIACIÓN PÚBLICA I+D OTROS.
Actividades de Innovación	ICO, ICOCAN, ICODIS, ICOFIJ, ICOPRO, IMO, IMOGE, IMOPE, IP, IPNC, IPNF, IPNM, IPR, IPRME, IPRPI, IPRTM.
Recursos Humanos	IILR, REEID, REPID.
Variable adicional	TEMPR2, TAMAÑO DE LA EMPRESA (INTERVALOS DE TAMAÑO)

Tabla 4.6. Variables seleccionadas para el análisis de minería de datos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

A las 52 variables de la Tabla 4.6 se añade una última variable categórica (booleana) que indica si la correspondiente empresa pertenece o no a un sector de alta o media-alta tecnología. Un total de 439 empresas de la muestra de 1540, distribuidas tal como muestra la Tabla 4.7 están asociadas a esos sectores y como tal fueron consideradas empresas de alta o media-alta tecnología. Las restantes empresas quedan agrupadas en el segmento de otros sectores industriales de manufactura (OSMA)

Sector	***Nº empresas***
Industria química y productos farmacéuticos (CNAE 21)	137
Productos informáticos, electrónicos y ópticos (CNAE 26)	31
Maquinaria y material eléctrico (CNAE 27)	74
Máquinas agrícolas e industriales (CNAE 28)	110
Vehículos de motor (CENAE 29)	87

Tabla 4.7. Número de empresas de alta o media-alta tecnología en la muestra disponible para la minería de datos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

4.2.2 Preprocesamiento

Con objeto de poder aplicar las técnicas de minería de datos usando el software Weka, se generó un archivo ARFF con los datos de las 1540 empresas y 52 atributos para cada una de ellas.

Para el análisis no se tiene en cuenta la variable que indica el sector de actividad de la empresa (NACECLIO) y sí se considera una nueva y última variable binaria que indica si la empresa es de alta y media-alta tecnología o de otros sectores industriales de manufactura, ello debido a que la variable NACECLIO identifica el sector industrial al cual pertenece la empresa, pero no ofrece información de si esta es o no considerada de alta o media-alta tecnología.

A fin de poder interpretar posteriormente mejor los resultados obtenidos, se realiza un proceso de transformación de los datos sustituyendo los valores numéricos que identificaban a las variables categóricas por expresiones literales descriptivas de cada valor. Los valores numéricos de algunos atributos son discretizados para trabajar con todas las variables de tipo categórico, específicamente las variables FPIDCA, FPIDES, FPIDOT, GTID, PATESP, PATEXT, la cuales se expresan en categorías baja, media, alta; para ello se ha utilizado el filtro de atributo de discretización disponible en Weka.

4.2.3 Aplicación de técnicas de explotación de datos

La primera de las técnicas que se aplica sobre los datos es la de selección de atributos, que permite seleccionar de manera priorizada los subconjuntos de atributos con mayor poder discriminante a la hora de clasificar las instancias a partir de un atributo de clasificación. Estas técnicas se basan en buscar atributos altamente correlacionados con la variable de clasificación y con poca correlación entre ellos (Hall 1999). En este caso, el motivo de utilizar esta técnica es poder determinar si existe un subconjunto con un menor número de atributos y que pueda ser utilizado para discriminar adecuadamente si las empresas son realmente de alta tecnología.

Para utilizar esta técnica se recurrió al uso del software *Weka*, que incorpora una gran variedad de técnicas de selección de atributos que tratan de explorar qué subconjuntos de atributos son los que mejor pueden clasificar la clase de la instancia. Esta selección de atributos tiene dos componentes:

1. Un método de evaluación que determina la calidad del conjunto de atributos para discriminar la clase. Se pueden distinguir dos categorías de métodos de evaluación, en la primera se utiliza directamente un clasificador específico para medir la calidad del subconjunto de atributos a través de la tasa de error del clasificador. Estos métodos necesitan un proceso completo de entrenamiento y evaluación en cada caso de búsqueda, por eso resultan de un elevado coste computacional. La alternativa es la utilización de métodos que no utilizan un clasificador específico, por ejemplo el método *CfsSubsetEval* que se encuentra implementado en *Weka* y que se basa en calcular la correlación de la clase con cada atributo, y eliminar atributos que tienen una correlación muy alta como atributos redundantes. Según este método los subconjuntos preferidos son aquellos altamente correlacionados con el atributo que define las clases y con poca correlación entre ellos.
2. Un método de búsqueda que determina la forma de realizar la búsqueda de conjuntos. La evaluación exhaustiva de todos los posibles subconjuntos se convierte en un problema combinatorio inabordable cuando el número de atributos es elevado. Por tanto, se necesitan

estrategias de búsqueda más eficientes. Una de las estrategias más efectiva, por su rapidez, es el *ForwardSelection*, que se basa en elegir primero el mejor atributo, y realizar un proceso iterativo de ir añadiendo atributos que aporten más información hasta llegar a la situación en la que añadir un nuevo atributo empeora la situación.

En el caso del análisis realizado se utilizó la técnica de evaluación de atributos *CfsSubsetEval*[11] *y RanksSearch*[12] como método de búsqueda, de esta forma se determina un ranking de atributos según su poder discriminante. En concreto, el ranking incluye los 52 atributos (o rasgos) y se obtiene un subconjunto de 9 variables para clasificar si una empresa es o no de alta tecnología, se presentan en su orden:

{CTCL, DCT, IIRL, IMPORT, GTID, REPID, FPIDES, APLIFN, EXPORT}

Estas variables corresponden a siete variables binarias que establecen si la empresa aplica incentivos fiscales para I+D+I (APLIFN), realiza colaboración tecnológica con clientes (CTCL), mantiene dirección o comité de tecnología (DCT), si realiza exportaciones (EXPORT), incorpora ingenieros y/o licenciados de graduación reciente (IILR), si realiza importaciones (IMPORT), recluta personal con experiencia profesional en el sistema público de I+D (REPID), la variable que considera el total de gastos en actividades de I+D (GTID) y finalmente la variable que considera los recursos financieros recibidos de la administración central para I+D (FPIDES).

La segunda de las técnicas de minería de datos que se aplica es la de generación de árboles de decisión. En este caso se trata de establecer de manera automática condiciones sobre los atributos que permitan realmente clasificar a las empresas. El algoritmo que se utiliza es el J48[13] también implementado en *Weka*.

---

[11] *CfsSubsetEval: para* mayor información Weka recomienda consultar a M. A. Hall (1999). Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand.

[12] *RanksSearch*: igualmente Weka recomienda para mayor información consultar a Mark Hall, Geoffrey Holmes (2003). Benchmarking attribute selection techniques for discrete class data mining. IEEE Transactions on Knowledge and Data Engineering. 15(6):1437-1447.

[13] Para mayor información sobre algoritmos de árbol de decisión C4.5 se puede consultar Salzberg (1994).

En este caso se opta por generar dos árboles, en uno de ellos se consideran los 9 atributos (o rasgos) más relevantes seleccionados por el ranking y se agrega la variable de clasificación. De manera complementaria se genera un segundo árbol con características similares y con el propósito de mejorar su interpretación.

En el primer caso, se obtiene un árbol (Figura 4.8) con la intervención de los atributos más relevantes obtenidos en el ranking: CTCL, DCT, FPIDES, GTID, REPID, IMPORT, EXPORT, IILR, APLIFN.

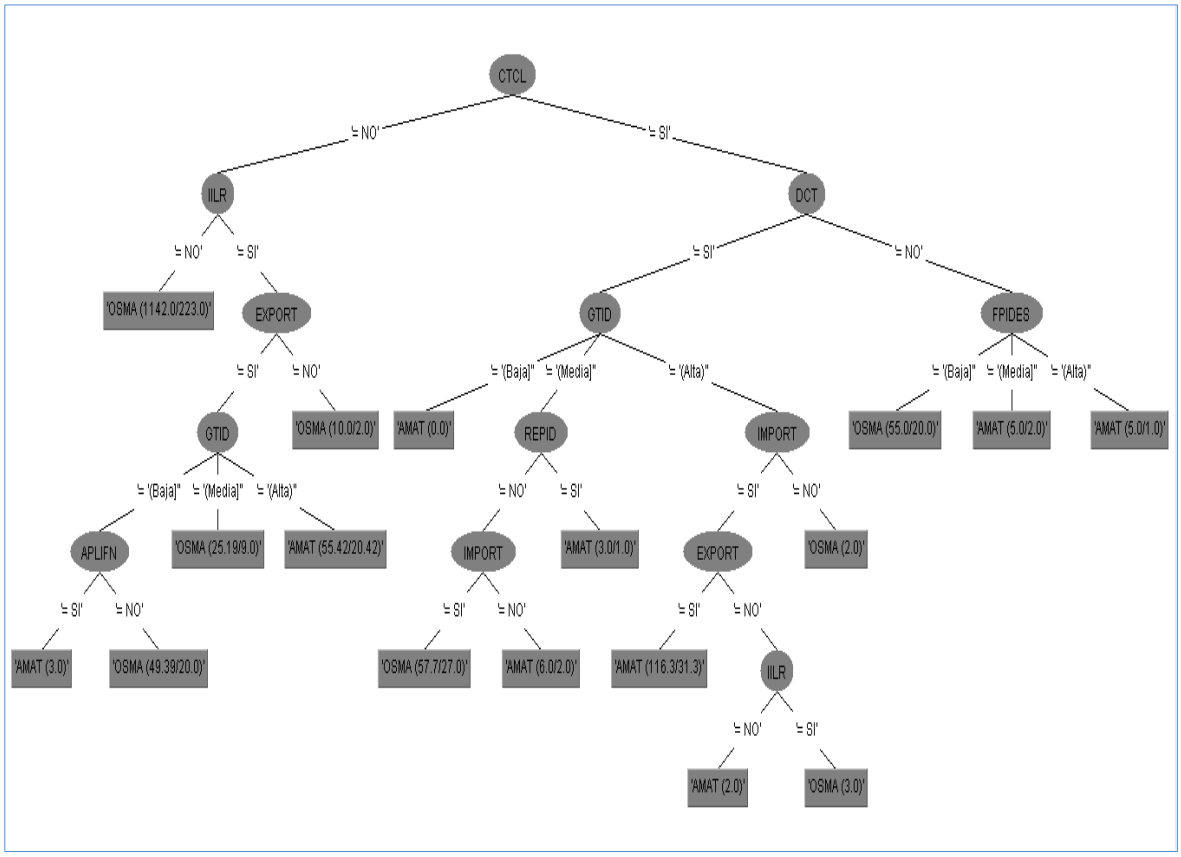


Figura 4.8. Árbol de decisión para la clasificación de empresas de alta tecnología a partir de los 9 atributos obtenidos en el ranking. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software *Weka*.

Este primer árbol clasifica como empresas de alta y media-alta tecnología aquellas que presentan uno de los siguientes patrones:

- Un primer patrón, presente en 116 empresas, caracterizado por aquellas empresas que realizan colaboración tecnológica con clientes (CTCL), mantienen dirección o comité de tecnología (DCT), presentan un gasto

total en I+D (GTID) alto, realizan importaciones (IMPORT) y realizan exportaciones (EXPORT).

- Un segundo patrón, presente en 55 empresas, compuesto por aquellas que no realizan colaboración tecnológica con clientes (CTCL), incorporan ingenieros y/o licenciados de graduación reciente (IILR), realizan exportaciones (EXPORT), y presentan un gasto total en I+D (GTID) alto.
- Un tercer patrón, presente en 10 empresas, constituido por aquellas que realizan colaboración tecnológica con clientes (CTCL), no mantienen dirección o comité de tecnología (DCT) y reciben recursos financieros de la administración central para I+D (FPIDES) en las categorías media y alta.
- Un cuarto patrón, presente en 6 empresas, formado por aquellas que realizan colaboración tecnológica con clientes (CTCL), mantienen dirección o comité de tecnología (DCT), presentan un gasto total en I+D (GTID) medio y reclutan personal con experiencia profesional en el sistema público de I+D (REPID).
- Un quinto patrón, presente en 6 empresas, a partir de aquellas que realizan colaboración tecnológica con clientes (CTCL), mantienen dirección o comité de tecnología (DCT), presentan un gasto total en I+D (GTID) alto, importan (IMPORT), no exportan (Export) y no incorporan ingenieros y/o licenciados de graduación reciente (IILR).
- Finalmente, un sexto patrón presente en 3 empresas, donde se ubican aquellas que no realizan colaboración tecnológica con clientes (CTCL), incorporan ingenieros y/o licenciados de graduación reciente (IILR), realizan exportaciones (EXPORT), presentan un gasto total en I+D (GTID) bajo y aplican incentivos fiscales para I+D+I (APLIFN).

Se presentan los parámetros de calidad más comúnmente usados y calculados por *Weka* para esta clasificación:

- Instancias correctamente clasificadas: 1181 (76,69%)
- Instancias incorrectamente clasificadas: 359 (23,31%)

Como puede apreciarse en la Figura 4.8, el árbol es un tanto complejo y clasifica correctamente el 76,69 % del conjunto de empresas presentes en la

muestra. Se resaltan como patrones de especial interés aquellos que con pocos rasgos (variables o atributos) están presentes en un mayor número de empresas; también son interesantes aquellos otros patrones que, al estar presente en pocas empresas, pueden pasar fácilmente desapercibidos, o pueden ser de difícil identificación.

Para el segundo caso y con el propósito de construir un árbol de decisión que permita realizar la clasificación, pero con una mejor interpretación, se fueron incorporando de manera iterativa al árbol los siguientes atributos en el ranking de atributos previamente presentado. A continuación, se presenta el árbol finalmente seleccionado.

El segundo árbol está formado a partir de un grupo de atributos (rasgos) que reúnen un mayor grupo de empresas por patrón identificado en comparación con los expuestos en el primer caso. Se parte de todos los atributos seleccionados en el ranking excepto GTID Y FPIDES y se obtiene un árbol de decisión adecuado para su interpretación y con una calidad de clasificación similar. El árbol obtenido puede verse en la Figura 4.9 y en él intervienen los atributos CTCL, DCT, IIRL, EXPORT, y APLFN.

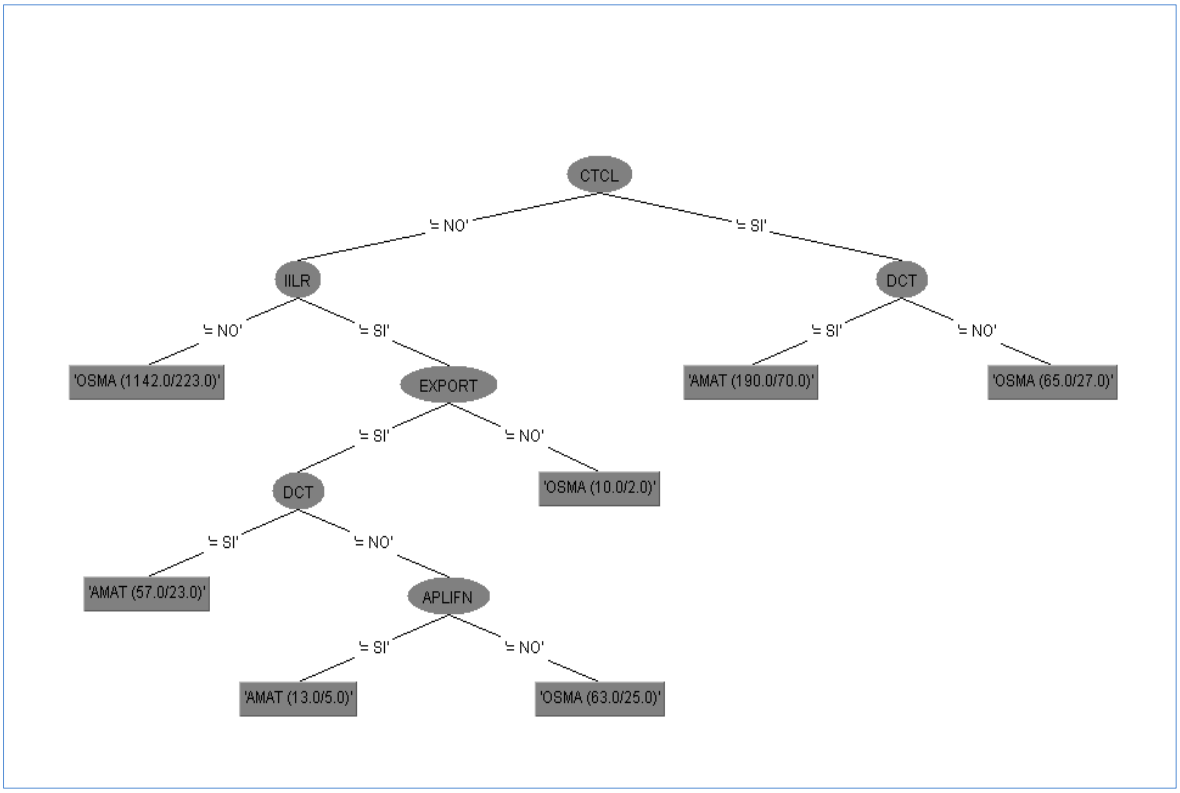


Figura 4.9. Árbol de decisión para la clasificación de empresas de alta tecnología considerando ocho de los principales atributos. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software *Weka*.

Este árbol clasifica como empresas de alta y media-alta tecnología aquellas que presentan uno de los tres siguientes patrones:

- En un primer patrón, presente en 190 empresas, se identifican como empresas de alta y media-alta tecnología aquellas que realizan colaboración tecnológica con clientes (CTCL) y mantienen dirección o comité de tecnología (DCT).
- En un segundo patrón, presente en 57 empresas, se incluyen aquellas que no realizan colaboración tecnológica con clientes (CTCL), pero incorporan ingenieros y/o licenciados de graduación reciente (IILR), realizan exportaciones (EXPORT) y mantienen dirección o comité de tecnología (DCT).
- En un tercer patrón de empresas tecnológica, identificado en 13 empresas, se encuentran aquellas que no realizan colaboración tecnológica con clientes (CTCL), no cuentan con una dirección o comité

de tecnología (DCT), pero incorporan ingenieros y/o licenciados de graduación reciente (IILR), realizan exportaciones (EXPORT) y aplican incentivos fiscales para I+D+I (APLIFN).

A continuación se presentan los parámetros de calidad que se obtienen después de aplicar el algoritmo J48 para esta clasificación:

- Instancias correctamente clasificadas: 1165 (75,65%)
- Instancias incorrectamente clasificadas: 375 (24,35%)

El árbol obtenido, que se muestra en la Figura 4.9, es más sencillo de interpretar y clasifica correctamente el 75,65% de las empresas presentes en la muestra; además reúne un mayor número de empresas (190, 57, 13, respectivamente) en comparación con los expuestos en el primer árbol (116, 55,10, 6,6 y 3 empresas respectivamente). Los resultados de las instancias correctamente clasificadas son muy similares entre el primer árbol y el segundo (76,68% y 75,65%, respectivamente), no obstante, la mejora en interpretación del segundo árbol puede justificar una disminución en 1,03% del número de empresas correctamente clasificadas.

Los dos árboles tienen en común algunas características comunes de las actividades en I+D+I en las empresas de alta tecnología, como son: la habitual colaboración tecnológica con clientes (CTCL), la existencia de dirección o comité de tecnología (DCT), la realización de exportaciones (EXPORT), la incorporación de ingenieros y/o licenciados de graduación reciente (IILR) y acogerse a incentivos fiscales para la I+D+I (APLIFN).

La tercera técnica de minería de datos que se aplica en este estudio es la generación de reglas de asociación. Parece claro que entre las variables o atributos presentes en la muestra hay reglas que relacionan sus valores, algunas reglas pueden resultar obvias, pero otras pueden estar más ocultas, y ese es precisamente el objetivo de aplicar un algoritmo de búsqueda de reglas de asociación, descubrir relaciones ocultas en los datos.

En este caso se utiliza el algoritmo el *Apriori*[14] implementado en *Weka*. A este algoritmo hay que indicarle, entre otras cosas, un soporte mínimo para las reglas, es decir, el número mínimo de objetos en la muestra en la que la regla se cumple, un número máximo de reglas a generar, y una confianza mínima, o lo que es lo mismo, un error máximo que se produciría al dar como cierta la regla con carácter general. Para la obtención de las reglas se fijó un soporte mínimo del 10% de las empresas de la muestra y una confianza mínima del 0,9.

En la primera de las ejecuciones del algoritmo *Apriori* se trabaja con las 9 variables seleccionadas en el ranking de atributos, adicionando la variable de clasificación de empresas de alta y media-alta tecnología (AMAT_OSMA).

Las 10 mejores reglas obtenidas tuvieron un soporte de al menos un 65%, es decir, 1101 empresas en las que se cumplían. Las 10 mejores reglas encontradas fueron:

1. CTCL=NO FPIDES='(Baja]' IILR=NO 1103 ==> REPID=NO 1102 conf:(1)

2. DCT=NO IILR=NO 1065 ==> REPID=NO 1064 conf:(1)

3. DCT=NO FPIDES='(Baja]' IILR=NO 1046 ==> REPID=NO 1045 conf:(1)

4. CTCL=NO DCT=NO IILR=NO 1022 ==> REPID=NO 1021 conf:(1)

5. APLIFN=NO CTCL=NO FPIDES='(Baja]' IILR=NO 1020 ==> REPID=NO 1019 conf:(1)

6. CTCL=NO DCT=NO FPIDES='(Baja]' IILR=NO 1008 ==> REPID=NO 1007 conf:(1)

7. FPIDES='(Baja]' IILR=NO 1211 ==> REPID=NO 1209 conf:(1)

8. DCT=NO 1171 ==> REPID=NO 1169 conf:(1)

9. DCT=NO FPIDES='(Baja]' 1143 ==> REPID=NO 1141 conf:(1)

10. CTCL=NO IILR=NO 1142 ==> REPID=NO 1140 conf:(1)

[14] *Weka* lo define como un algoritmo que reduce iterativamente el soporte mínimo hasta encontrar el número requerido de reglas con la confianza mínima dada. Para mayor información consultar las obras de Liu, Hsu, & Ma (1998) y Agrawal & Srikant (1994).

Se observa un conjunto de reglas en las que no está presente la variable de clasificación.

En la segunda de las ejecuciones del algoritmo, utilizando las variables que seleccionó el ranking más la variable de clasificación, puede ser interesante analizar aquellas 10 primeras reglas entre las 1000 mejores en las que intervenía la variable de clasificación. Las 10 mejores reglas obtenidas presentan un soporte de al menos un 55%, es decir, 847 empresas en las que se cumplían.

De nuevo se observa que esas reglas permiten identificar características de las empresas que no pueden ser consideradas de alta tecnología y media-alta tecnología:

163. DCT=NO AMAT_OSMA=OSMA 930 ==> REPID=NO 929 conf:(1)

164. CTCL=NO IILR=NO AMAT_OSMA=OSMA 919 ==> REPID=NO 918 conf:(1)

165. DCT=NO FPIDES='(Baja]' AMAT_OSMA=OSMA 914 ==> REPID=NO 913 conf:(1))

193. CTCL=NO FPIDES='(Baja]' IILR=NO AMAT_OSMA=OSMA 895 ==> REPID=NO 894 conf:(1)

194. CTCL=NO DCT=NO AMAT_OSMA=OSMA 892 ==> REPID=NO 891 conf:(1)

224. DCT=NO IILR=NO AMAT_OSMA=OSMA 869 ==> REPID=NO 868 conf:(1)

225. APLIFN=NO DCT=NO FPIDES='(Baja]' AMAT_OSMA=OSMA 866 ==> REPID=NO 865 conf:(1)

226. DCT=NO FPIDES='(Baja]' IILR=NO AMAT_OSMA=OSMA 858 ==> REPID=NO 857 conf:(1)

227. APLIFN=NO CTCL=NO IILR=NO AMAT_OSMA=OSMA 854 ==> REPID=NO 853 conf:(1)

228. APLIFN=NO CTCL=NO DCT=NO AMAT_OSMA=OSMA 850 ==> REPID=NO 849 conf:(1)

En la tercera de las ejecuciones del algoritmo se opta por trabajar con un conjunto reducido de atributos seleccionados en el ranking. En concreto, se selecciona las tres características (o variables) más importantes: colaboración tecnológica con clientes (CTCL), dirección o comité de tecnología (DCT) e incorporación de ingenieros y/o licenciados de graduación reciente (IILR). Con esas variables se solicita al algoritmo que busque un total de 100 reglas, que relacionen esas variables con la variable de clasificación.

Las reglas obtenidas presentan un soporte de al menos un 10%, es decir, 154 empresas en las que se cumplían, en este caso las reglas son:

1. DCT=NO IILR=NO AMAT_OSMA=OSMA 869 ==> CTCL=NO 843 conf:(0.97)

3. DCT=NO AMAT_OSMA=OSMA 930 ==> CTCL=NO 892 conf:(0.96)

4. CTCL=NO DCT=NO AMAT_OSMA=OSMA 892 ==> IILR=NO 843 conf:(0.95)

6. DCT=NO AMAT_OSMA=OSMA 930 ==> IILR=NO 869 conf:(0.93)

7. IILR=NO AMAT_OSMA=OSMA 989 ==> CTCL=NO 919 conf:(0.93)

8. CTCL=NO AMAT_OSMA=OSMA 993 ==> IILR=NO 919 conf:(0.93)

10. CTCL=NO IILR=NO AMAT_OSMA=OSMA 919 ==> DCT=NO 843 conf:(0.92)

13. DCT=NO AMAT_OSMA=OSMA 930 ==> CTCL=NO IILR=NO 843 conf:(0.91)

14. AMAT_OSMA=OSMA 1101 ==> CTCL=NO 993 conf:(0.9)

Como puede apreciarse, la totalidad de estas reglas consideran la variable de clasificación, e informan sobre aquellas empresas que no pueden ser consideradas de alta y media-alta tecnología.

Finalmente, la última técnica de minería de datos que se aplica sobre los datos de la encuesta es la de *clustering* (agrupamiento), también conocida como clasificación no supervisada. En este caso se trataba de buscar grupos de empresas relacionadas a partir de los valores de los atributos. Existe una gran variedad de algoritmos de *clustering*, de todos ellos se eligieron dos de los que se encuentran implementados en *Weka*: el algoritmo EM y el de las k-medias. La ventaja del primero es que permite identificar de manera automática el número de grupos, mientras que el segundo destaca por su sencillez y rapidez, pero exige determinar de antemano el número de grupos a crear. De acuerdo con Weka el algoritmo EM asigna a cada instancia una distribución de probabilidad que indica la probabilidad de pertenencia a cada cluster. La elección del número de clusters se realiza mediante una validación cruzada, aunque también se puede dar un número prefijado. En el caso del algoritmo de las k-medias, se seleccionan aleatoriamente *k* centroides para cada cluster y de manera iterativa se van asignando las instancias a los *clusters* con centroides más cercanos, se actualizan las posiciones de los centroides y se repite el proceso hasta que no sean necesarias más recolocaciones de instancias.

Se aplican ambos algoritmos de *clustering* considerando las 9 variables seleccionadas por el ranking (CTCL, DCT, IIRL, IMPORT, GTID, REPID, FPIDES, APLIFN,

EXPORT) y tomando la variable de clasificación únicamente para evaluar los grupos creados. En primer lugar, se aplica el algoritmo EM solicitándole que se genere el número de grupos de forma automática.

En concreto son creados 7 grupos, uno de ellos, el más numeroso, con el 34% de empresas de la muestra fue identificado como el grupo de empresas que no pertenecían al sector de alta tecnología (OSMA).

El grupo que el algoritmo asocia a las empresas de alta tecnología está compuesto por un 33% de las empresas. Además de esos dos grupos se generan otros cinco con un 9%, 12%, 0,32%, 11% y 15% de las empresas, y que el algoritmo no pudo asociarles ninguna clase. El número de empresas que quedan mal clasificadas es alto: un 60,45%.

Después de que el algoritmo EM realiza la clasificación anterior de forma automática, se le solicita que únicamente genere dos grupos, entonces clasifica las empresas en dos categorías. La primera con el 67% de la muestra y segunda con el 33%, esta última es la que el algoritmo asocia a las empresas de alta tecnología. El porcentaje de error en este caso es sensiblemente menor: un 27.57%.

Igualmente se decide contrastar la clasificación utilizando el algoritmo de las *k-medias* solicitando la creación de dos grupos. El porcentaje de error en este caso fue menor: un 25,19% (74,80 % de acierto). Además, este algoritmo proporciona información sobre los centroides, que pueden verse en este caso como prototipos que representan a las empresas de cada grupo. La Tabla 4.8 muestra la información de los dos grupos de empresas obtenidos por este algoritmo. El primer grupo (cluster 0) es identificado por el algoritmo como empresas de alta y media-alta tecnología (AMAT), mientras que el segundo grupo (cluster 1) representa a las empresas que de sectores menos tecnológicos (OSMA).

Características	**Cluster 0**	**Cluster 1**
Número total de empresas	439	1101
Porcentaje	24%	76%
Valor atributo APLIFN	SI	NO
Valor atributo CTCL	SI	NO
Valor atributo DCT	SI	NO
Valor atributo EXPORT	SI	SI
Valor atributo FPIDES	BAJA	BAJA

Características	Cluster 0	Cluster 1
Valor atributo GTID	ALTA	BAJA
Valor atributo IILR	NO	NO
Valor atributo IMPORT	SI	SI
Valor atributo REPID	NO	NO

Tabla 4.8. Características de los dos grupos de empresas generados por el algoritmo de las k medias. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y utilizando el software *Weka*.

Considerando solo las características en las que su valor difiere para cada cluster, se pueden identificar como variables relevantes para describir el cluster (0) correspondiente a las empresas de alta y media alta tecnología a: si la empresa aplica incentivos fiscales para I+D e innovación tecnológica (APLIFN), si realiza colaboración tecnológica con clientes (CTCL), si mantiene una dirección o comité de tecnología (DCT) y si el gasto total en I+D se realiza en la categoría "alta" (GTID).

4.2.4 Análisis y evaluación de resultados.

A la vista de los resultados obtenidos por los diferentes algoritmos parece claro que existe un conjunto de rasgos (variables o atributos) que mejor describen a las empresas de alta y media alta tecnología, tal cual lo evidencia el ranking de atributos seleccionados. Dentro de las empresas de sectores de alta o media-alta tecnología se perciben diferentes patrones cada uno con rasgos característicos en actividades en I+D+I, como se observa a través de los algoritmos de árboles de decisión. Las reglas de asociación evidencian que resulta más sencillo identificar características que automáticamente permitan identificar aquellas empresas que no se puedan considerar como de alta y media-alta tecnología. Ha resultado imposible obtener reglas simples que permitan señalar categóricamente cuando una empresa pertenece a un sector de alta tecnología. Los resultados después de ejecutar los algoritmos de *clustering* permiten identificar la existencia de dos grupos de empresas, uno de los cuales se puede considerar de alta y media-alta tecnología. Se destaca como rasgo de especial aportación a las empresas de alta y media alta tecnología la colaboración tecnológica con clientes (CTCL).

Finalmente, también es interesante destacar que en general las tasas de error al clasificar las empresas haciendo uso de los valores de las variables relacionadas con el I+D son aceptables, teniendo en cuenta la dificultad que supone la clasificación a priori de la muestra, la cual se realiza únicamente considerando el sector manufacturero al que pertenece la empresa, tomando

como referencia la Clasificación Nacional de Actividades Económicas (CNAE-2009). De igual manera dentro de un mismo sector pueden coexistir empresas que respondan de manera muy diferente a acciones vinculadas con la innovación, la I+D, la gestión de contratación de personal, etc. En definitiva, empresas de sectores a priori de no alta y media-alta tecnología pueden responder perfectamente a perfiles de empresas alta y media-alta tecnología en lo que se refiere a sus actividades en I+D+I.

4.2.5 Conclusiones del análisis de aplicación de técnicas de minería de datos

El uso de técnicas de minería de datos resulta útil para tratar de descubrir asociaciones o relaciones ocultas en los datos. En este caso se usan para sistematizar el proceso de clasificación de empresas de alta y media-alta tecnología, tratando de identificar rasgos o características de estas empresas que las puedan diferenciar de otras de sectores con una componente tecnológica menor.

Los árboles de decisión generados son capaces de clasificar adecuadamente más del 76% de las empresas de alta y media-alta. Además, es de destacar que para la clasificación elaborada se ha requerido un numero número muy limitado de variables o indicadores relevantes, aspecto que se considera deseable.

Las técnicas de *clustering* han resultado ser útiles para identificar rasgos discriminantes de las empresas de carácter más tecnológico. Al igual que los árboles de decisión, la clasificación no supervisada o *clustering* también obtuvo buenos resultados, con tasas de acierto en torno al 74,8% de las empresas de la muestra.

En definitiva, los análisis realizados ofrecen información sobre los vínculos entre I+D e innovación tecnológica en la empresa, en el contexto de la alta y media alta tecnología del sector industrial de manufactura.

4.3 ANÁLISIS FACTORIAL EXPLORATORIO

El análisis factorial, dependiendo del ámbito de la investigación y la finalidad de ésta, puede ser considerado una técnica descriptiva si es utilizada para reducir información (García et al. 2000). Esta reducción consiste en hallar factores (a partir de las variables originales) con una mínima pérdida de información.

En el contexto de la presente investigación, y acorde a los objetivos planteados, el análisis factorial exploratorio se utiliza para establecer la existencia o no de patrones ocultos (o factores latentes) en las actividades de I+D+I en las empresas de alta y media-alta tecnología del sector industrial de manufactura. De nuevo, el análisis se realiza a partir de un conjunto amplio de empresas industriales españolas. De existir estos factores latentes (o patrones ocultos) se consideran una reducción de las variables originales.

Para la exploración y descripción de patrones se cuenta con un grupo de variables (atributos o rasgos) de índole numérica o de escala. La selección de las variables se realiza acorde al agrupamiento previo en los cinco bloques principales de actividades en I+D+I:

1) Actividades de Comercialización (18 variables)

2) Actividades Innovadoras y de apoyo tecnológico (11variables)

3) Incentivos y financiación I+D (10 variables)

4) Actividades de Innovación (1 variables)

5) Recursos Humanos (12variables)

Se dispone sólo de variables (52 en total) de índole cuantitativa (numérica) puesto que es un requerimiento de entrada de la técnica de análisis factorial. Específicamente, se utilizan los datos correspondientes a las 439 empresas presentes en los sectores industriales de manufactura de alta y media-alta tecnología.

De acuerdo con Uriel (1995), se procede a destacar un conjunto principal de hipótesis de índole estadística asociadas a los contrastes con los cuales se

analiza la pertinencia de la aplicación del análisis factorial a un conjunto de variables observables:

1. La cuestion previa a la realización del análisis factorial que se plantea es la siguiente: ¿Están correlacionadas entre sí las variables originales?. Si no lo estuvieran, no existirian factores comunes y, por lo tanto, no tendría sentido aplicar el análisis factorial. Para dar respuesta se utiliza el contraste de Esfericidad de Barlet. La estructura de la hipótesis es la siguiente:
    - Hipótesis nula (Ho): los coeficientes de correlación son nulos.
    - Hipótesis alterna (H1): los coeficientes de correlación son no nulos.

    La medida de Esfericidad de *Barlett* va acompañada de la medida de adecuación muestral de *Kaiser*, *Meyer* y *Olkin* (KMO), en el caso de que exista adecuación muestral de los datos a un análisis factorial el KMO estará cercano a uno, un valor cercano a 0,90 es muy bueno y un valor por debajo de 0,5 no es aceptable.
2. Se pueden realizar contrastes estadísticos formales de bondad del ajuste en el caso de que el método de extracción seleccionado sea el de máxima verosimilitud. Las hipótesis nula y alternativa en el contraste de bondad de ajuste son:
    - Ho: Existen m factores comunes.
    - H1: No existen factores comunes.

De acuerdo con Comrey (1985), se procede a realizar el análisis factorial siguiendo las etapas de selección de variables a utilizar, cálculo de la matriz de correlaciones entre variables, extracción de factores e interpretación de la matriz de los factores rotados.

4.3.1 Estadísticas descriptivas de la totalidad de variables disponibles

En la Tabla 4.9 se presenta la descripción estadística de cada una de las variables que se utilizarán en el análisis factorial exploratorio. Para ello se contempla: descripción de la variable, valores mínimos y máximos, media y la desviación típica.

	Mínimo	Máximo	Media	Desv. típ.
AUTONOMÍA TECNOLÓGICA	,0000	100,00	58,61	48,18
CONCENTRACIÓN DE CLIENTES	3,00	100,00	48,41	28,79

	Mínimo	Máximo	Media	Desv. típ.
COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	,00	69,00	5,10	11,27
COMERCIALIZACIÓN PRODUCTOS NACIONALES	,00	50,00	2,51	6,51
COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	,00	71,00	7,61	13,58
DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES	,00	20615987,00	228314,38	1,56E6
DEDUCCIONES DE I+D SOBRE GASTOS I+D	,00	233,90	12,75	27,19
DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES	,00	4817153,00	23210,04	2,43E5
DEDUCCIONES DE INNOVACIÓN TECNOLÓGICA SOBRE GASTOS I+D	,00	123,70	1,72	9,24
DEDUC. TOTALES I+D E INNOV. SOBRE GASTOS TOTALES I+D	,00	233,90	14,47	30,05
DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES	,00	21378646,00	251524,43	1,60E6
EXPORTACIÓN DE TECNOLOGÍA	,00	13066,90	60,93	670,65
FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS	,00	3588,00	40,42	235,69
FINANCIACIÓN PÚBLICA I+D ESTADO	,00	26066,50	174,08	1406,17
FINANCIACIÓN PÚBLICA I+D OTROS	,00	1433,00	10,05	79,60
FINANCIACIÓN PÚBLICA I+D TOTAL	,00	27427,60	224,56	1507,20
GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS	,00	555962,00	14243,75	46851,61
GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA	,00	1945867,00	16179,64	1,01E5
GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR	,00	374,20	34,12	62,51
GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	,00	535,40	45,20	73,48
GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN	,00	1667885,00	6699,40	80864,33
GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL	,00	,80	,01	,06
GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR	,00	275,40	8,67	25,73
GASTOS EXTERNOS TOTALES EN FORMACIÓN	,00	5559616,00	61339,86	2,98E5
GASTOS TOTALES EXTERNOS EN FORMACIÓN SOBRE COSTES DE PERSONAL	,00	2,40	,31	,35
GASTOS TOTALES EXTERNOS EN FORMACIÓN POR TRABAJADOR	,00	1342,00	147,28	181,01
GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MÁRKETING	,00	444767,00	3620,26	29655,09
GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MÁRKETING POR TRABAJADOR	,00	402,60	9,75	36,99
GASTOS EXTERNOS I+D	,00	2,95E8	1,708E6	1,568E7
GASTOS EXTERNOS DE I+D A EMPRESAS	,00	295364,90	1130,42	14421,35
GASTOS INTERNOS I+D	,00	1,42E8	1,54E6	8,68E6

	Mínimo	Máximo	Media	Desv. típ.
GASTOS DE PUBLICIDAD E I+D SOBRE VENTAS	,00	46,20	3,07	5,55
GASTOS DE PUBLICIDAD SOBRE VENTAS	,00	29,70	,86	2,22
GASTOS TOTALES I+D	,00	4,37E8	3,25E6	2,38E7
INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	,00	219,10	2,77	11,21
ÍNDICE DE DINAMISMO DE LOS MERCADOS	,00	100,00	35,53	31,23
GASTOS DE I+D SOBRE VENTAS	,00	43,40	2,21	4,70
ÍNDICE DE GLOBALIZACIÓN MERCADOS	,00	100,00	36,44	45,87
IMPORTACIÓN DE TECNOLOGÍA	,00	168207,40	1189,71	11841,91
MODELOS DE UTILIDAD	0	17	,11	,93
NÚMERO DE INNOVACIONES DE PRODUCTO	0	67	1,50	5,13
NÚMERO DE ESTABLECIMIENTOS	1	32	2,24	3,37
NÚMERO DE MERCADOS	1	5	2,10	1,18
PATENTES REGISTRADAS EN ESPAÑA	0	27	,31	1,93
PATENTES REGISTRADAS EN EL EXTRANJERO	0	308	1,13	15,29
INTENSIDAD IMPORTADORA	,0	68,6	17,43	17,14
PROPENSIÓN EXPORTADORA	,0	100,0	43,28	32,86
RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN	,0	2339,7	12,46	113,64
SALDO TECNOLOGICO RELATIVO	-200,30	48,90	-2,13	13,40
VENTAS	159558	6086889990	1,29E8	5,25E8
VALOR DE LAS EXPORTACIONES	0	5080462318	82960916,51	4,10E8
VALOR DE LAS IMPORTACIONES	0	2267844288	40647752,41	1,99E8
N válido (según lista)				

Tabla 4.9. Análisis Factorial Exploratorio, estadísticas descriptivas de las variables disponibles. Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Se observa con especial atención durante al análisis factorial exploratorio aquellas variables con un número menor de observaciones a 439. No obstante, si se considera pertinente remplazar dichos valores faltantes o valores perdidos, según sea el caso, se remplazan por el valor medio de la variable. Se remplazan únicamente valores perdidos en aquellas variables en las que este valor sea próximo o inferior al 1% de la muestra. Para valores superiores se procede a la exclusión de la variable del análisis. Se presenta en la Tabla 4.10 el resumen de valores perdidos y conteo de valores extremos.

Nombre de la variable		N	Perdidos		No de extremos[b]	
			Recuento	Porcentaje	Bajos	Altos
1	AUTEC	439	0	,0	0	0
2	CCLIE	439	0	,0	0	0
3	COMPE	439	0	,0	0	76
4	COMPN	439	0	,0	0	90

Nombre de la variable		N	Perdidos		No de extremos[b]	
			Recuento	Porcentaje	Bajos	Altos
5	COMSV	439	0	,0	0	57
6	DEDID	439	0	,0	0	96
7	DEDIDG	270	169	38,5	0	14
8	DEDIT	439	0	,0	.	.
9	DEDITG	270	169	38,5	.	.
10	DEDTOGN	270	169	38,5	0	17
11*	DEDTON	439	0	,0	0	92
12	EXPTEC	439	0	,0	.	.
13	FPIDCA	439	0	,0	.	.
14	FPIDES	439	0	,0	.	.
15	FPIDOT	439	0	,0	.	.
16*	FPIDTO	439	0	,0	0	108
17*	GEFI	439	0	,0	0	63
18*	GEFIF	439	0	,0	0	55
18*	GEFIFN	439	0	,0	0	50
20*	GEFIN	439	0	,0	0	32
21*	GEFIT	439	0	,0	0	92
22	GEFITC	439	0	,0	.	.
23*	GEFITN	439	0	,0	0	87
24 *	GEFT	439	0	,0	0	40
25 *	GEFTC	439	0	,0	0	12
26 *	GEFTN	439	0	,0	0	18
27	GEFVM	439	0	,0	.	.
28	GEFVMN	439	0	,0	.	.
28 *	GEID	439	0	,0	0	77
39 *	GEIDE	439	0	,0	0	79
31 *	GIID	439	0	,0	0	58
32 *	GPIDSV	439	0	,0	0	38
33 *	GPV	439	0	,0	0	44
34 *	GTID	439	0	,0	0	64
35 *	IBESV	439	0	,0	0	32
36	IDMERPN	439	0	,0	0	0
37 *	IDV	439	0	,0	0	37
38	IGM	439	0	,0	0	0
39	IMPTEC	439	0	,0	.	.
40	MODUT	439	0	,0	.	.
41 *	*NIP*	*434*	*5*	*1,1*	0	63
42 *	NTES	439	0	,0	0	58
43	NUMER	439	0	,0	0	0
44	*PATESP*	*438*	*1*	*,2*	.	.
45	PATEXT	439	0	,0	.	.
46 *	PM	439	0	,0	0	7
47	PX	439	0	,0	0	0
48 *	RXM	439	0	,0	0	44
49	STEREL	439	0	,0	.	.
50 *	VENTAS	439	0	,0	0	49
51 *	VEXPOR	439	0	,0	0	53
52 *	VIMPOR	439	0	,0	0	65

Tabla 4.10. Valores perdidos. Elaboración propia a partir de datos de la Fundación SEPI (2012).

Se excluyen del análisis factorial exploratorio las variables de deducciones por I+D en el impuesto de sociedades (DEDID), deducción en I+D sobre gastos en I+D (DEDIDG) y deducciones totales en el impuesto de sociedades (DEDTON), debido a que presentan porcentajes superiores a 1.5 % de valores perdidos.

Se admiten las variables patentes registradas en España (PATESP) y patentes registradas en el extranjero (PATEX), debido a que los datos perdidos están próximos 1.0%. En los casos señalados, los porcentajes son inferiores al 1,2%, durante el análisis factorial estas variables demandan especial atención. Las variables acompañadas con el carácter (*) se mantienen bajo observación y especial interés debido al número de casos con valores extremos en relación a la media.

En los Anexos 6.2 y 6.3 de la presente tesis se encuentra disponible el resumen de los estadísticos de la totalidad de las variables disponibles para el Análisis Factorial Exploratorio y asimismo el resumen de los gráficos Q-Q requeridos para la inspección visual de que tanto la distribución de los datos presentes en las variables se ajusta a una distribución normal.

4.3.2 Análisis Factorial por grupos de variables

De acuerdo con Hair et al (1999), el proceso de decisión del análisis factorial considera diferentes etapas y criterios de decisión, siendo posible, una vez evaluado cada criterio, regresar a etapas anteriores, ya sea para incluir o excluir variables, o para establecer un método diferente de extracción de factores, o de rotación de los mismos.

Atendiendo al propósito de identificar si existen o no factores latentes (u ocultos) al interior del conjunto de variables cuantitativas que describen las actividades en I+D+I en las empresas de alta y media-alta tecnología, se selecciona el método de extracción de máxima verosimilitud debido a que este método, frente a otros, tiene la ventaja de que permite contrastar el ajuste del modelo a los datos a través de un índice que sigue una distribución *ji* cuadrado. Igualmente permite obtener errores típicos y pruebas de significación alrededor de los parámetros estimados, aunque tiene como requerimiento que la muestra proceda de una muestra normal multivariada (Lloret-Segura et al. 2014).

Los criterios relevantes de decisión para evaluar la aplicación del AFE que se analizan son:

- Determinante de la matriz de correlaciones, el cual debe presentar valores cercanos a cero, pero nunca cero.
- La prueba de esfericidad de *Barlett*, con la cual se contrasta la hipótesis de que los coeficientes de correlación son nulos, y el KMO permite evaluar la medida de adecuación muestral. Para no rechazar la hipótesis nula se consideran niveles de significación (Sig) mayores a 0,05 y para evaluar la medida de adecuación muestral valores para el KMO superiores a 0,5 y de preferencia cercanos a uno.
- Comunalidades, si éstas son mayores que uno, es un indicativo de que los resultados deben ser interpretados con precaución.
- Finalmente, la prueba de bondad de ajuste con la cual se contrasta la hipótesis de que existen los factores. Para no rechazar la hipótesis nula se consideran niveles de significación (Sig) mayores a 0,05.

A continuación se plantean dos fases para la identificación de factores latentes (o patrones ocultos) al interior de las actividades en I+D+I.

En la primera fase, se analiza la existencia o no de los factores por grupo de actividades. Se parte de las variables relacionadas con las actividades de comercialización, actividades innovadoras y de apoyo tecnológico, de incentivos y financiación de I+D+I y recursos humanos. Para las actividades de innovación sólo se cuenta con la variable de número de innovaciones de producto (NIP), variable que se incluirá en los análisis de la siguiente etapa.

En la segunda fase, se analiza el total de variables provenientes de todos los grupos de actividades en I+D+I, seleccionando aquellas que presenten correlaciones altas con las demás variables. Las seleccionadas son:

{DEDID, DEDIT, FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, MODUT, PATESP, PATEXT, PM, PX, VEXPOR, VIMPOR}.

A partir de este grupo de variables se va aplicando el análisis factorial identificando aquellas variables con las correlaciones más bajas entre ellas y que al excluirlas se refleje una mejoría en alguno de los valores de los criterios de decisión.

Tanto en la primera fase como en la segunda se excluyen variables, donde cada iteración da origen a un modelo en evaluación, el cual se enumera para su identificación. Los resultados en cada caso se presentan en tablas donde se especifica las variables seleccionadas y los valores obtenidos para los criterios de decisión. Adicionalmente en esta fase se presentan las estadísticas descriptivas con los valores tipificados de las variables.

Seleccionadas las variables y el método de extracción, se procede en su orden a evaluar el valor del determinante de las correlaciones, el valor de la prueba de la esfericidad de *Barlett*, la presencia o no de comunalidades mayores a uno y la prueba de bondad de ajuste. En cada tabla igualmente aparecen las variables involucradas por sus siglas, en las tablas de estadísticos descriptivos los nombres van precedidos por una z, para indicar que los valores de la variable han sido tipificados.

Finalmente, se presenta una tabla resumen donde se informa de la existencia o no de factores latentes, atendiendo a los criterios de aceptación o rechazo de la hipótesis nula para los coeficientes de correlación y la bondad de ajuste.

4.3.3 Actividades de comercialización

El primer bloque de actividades en I+D+I en los cuales se analiza la existencia o no de patrones ocultos (factores latentes) corresponde a las variables relacionadas con las actividades de comercialización, y se compone por 15 variables.

Estadísticos descriptivos			
	Media	Desviación típica	N del análisis
Z_CONCENTRACIÓN DE CLIENTES	,03	,98	439
Z_COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	,27	1,33	439
Z_COMERCIALIZACIÓN PRODUCTOS NACIONALES	-,07	,82	439
Z_COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	,14	1,12	439
Z_EXPORTACIÓN DE TECNOLOGÍA	,00	,52	439
Z_GASTOS DE PUBLICIDAD SOBRE VENTAS	-,02	,88	439
Z_IMPORTACIÓN DE TECNOLOGÍA	,13	1,9	439
Z_NÚMERO DE ESTABLECIMIENTOS	,01	,45	439
Z_NÚMERO DE MERCADOS	,19	1,10	439
Z_ INTENSIDAD IMPORTADORA	,48	1,16	439
Z_ PROPENSIÓN EXPORTADORA	,56	1,06	439

Estadísticos descriptivos			
	Media	Desviación típica	N del análisis
Z_ RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN	-,007	,44	439
Z_VENTAS	,22	1,77	439
Z_VALOR DE LAS EXPORTACIONES	,24	1,90	439
Z_VALOR DE LAS IMPORTACIONES	,21	1,79	439

Tabla 4.11. Estadísticos descriptivos de las actividades de comercialización. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Como se observa en la Tabla 4.11, debido al requerimiento de normalidad, y dado que las variables han sido tipificadas, las variables con desviaciones típicas lejanas a 1 se deben analizar con especial atención y cuidado. Estas variables son: comercialización de productos extranjeros, importación de tecnología, ventas, valor de las exportaciones y de las importaciones.

Se presenta en la Tabla 4.12 el resumen de los modelos evaluados a partir del conjunto anterior de variables.

Modelo	Variables seleccionadas	Método Extracción / Método Rotación	Valor del de determinante de la matriz de correlaciones	Prueba de esfericidad de Barlett	KM0 > 0,5	Comunalidades (mayores a uno)	Prueba de bondad de ajuste
M_1	CCLIE, COMPE, COMPN, COMSV, EXPTEC, GPV, IMPTEC, NTES, NUMER, PM, PX, RXM, VENTAS, VEXPOR, VIMPOR.	Máxima verosimilitud / Sin Rotación.	0,000 Matriz no definida positiva	-	-	-	-
M_2	CCLIE, COMPE, COMPN, EXPTEC, IMPTEC, PX, VENTAS, VEXPOR, VIMPOR	Máxima verosimilitud / Sin Rotación.	0,004	0,000	0,674	SI	-
M_3	VENTAS, VEXPOR, VIMPOR	Máxima verosimilitud / Sin Rotación.	0,07	0,000	0,659	SI	*
M_4	EXPTEC, IMPTEC, VENTAS	Máxima verosimilitud / Sin Rotación.	0,851	0,00	0,5	SI	*

Tabla 4.12. Resumen de los modelos evaluados. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En los anteriores cuatro modelos se observa que en ninguno de ellos se puede establecer si existen o no factores latentes o patrones ocultos. En el primer caso la matriz es no definida positiva (no tiene inversa) y no se puede continuar con los análisis y por tanto no se acepta o rechaza hipótesis alguna. En estos casos el AFE no es adecuado para el conjunto de datos disponibles. Se procede a excluir aquellas variables, que tengan correlaciones muy bajas entre ellas.

En el segundo modelo, una vez analizada la matriz de correlaciones, se selecciona un conjunto de 9 variables compuesto por: concentración de clientes, comercialización de productos extranjeros, comercialización de productos nacionales, exportación de tecnología, intensidad importadora, ventas, valor de las exportaciones, valor de las importaciones. En este caso, el valor del determinante es cercano a cero (sin llegar a ser cero), con KMO (0,76) mayor a 0,5, aspectos favorables, pero se presentan comunalidades mayores a uno, aspecto bajo el cual, lo recomendable es afirmar que el modelo no es adecuado. Adicionalmente no es posible aplicar la prueba estadística para la bondad de ajuste.

En los modelos M_3 y M_4, donde se analiza la posibilidad de identificar o no patrones ocultos a partir de la información de ventas, valor de las exportaciones y el valor de las importaciones, así mismo, ventas, exportación de tecnología, e importación de tecnología; la conclusión es la misma, existen comunalidades mayores a uno, y no se puede establecer un valor para evaluar la hipótesis correspondiente al criterio de bondad de ajuste.

4.3.4 Actividades Innovadoras y de apoyo tecnológico

El segundo bloque de actividades en I+D+I en los cuales se analiza la existencia o no de patrones ocultos (factores latentes) corresponde a las variables relacionadas con las actividades de Innovación y de apoyo, se compone por 11 variables.

En las variables presentes en la Tabla 4.13, se identifica un dato perdido por cada variable y será remplazado por el valor medio de los datos presente en la variable.

Estadísticos descriptivos			
	Media	Desviación típica	N del análisis
Z_AUTONOMÍA TECNOLÓGICA	-,001	1,00	438
Z_GASTOS EXTERNOS I+D	,00	1,00	438
Z_GASTOS EXTERNOS DE I+D A EMPRESAS	,00	1,00	438
Z_GASTOS INTERNOS I+D	,00	1,00	438
Z_GASTOS DE PUBLICIDAD E I+D SOBRE VENTAS	-,0006	1,00	438
Z_ GASTOS TOTALES I+D	,00	1,00	438
Z_ INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	,00	1,00	438
Z_GASTOS DE I+D SOBRE VENTAS	-,001	1,00	438
Z_MODELOS DE UTILIDAD	,00	1,00	438
Z_PATENTES REGISTRADAS EN ESPAÑA	,00	1,00	438
Z_PATENTES REGISTRADAS EN EL EXTRANJERO	,00	1,00	438

Tabla 4.13. Estadísticos descriptivos de las actividades de innovación y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En la Tabla 4.14 se presenta el resumen de los modelos elaborados.

Modelo	Variables seleccionadas	Método Extracción / Método Rotación	Valor del de determinante de la matriz de correlaciones	Prueba de esfericidad de Barlett	KM0 > 0,5	Comunalidades (mayores a uno)	Prueba de bondad de ajuste
M_5	AUTEC, GEID, GEIDE, GIID, GPIDSV, GTID, IBESV, IDV, MODUT, PATESP, PATEXT.	Máxima verosimilitud / Sin Rotación.	0.000 Matriz no definida positiva	-	-	-	-
M_6	AUTEC, IBESV, MODUT, PATESP, PATEXT	Máxima verosimilitud / Sin Rotación.	0,486	0,000	0,45	SI	0,546
M_7	GEID, GEIDE, GIID, GPIDSV, GTID, IDV	Máxima verosimilitud / Sin Rotación.	0.000 Matriz no definida positiva	-	-	-	-

Tabla 4.14. Resumen de los modelos elaborados en las actividades de innovación y de apoyo tecnológico. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Los modelos M5, M6 y M7 se descartan debido a que, utilizando el método de extracción de máxima verosimilitud y sin rotación, tanto la prueba del KMO como la presencia de comunalidades mayores que 1, orientan a que los modelos no reflejan una estructura que permita resumirlos.

4.3.5 Incentivos y financiación de la I+D e Innovación

El tercer bloque de actividades en I+D+I en los cuales se analiza la existencia o no de patrones ocultos (factores latentes) corresponde a las variables relacionadas con las actividades de incentivos y financiación de I+D e Innovación, se compone por 7 variables.

Estadísticos descriptivos			
	Media	**Desviación típica**	**N del análisis**
Z_DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES	,00	1,00	439
Z_DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES	,00	1,00	439
Z_DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES	,00	1,00	439
Z_FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS	,00	1,00	439
Z_FINANCIACIÓN PÚBLICA I+D ESTADO	,00	1,00	439
Z_FINANCIACIÓN PÚBLICA I+D OTROS	,00	1,00	439
Z_FINANCIACIÓN PÚBLICA I+D TOTAL	,00	1,00	439

Tabla 4.15. Estadísticos descriptivos de las actividades en incentivos y financiación en I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Para el conjunto de variables presente en el grupo de actividades de deducción y financiación Tabla 4.15, se observa que las variables tienen media de valor cero y desviación típica de valor uno (variables con una distribución normal), y no se presentan valores perdidos. En la Tabla 4.16 se presenta el resumen de los modelos.

Modelo	**Variables seleccionadas**	**Método Extracción /Método Rotación**	**Valor del de determinante de la matriz de correlaciones**	**Prueba de esfericidad de Barlett**	**KM0 > 0,5**	**Comunalidades (mayores a uno)**	**Prueba de bondad de ajuste**
M_8	DEDID, DEDIDG, DEDIT, DEDITG, DEDTOGN, DEDTON, FPIDCA, FPIDES, FPIDOT, FPIDTO.	Máxima verosimilitud/ Sin Rotación	0.000 Matriz no definida positiva	-	-	-	-
M_9	DEDID, DEDIT, FPIDES, FPIDOT, FPIDCA	Máxima verosimilitud/ Sin Rotación	0,675	0,00	0,481	SI	0,727

Tabla 4.16. Resumen de los modelos en actividades de incentivos y financiación de la I+D. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Los modelos M8 y M9 se descartan debido a que, utilizando el método de extracción de máxima verosimilitud, y sin rotación, tanto la prueba del KMO con valores inferiores 0.5, y la presencia de comunalidades mayores que 1, orientan a que los modelos no reflejan una estructura que permita resumirlos, y que no es posible rechazar o no las hipótesis bondad de ajuste.

4.3.6 Talento Humano

El cuarto bloque de actividades en I+D+I en los cuales se analiza la existencia o no de patrones ocultos (factores latentes) corresponde a las variables relacionadas con el talento humano, se compone por 12 variables.

Estadísticos descriptivos			
	Media	Desviación típica	N del análisis
Z_GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR	,00	1,00	439
Z_GASTOS EXTERNOS TOTALES EN FORMACIÓN	,00	1,00	439
Z_GASTOS TOTALES EXTERNOS EN FORMACIÓN SOBRE COSTES DE PERSONAL	,00	1,00	439
Z_GASTOS TOTALES EXTERNOS EN FORMACIÓN POR TRABAJADOR	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MÁRKETING	,00	1,00	439
Z_GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MÁRKETING POR TRABAJADOR	,00	1,00	439

Tabla 4.17. Estadísticos descriptivos de las actividades en relación al talento humano. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

Para el conjunto de variables presente en el grupo de talento humano Tabla 4.17, se observa que las variables tienen media de valor cero y desviación típica de valor uno (variables con una distribución normal), y no se presentan valores perdidos. Se presenta el resumen de los modelos en la Tabla 4.18

Modelo	Variables seleccionadas	Método Extracción / Método Rotación	Valor del de determinante de la matriz de correlaciones	Prueba de esfericidad de Barlett	KM0 > 0,5	Comunalidades (mayores a uno)	Prueba de bondad de ajuste
M_{10}	GEFI, GEFIF, GEFIFN, GEFIN, GEFIT, GEFITC, GEFITN, GEFT, GEFTC, GEFTN, GEFVM, GEFVMN	Máxima verosimilitud / Sin Rotación.	1,82E-006	0,00	0,687	SI	-
M_{11}	GEFI, GEFIF, GEFIT, GEFT, GEFTN, GEFVM	Máxima verosimilitud / Sin Rotación.	0,02	0,00	0,825	SI	0,000 rechaza Ho
M_{12}	GEFI, GEFIF, GEFIT, GEFT, GEFVM	Máxima verosimilitud / Sin Rotación.	0,03	0,00	0,824	SI	0,00 rechaza Ho

Tabla 4.18. Resumen de los modelos elaborados para las actividades del talento humano. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En el modelo M_{10} utilizando el método de extracción de máxima verosimilitud, y sin rotación, no obstante, un valor del determinante cercano a cero (aspecto deseable), un valor de la prueba de esfericidad de Barlett de 0,000, lo

cual permite rechazar la hipótesis nula correspondientes y un valor de 0,687 para el criterio KMO superior a 0,5 que es el valor límite, se descarta el modelo debido a la presencia de comunalidades mayores a uno. No se puede entonces establecer la existencia o no de un factor oculto o patrón.

De igual manera se concluye para los modelos M_{11} y M_{12}, la presencia de comunalidades es un indicador de precaución, adicionalmente los valores de las pruebas de bondad de ajuste permiten no aceptar la hipótesis nula correspondiente a la existencia de los factores.

Por lo tanto, se puede afirmar que no hay un patrón oculto asociado a las variables de: gastos externos en formación en idiomas (GEFI); gastos externos en formación en ingeniería y formación técnica (GEFIF); gastos externos en formación en informática y tecnologías de la información (GEFIT); gastos externos totales en formación (GEFT); gastos externos en formación de en ventas y marketing por trabajador (GEFVM).

Se presenta en la Tabla 4.19 el resumen de los modelos finales.

Modelo	**Variables seleccionadas**	**Método Extracción / Método Rotación**	**Valor del de determinante de la matriz de correlaciones**	**Prueba de esfericidad de Barlett**	**KM0 > 0,5**	**Comunalidades (mayores a uno)**	**Prueba de bondad de ajuste**
M_{13}	DEDIT, FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, MODUT, PATESP, PATEXT, PM, PX, VEXPOR, VIMPOR.	Máxima verosimilitud / Sin Rotación.	1,05E-007	0,000	0,813	SI	0,000
M_{14}	FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, MODUT, PATESP, PATEXT, VEXPOR, VIMPOR.	Máxima verosimilitud / Sin Rotación.	1,57E-007	0,000	0,827	SI	0,000
M_{15}	FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, MODUT, VEXPOR, VIMPOR.	Máxima verosimilitud / Sin Rotación.	5,38 E-007	0,000	0,844	NO	0,000
M_{16}	FPIDES, GEFI, GEFIF, GEFIT, GEFT,	Máxima verosimilitud	2,72 E-006	0,000	0,831	NO	0,000

Modelo	Variables seleccionadas	Método Extracción / Método Rotación	Valor del de determinante de la matriz de correlaciones	Prueba de esfericidad de Barlett	KM0 > 0,5	Comunalidades (mayores a uno)	Prueba de bondad de ajuste
	GEFVM, GEID, GIID, VEXPOR, VIMPOR.	/ Sin Rotación.					
M_{17}	FPIDES, GEFI, GEFIF, GEFIT GEFT, GEFVM, GEID, GIID.	Máxima verosimilitud / Sin Rotación.	3,34 E-005	0,000	0,866	NO	0,000
M_{18}	FPIDES, GEFIF, GEFIT GEFT, GEFVM, GEID, GIID	Máxima verosimilitud / Sin Rotación.	0,000	0,000	0,846	N0	0,000

Tabla 4.19. Resumen de los modelos finales. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En relación a los modelos M_{13} y $M_{14,}$ las comunalidades mayores que uno y la imposibilidad de calcular un valor para la prueba de bondad de ajuste no permiten concluir sobre la existencia o no de factores latentes relacionados con las variables presentes en el análisis.

En cuanto al modelo M_{15} además de la presencia de comunalidades mayores que uno, el valor de la prueba de bondad de ajuste permite rechazar la hipótesis nula de la existencia de factores, por tanto se puede afirmar que no hay factores ocultos presentes en el conjunto de variables compuesta por (FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, MODUT, VEXPOR, VIMPOR.)

En el modelo M_{16}, se cumplen todos los criterios de toma de decisiones para evaluar la hipótesis nula de

Ho: Existen m factores comunes.

H1: No existen factores comunes.

Dado el valor de Sig (0,000) < 0,05 se rechaza la hipótesis nula de que existen factores comunes a partir de las variables (FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID, VEXPOR, VIMPOR).

En el modelo M_{17}, se cumplen todos los criterios de toma de decisiones para evaluar la hipótesis nula de

Ho: Existen m factores comunes.

H1: No existen factores comunes.

Dado el valor de Sig (0,000) < 0,05 se rechaza la hipótesis nula de que existen factores comunes a partir de las variables (FPIDES, GEFI, GEFIF, GEFIT, GEFT, GEFVM, GEID, GIID).

En el modelo $M_{18,}$ ya no es posible evaluar ninguno de los criterios de decisión. Debido a ello, ya no es posible seguir con la estrategia de excluir variables con aquellas correlaciones de menor valor con las demás. Se culmina así, la búsqueda de factores latentes o patrones ocultos.

4.3.7 Conclusiones del análisis factorial exploratorio utilizado para identificar la existencia no de patrones ocultos

En conclusión, el análisis factorial ha demostrado ser una técnica útil en el análisis de datos, en especial si el objetivo de uso es la identificación de patrones ocultos.

Los criterios de decisión a partir del análisis factorial, tales como el valor del determinante de la matriz de correlaciones, la prueba de esfericidad de Barlett, el KMO, la existencia o no de comunalidades mayores a uno y la prueba de bondad de ajuste, son suficientes para establecer la existencia o no de factores latentes.

En concreto, a partir los aspectos específicos considerados para cada factor, no existe evidencia empírica para afirmar la existencia de factores latentes ocultos partiendo de los grupos de variables establecidos previamente. En consecuencia, se considera adecuado que los grupos de actividades sean estudiados a partir de las variables específicas, con el propósito de interpretar y describir las actividades en I+D+I de las empresas de alta y media-alta tecnología del sector industrial manufacturero.

4.4 ANÁLISIS DE REGRESIÓN LOGÍSTICA BINARIA MULTIVARIANTE

El objetivo central de aplicar este modelo de regresión logística es identificar los aspectos (variables) que mejor permitan clasificar en Alta-Media-Alta Tecnologías y Otros Sectores de Manufactura. Se parte del conjunto de 1540

empresas clasificadas previamente en función del sector productivo al que pertenecen, 439 empresas en el grupo AMAT y 1101 empresas en el grupo OSMA.

De acuerdo con Pardo-Merino & Ruiz (2002), las fases para la elaboración de un modelo de clasificación mediante regresión logística binaria multivariante son la selección de las variables apoyada por criterios estadísticos, estimación de los coeficientes de las variables seleccionadas, elaboración del modelo y análisis de resultados.

4.4.1 Selección de variables para la clasificación de empresas a partir de actividades de I+D+I requeridas en la regresión logística binaria multivariada.

Se identifica como variable dependiente a la variable de clasificación de las empresas de alta y media-alta tecnología y otros sectores. Esta variable estará identificada como {VAR_CLAS_AMAT_OSMA} y presenta dos categorías AMAT (empresas de alta y media-alta tecnología) y OSMA (empresas de otros sectores industriales de manufactura).

Se parte de un conjunto inicial de variables consideradas independientes. El conjunto de variables de partida está compuesto por 101 variables de índole numérica y categórica, dado que una de las ventajas del análisis mediante regresión logística es que las variables independientes pueden ser de ambos tipos y se pueden analizar conservando sus unidades de medición originales. La Tabla 4.20 presenta las variables, incluyendo una descripción y el tipo de variable.

Variable	Descripción de la variable	Tipo de variable
CCLIE	CONCENTRACIÓN DE CLIENTES	Numérica
COMPE	COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	Numérica
COMPE1	COMERCIALIZACIÓN DE PRODUCTOS EXTRANJEROS	Categórica
COMPN	COMERCIALIZACIÓN PRODUCTOS NACIONALES	Numérica
COMPN1	COMERCIALIZACIÓN DE PRODUCTOS NACIONALES	Categórica
COMSV	COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	Numérica
DMERPN	DINAMISMO DE LOS MERCADOS	Categórica
EXPORT	EXPORTACIÓN	Categórica
EXPTEC	EXPORTACIÓN DE TECNOLOGÍA	Numérica
IDMERP1N	ÍNDICE DE DINAMISMO DE LOS MERCADOS	Categórica
IDMERPN	ÍNDICE DE DINAMISMO DE LOS MERCADOS	Numérica
IGM	ÍNDICE DE GLOBALIZACIÓN MERCADOS	Numérica
IMPORT	IMPORTACIÓN	Categórica
IMPTEC	IMPORTACIÓN DE TECNOLOGÍA	Numérica
NTES	NÚMERO DE ESTABLECIMIENTOS	Numérica

Variable	Descripción de la variable	Tipo de variable
NUMER	NÚMERO DE MERCADOS	Numérica
PM	INTENSIDAD IMPORTADORA	Numérica
PX	PROPENSIÓN EXPORTADORA	Numérica
RXM	RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN	Numérica
STEREL	SALDO TECNOLÓGICO RELATIVO	Numérica
VENTAS	VENTAS	Numérica
VEXPOR	VALOR DE LAS EXPORTACIONES	Numérica
VIMPOR	VALOR DE LAS IMPORTACIONES	Numérica
WEBB2B	VENTAS A EMPRESAS POR INTERNET	Categórica
WEBB2C	VENTAS A CONSUMIDORES FINALES POR INTERNET	Categórica
WEBVEN	INCIDENCIA DE INTERNET SOBRE LAS VENTAS	Categórica
ACT	ACUERDOS DE COOPERACIÓN TECNOLÓGICA	Categórica
ADBEM	ADQUISICIÓN DE BIENES DE EQUIPO PARA MEJORA DE PRODUCTOS	Categórica
AID	ACTIVIDADES DE I+D	Categórica
AUTEC	AUTONOMÍA TECNOLÓGICA	Numérica
AUTEC1	AUTONOMÍA TECNOLÓGICA	Categórica
CTCL	COLABORACIÓN TECNOLÓGICA CON CLIENTES	Categórica
CTCO	COLABORACIÓN TECNOLÓGICA CON COMPETIDORES	Categórica
CTPR	COLABORACIÓN TECNOLÓGICA CON PROVEEDORES	Categórica
CUCT	COLABORACIÓN CON UNIVERSIDADES Y/O C. TECNOLÓGICOS	Categórica
DCT	DIRECCIÓN O COMITÉ DE TECNOLOGÍA	Categórica
EPCT	EVALUACIÓN PERSPECTIVAS CAMBIO TECNOLÓGICO	Categórica
ETAE	EVALUACIÓN DE TECNOL. ALTERNATIVAS	Categórica
GEID	GASTOS EXTERNOS I+D	Numérica
GEIDE	GASTOS EXTERNOS DE I+D A EMPRESAS	Numérica
GIID	GASTOS INTERNOS I+D	Numérica
GPIDSV	GASTOS DE PUBLICIDAD E I+D SOBRE VENTAS	Numérica
GTID	GASTOS TOTALES I+D	Numérica
IBESV	INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	Numérica
IBESV1	INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	Categórica
IDV	GASTOS DE I+D SOBRE VENTAS	Numérica
IDV1	GASTOS DE I+D SOBRE VENTAS	Categórica
IRI	INDICADORES DE RESULTADOS DE LA INNOVACIÓN	Categórica
MODUT	MODELOS DE UTILIDAD	Numérica
PAI	PLAN DE ACTIVIDADES DE INNOVACIÓN	Categórica
PATESP	PATENTES REGISTRADAS EN ESPAÑA	Numérica
PATEXT	PATENTES REGISTRADAS EN EL EXTRANJERO	Numérica
PEIT	PARTICIPACIÓN EN EMPRESAS CON INNOVACIÓN TECNOL.	Categórica
PIUE	PARTICIPACIÓN PROGRAMA INVESTIGACIÓN DE UE	Categórica
SPR	SUBCONTRATACIÓN PRODUCTOS O COMPONENTES	Categórica
UAIT	UTILIZACIÓN DE ASESORES PARA INFORMARSE SOBRE TECNOL.	Categórica
APLIFN	APLICAR INCENTIVOS FISCALES I+D	Categórica
BSFE	BUSCAR SIN ÉXITO FINANCIAC. EXTERNA DE LA INNOVAC.	Categórica
DEDID	DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES	Numérica
DEDIDG	DEDUCCIONES DE I+D SOBRE GASTOS I+D	Numérica
DEDIT	DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES	Numérica
DEDITG	DEDUCCIONES DE INNOVACIÓN TECNOLÓGICA SOBRE GASTOS I+D	Numérica
DEDTOGN	DEDUC. TOTALES I+D E INNOV. SOBRE GASTOS TOTALES I+D	Numérica
DEDTON	DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES	Numérica
FICS	FINANCIACIÓN DE LA INNOVACIÓN CON CRED. SUBVENC.	Categórica

Variable	Descripción de la variable	Tipo de variable
FPIDCA	FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS	Numérica
FPIDES	FINANCIACIÓN PÚBLICA I+D ESTADO	Numérica
FPIDOT	FINANCIACIÓN PÚBLICA I+D OTROS	Numérica
FPIDTO	FINANCIACIÓN PÚBLICA I+D TOTAL	Numérica
ICO	INNOVACIONES DE COMERCIALIZACIÓN	Categórica
ICOCAN	INNOVACIONES EN CANALES DE VENTA	Categórica
ICODIS	INNOVACIONES DE COMERCIALIZACIÓN POR DISEÑO	Categórica
ICOFIJ	INNOVACIONES EN LA FIJACIÓN DE LOS PRECIOS	Categórica
ICOPRO	INNOVACIONES DE COMERCIALIZACIÓN POR PROMOCIÓN	Categórica
IMO	INNOVACIONES DE MÉTODOS ORGANIZATIVOS	Categórica
IMOGE	INNOVACIONES EN LA GESTIÓN DE LAS RELACIONES EXTERNAS	Categórica
IMOPE	INNOVACIONES EN LA ORGANIZACIÓN DEL TRABAJO	Categórica
IP	OBTENCIÓN INNOVACIONES DE PRODUCTO	Categórica
IPNC	INNOVACIÓN PRODUCTO POR NUEVOS COMPONENTES	Categórica
IPNF	INNOVACIÓN PRODUCTO POR NUEVAS FUNCIONES	Categórica
IPNM	INNOVACIÓN PRODUCTO POR NUEVOS MATERIALES	Categórica
IPR	OBTENCIÓN INNOVACIONES DE PROCESO	Categórica
IPRME	INNOVACIONES DE PROCESO POR NUEVOS EQUIPOS	Categórica
IPRPI	INNOVACIONES DE PROCESO POR PROGRAMAS INFORMÁTICOS	Categórica
IPRTM	INNOVACIONES DE PROCESO POR NUEVAS TÉCNICAS	Categórica
NIP	NÚMERO DE INNOVACIONES DE PRODUCTO	Numérica
GEFI	GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS	Numérica
GEFIF	GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA	Numérica
GEFIFN	GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR	Numérica
GEFIN	GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	Numérica
GEFIT	GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN	Numérica
GEFITC	GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL	Numérica
GEFITN	GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR	Numérica
GEFT	GASTOS EXTERNOS TOTALES EN FORMACIÓN	Numérica
GEFTC	GASTOS TOTALES EXTERNOS EN FORMACIÓN SOBRE COSTES DE PERSONAL	Numérica
GEFTN	GASTOS TOTALES EXTERNOS EN FORMACIÓN POR TRABAJADOR	Numérica
GEFVM	GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MÁRKETING	Numérica
GEFVMN	GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MÁRKETING POR TRABAJADOR	Numérica
IILR	INCORPOR. DE INGENIEROS Y/O LICENCIADOS RECIENTES	Categórica
REEID	RECLUTAR PERS. EXPERIENCIA EMPRESARIAL I+D	Categórica
REPID	RECLUTAR PERS. EXPERIENCIA PROF. S.P. I+D	Categórica

Tabla 4.20. Definición de variables iniciales seleccionadas a incluir en el análisis del modelo de regresión logística binaria multivariada. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

En el Anexo 6.1 se muestra una descripción de las variables y las unidades de medidas. Después de identificar las variables de partida se requiere identificar mediante técnicas estadísticas el conjunto final de variables presentes en el

modelo de clasificación, y en El Anexo 6.4 se encuentra disponible el resumen de valores perdidos de las variables seleccionadas para participar en el modelo final de clasificación.

4.4.2 Modelo de clasificación de empresas de alta y media alta tecnología a partir de actividades en I+D+I: modelo de regresión logística binaria multivariada.

Para la selección de variables y construcción de un modelo de clasificación mediante la técnica de regresión logística binaria multivariada, se recurre al método de selección por pasos, el cual utiliza criterios estadísticos para identificar las variables que son significativas y dejar por fuera las que no lo son, para ello se cuenta con unos criterios de decisión.

De acuerdo con Pardo Merino & Ruíz Díaz (2002), se plantean los siguientes criterios de decisión para la selección de las variables en actividades de I+D+I y construcción del modelo de clasificación de las empresas de alta y media alta tecnología:

- *Adelante condicional*. Método de selección por pasos hacia adelante que, partiendo del modelo nulo, va incorporando aquellas variables cuyo estadístico de puntuación, siendo significativo, posee la probabilidad asociada más pequeña. Tras incorporar al modelo una nueva variable, todas las variables incluidas hasta el momento son revisadas para determinar si existe alguna que debe ser excluida. El proceso se detiene cuando entre las variables no incluidas en el modelo no queda ninguna cuyo estadístico de puntuación sea significativo.
- *Prueba Ommnibus*. Es una prueba de ajuste global que permite contrastar la hipótesis en la cual se considera que la mejora obtenida en el ajuste es nula. Y va acompañada en este caso de la prueba de *Hosmer Lemeshow*. Este índice es útil para evaluar el ajuste global del modelo, particularmente cuando se dispone de muchas variables independientes o cuando algunas de las variables independientes son continuas. El índice permite contrastar la hipótesis nula de igualdad de distribuciones, es decir, no existen diferencias entre las frecuencias observadas y las esperadas. En ambos casos no se acepta la hipótesis nula cuando los valores de significación (Sig.) son iguales o inferiores a 0,05.

- *Porcentaje global de clasificación*. Este es el valor global de acierto en la clasificación, está presente en la matriz de confusión (o matriz de clasificación correcta elaborada a partir del modelo de regresión logística), la cual contiene además de los casos correctamente clasificados por el modelo, los porcentajes de acierto en cada una de las categorías de la variable de clasificación. La matriz se calcula considerando un punto de corte que se utiliza para para clasificar los casos en una categoría u otra de la variable dependiente.
- Finalmente, la presencia de variables o de la constante del modelo, que sean no significativas en el modelo final. Una vez la técnica se detiene producto de no contar con variables independiente significativas para introducir en el modelo, es posible que alguna de las variables finales en presencia de las demás no sea significativa, por lo tanto, el modelo no es adecuado para clasificar. En estos casos se procede a excluir del análisis las variables no significativas, o lo que es lo mismo, a seleccionar las variables restantes como candidatas a incluir en un nuevo modelo. De esta manera se itera hasta obtener un modelo de clasificación que cumpla con todos los criterios de decisión y que sea interpretable.

Se establecen dos fases de análisis, en la primera fase se realiza por grupo de actividades en I+D+I con el propósito de seleccionar un conjunto de variables (o rasgos) que puedan ser incluidos en un modelo de clasificación y/o detectar algún modelo válido de clasificación, y la segunda se realiza con el fin de elaborar el modelo de clasificación de las empresas de alta y media-alta tecnología, a partir del conjunto de variables sugeridas en la primera fase.

Se presenta en la Tabla 4.21 el resumen correspondiente a la primera fase, identificando el modelo, las variables de partida, el valor de las pruebas de hipótesis, el porcentaje global de clasificación (considerando un punto de corte de 0,5, el cual solo toma valores entre cero y uno), si existe o no la presencia de variables o de la constante en el modelo final que no son significativas, si el modelo es válido o no para clasificar, de no ser válido el modelo, se presenta un subconjunto de variables candidatas a incluir en el nuevo modelo de clasificación.

	Variables de partida	Prueba Ommnibus / Prueba de Hosmer y Lemeshow	Porcentaje global de clasificación. Punto de corte 0,5	Presencia de Variables o cte en el modelo final no significativas. Sig.>0,05 / Modelo válido	Variables significativas en el modelo y candidatas a incluir en un nuevo modelo
M1	CCLIE, COMPE, COMPE1, COMPN, COMPN1, COMSV, DMERPN, EXPORT, EXPTEC, IMPORT, IMPTEC, NTES, NUMER, PM, PX, RXM, VENTAS, VEXPOR, VIMPOR, WEBB2B, WEBB2C, WEBVEN (22 var)	0,000 / 0, 797	76,8%	Si / Modelo no válido	CCLIE, COMPE, IMPORT, PX
M2	ACT, ADBEM, AID, CTCL, CTCO, CTPR, CUCT, DCT, EPCT, ETAE, GEID, GEIDE, GIID, GTID, IBESV, IDV, IRI, MODUT, PAI, PATESP, PATEXT, PEIT, PIUE, SPR, UAIT (25 var)	0,000 / 0,681	77,1%	Si / Modelo no válido	CTCL, DCT, GIID, IDV, PAI, SPR
M3	APLIFN, BSFE, DEDID, DEDIT, DEDTON, FICS, FPIDCA, FPIDES, FPIDOT, FPIDTO	0,000 / 0,104	73,4%	Si / Modelo no válido	APLIFN, BSFE, FPIDCA, FPIDES
M4	ICO, ICOCAN, ICODIS, ICOFIJ, ICOPRO, IMO, IMOGE, IMOPE, IP, IPNC, IPNF, IPNM, IPR, IPRME, IPRPI, IPRTM, NIP	0,000 / 0,076	73,1%	Si / Modelo no válido	ICOCAN, IMO, IPNC, IPRTM
M5	GEFI, GEFIF, GEFIFN, GEFIN, GEFIT, GEFITC, GEFITN, GEFT, GEFTC, GEFTN, GEFVM, GEFVMN, IILR, REEID, REPID	0,000 / 0,000	(75,8%)	Si / Modelo no válido	IIRL

Tabla 4.21. Resumen de los modelos de la primera fase de selección de variables. Elaboración propia. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

La totalidad de los modelos de clasificación evaluados (M_1-M_5), elaborados a partir de los grupos de variables presentes en las actividades de comercialización (M_1), actividades innovadoras y de apoyo tecnológico (M_2), incentivos y financiación de I+D+I (M_3), las actividades de innovación (M_4), y talento humano (M_5), reportan variables no significativas, razón por la cual no se pueden considerar válidos para adelantar tareas de clasificación de la empresa de alta y media alta tecnología. Sin embargo, aquellas otras variables que son significativas al interior de cada modelo se consideran candidatas a incluir en un nuevo modelo general. Estas variables son:

{CCLIE, COMPE, IMPORT, PX, CTCL, DCT, GIID, IDV, PAI, SPR, APLIFN, BSFE, FPIDCA, FPIDES, ICOCAN, IMO, IPNC, IPRTM, IIRL}

Se procede a evaluar un grupo nuevo de modelos a partir de las anteriores variables candidatas con el propósito de identificar un conjunto final de variables que permitan elaborar un modelo de clasificación de las empresas de alta y media-alta tecnología del sector manufacturero. Se realizan varias iteraciones, eliminando en cada caso del modelo en evaluación, aquellas variables que en

presencia de las demás no son significativas. Esta ronda de iteraciones provee un nuevo conjunto de variables que se presentan en la Tabla 4.22.

	Variables de partida	**Prueba Ommnibus / Prueba de Hosmer y Lemeshow**	**Porcentaje global de clasificación. Punto de corte 0,5**	**Presencia de Variables o cte en el modelo final no significativas. Sig.>0,05 / Modelo válido**	**Variables significativas en el modelo seleccionado**
M6	IMPORT, CTCL, DCT, SPR, APLIFN, BSFE, FPIDES, IMO, IPRTM, IILR IDV	0,000/ 0,776	76,9	No / Modelo válido para su análisis	IMPORT, CTCL, DCT, IIRL, IDV

Tabla 4.22. Resumen de los criterios de decisión para la selección del modelo de regresión logística. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012).

El conjunto de las variables independientes de partida para la construcción del modelo de clasificación de empresas de alta y media-alta tecnologías y otros sectores de manufactura, son: La realización de importaciones (IMPORT), la colaboración tecnológica con clientes (CTCL), mantener una dirección o comité de tecnología (DCT), la subcontratación de productos terminados o componentes (SPR), la aplicación de incentivos fiscales para I+D+I (APLIFN), la búsqueda sin éxito de financiación externa de la innovación (BSFE), el monto de los recursos recibidos de la administración central para la I+D (FPIDES), la introducción de innovación en métodos organizativos (IMO), la introducción de innovaciones de proceso por nuevas técnicas (IPRTM), la incorporación de ingenieros o licenciados recientes (IIRL) y el porcentaje que representan los gastos totales en I+D sobre el volumen de ventas (IDV).

Identificadas las variables, se requiere analizarlas a la luz de un contraste de hipótesis exclusivo para cada variable independiente, en este caso con la hipótesis nula de que el efecto de la variable es nulo. Se presenta la síntesis de la prueba en la Figura 4.10.

			Puntuación	gl	Sig.
Paso 0	Variables	IMPORT(1)	122,143	1	,000
		CTCL(1)	129,130	1	,000
		DCT(1)	154,254	1	,000
		SPR(1)	25,709	1	,000
		APLIFN(1)	94,554	1	,000
		BSFE(1)	5,460	1	,019
		FPIDES	13,619	1	,000
		IMO(1)	27,363	1	,000
		IPRTM(1)	41,923	1	,000
		IILR(1)	116,971	1	,000
		IDV	122,971	1	,000
	Estadísticos globales		300,170	11	,000

Figura 4.10. Variables a incluir en el análisis del modelo de regresión logística. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Como el nivel crítico asociado al contraste (Sig.) de cada variable es menor que 0,05 se rechaza la hipótesis nula y se concluye que cada variable independiente por separado contribuye significativamente a explicar el comportamiento de la variable de clasificación de alta y media-alta tecnología, y otros sectores industriales de manufactura.

El conjunto de criterios a evaluar está compuesto por el ajuste global del modelo mediante la prueba *Omnibus* que observa en la Figura 4.11, y el ajuste global del modelo mediante la prueba de *Hosmer* y *Lemeshow* que se detalla en la Figura 4.12 se resalta que el modelo realiza una iteración por variable incluida, en total se realizaron cinco iteraciones o pasos.

		Chi cuadrado	gl	Sig.
Paso 5.	Modelo	306,659	5	,000

Figura 4.11. Pruebas Omnibus. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Como se observa, en la prueba *Omnibus* el valor del Chi cuadrado (o Ji-cuadrado) es de 306,659 y tiene un nivel crítico Sig. (0,000 < 0,05%). Por tanto, se rechaza la hipótesis nula de que el ajuste global del modelo es nulo y se concluye

que al ingresar las variables independientes al modelo se consigue un ajuste significativo.

Paso	Chi cuadrado	gl	Sig.
5	2,500	5	,776

Figura 4.12. Prueba de *Hosmer* y *Lemeshow*. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

En la prueba de *Hosmer* y *Lemeshow* el estadístico Chi cuadrado (o Ji-cuadrado) toma el valor 2,500 y tiene asociado un nivel crítico (Sig.) de 0,776 mayor que 0,05. Por tanto, no se rechaza la hipótesis nula de igualdad de distribuciones y, en consecuencia, se puede asumir que no existen diferencias significativas entre las frecuencias observadas y las esperadas.

El criterio de evaluación consiste en detallar la matriz de confusión presente Figura 4.13 a partir del valor de punto de corte que mejora los porcentajes de aciertos en las dos categorías de la variable de clasificación.

Observado			Pronosticado		
			VAR_CLAS_AMAT_OSMA		Porcentaje correcto
			OSMA	AMAT	
Paso 5	VAR_CLAS_AMAT_OSMA	OSMA	848	248	77,4
		AMAT	171	268	61,0
	Porcentaje global				72,7

a. El valor de corte es ,285

Figura 4.13. Matriz de confusión. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Mediante iteraciones sucesivas se establece que el punto de corte indicado para el análisis es 0,285; a partir de este valor se distribuye de mejor manera los porcentajes de acierto entre las dos categorías de la variable. Los casos cuya probabilidad pronosticada es mayor 0,285 son clasificados como alta y media-alta tecnología, los casos cuya probabilidad pronosticada es menor que los puntos de corte son clasificados en la categoría de otros sectores industriales de manufactura.

El modelo de regresión logística clasifica, a partir del conjunto de variables finalmente incluidas, de manera global el 72,7% del total de empresas (1540

disponibles), 61% de las empresas de alta y media alta tecnología (268 empresas de 439 disponibles en esta categoría) y clasifica de manera igualmente correcta 77,4 % de las empresas de otros sectores industriales de manufactura (848 empresas de 1101disponibles).

Finalmente, el conjunto de las variables presentes en la Tabla 4.23, los estadísticos y sus coeficientes permiten describir el modelo de regresión logística seleccionado para clasificar las empresas de alta y media-alta tecnología y otros sectores industriales de manufactura.

Variable	B	E.T.	Wald	gl	Sig.	Exp(β)
Constante	-2,410	,162	221,766	1	,000	,090
IMPORT	1,220	,180	45,879	1	,000	3,386
IILR	,764	,162	22,187	1	,000	2,147
CTLC	,526	,177	8,852	1	,003	1,693
DCT	,385	,167	5,294	1	,021	1,469
IDV	,198	,038	26,782	1	,000	1,219

Tabla 4.23. Variables y coeficientes del modelo de regresión logística de clasificación. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Como se observa en la Tabla 4.23, cada una de las variables en presencia de la constante y las demás variables son significativas para el propósito de clasificar las empresas de alta y media-alta tecnología, y otros sectores industriales de manufactura.

Las variables presentes en el modelo corresponden a la realización o de importaciones (IMPORT), la incorporación de ingenieros o licenciados recientes (IIRL), la colaboración tecnológica con clientes (CTCL), a si la empresa mantiene una dirección o comité de tecnología (DCT) y al porcentaje que representan los gastos totales en I+D sobre el volumen de ventas (IDV). Todas estas variables se consideran de interés al ser rasgos las actividades de I+D+I que permiten clasificar entre las empresas de alta y media-alta tecnología y otros sectores industriales de manufactura.

De acuerdo al modelo general para regresión logística presentada en la subsección 3.3.2.3 de la presente tesis, el modelo de regresión logística binaria multivariante identificado para clasificar las empresas de alta y media alta tecnologías es:

$$\Pr(VAR_CLAS_AMAT_OSMA) = \frac{1}{1 + e^{-(-2.410 + 1.22Import + 0.526CTCL + 0.385DCT + 0.764IILR + 0.198IDV)}}$$

Al analizar los signos de los coeficientes que acompañan las diferentes variables, se aprecia que todas tienen signos positivos en el coeficiente β (excepto la constante), lo que significa que en las empresas que realizan importaciones, incorporan ingenieros y licenciados recientemente, realizan colaboración tecnológica con clientes, mantienen una dirección o comité de tecnologías en presencia de mayores gastos totales en I+D en relación al volumen de ventas, se incrementa la probabilidad de que la empresa sea clasificada como de alta o media-alta tecnología.

4.4.3 Conclusiones del análisis de regresión logística binaria multivariada

Para interpretar los resultados obtenidos en el modelo es necesario indicar que, para todas las variables independientes dicotómicas, la categoría de referencia para el análisis es el valor de uno ("1") correspondiente al estado de la variable de "SI", es decir, la actividad representada por la variable, se realiza por parte de las empresas de alta y media-alta tecnología.

En este sentido, a partir de los valores correspondientes a *Exp(β)* o e^{β} se tiene que las empresas importadoras (IMPORT) presentes en la muestra tienen 3.386 veces más probabilidad de ser de alta tecnología y media-alta tecnología que las empresas que no realizan importaciones (es de desatacar que la probabilidad que representa el *Exp(β)* para cada variable es en presencia de valores cero para las demás variables, así en todos los casos).

De igual manera, la probabilidad de que una empresa que incorpora ingenieros o licenciados recientemente (IILR) sea considerada de alta tecnología es 2.147 veces mayor que aquellas que no contratan recientemente a este tipo de talento humano.

Para aquellas empresas que tienen colaboración tecnológica con clientes (CTLC) la posibilidad de ser de alta y media alta-tecnología es 1.693 veces mayor que las que no realizan tal colaboración.

Con respecto a las si la empresa mantiene una dirección o comité de tecnología (DCT) tienen la probabilidad de 1.469 veces superior que aquellas que no mantienen dirección o comité de tecnología alguno.

De acuerdo al porcentaje que representan los gastos totales en I+D sobre el volumen de ventas (IDV) se puede decir, que la probabilidad de que la empresa sea catalogada de alta tecnología es respectivamente, 1.219 veces superior a aquellas que no realizan ningún gasto en I+D

Por cada unidad de incremento en el porcentaje que representan los gastos totales en I+D sobre el volumen de ventas (IDV), se puede decir que la posibilidad de que la empresa sea considerada de alta y media alta-tecnología se incrementa en 1.219 veces.

Al resolver la ecuación del modelo de regresión logística dado anteriormente, se puede decir que cuando las empresas incorporan ingenieros o licenciados recientemente (IIRL), realizan colaboración tecnológica con clientes (CTCL), mantienen una dirección o comité de tecnología (DCT) y realizan mayores gastos totales en I+D en relación a su volumen de ventas (IDV), la probabilidad de que la empresa sea catalogada como de alta tecnología es de 66,44%.

Por otra parte, la probabilidad de que la empresa no sea catalogada como de alta tecnología cuando no son importadoras (IMPORT), no incorporan ingenieros o licenciados recientemente (IIRL), no realizan colaboración tecnológica con clientes (CTCL), no mantienen una dirección o comité de tecnología (DCT) y realizan muy bajos gastos totales en I+D en relación a su volumen de ventas (IDV), es aproximadamente el 9%.

Un aspecto adicional a considerar es el comportamiento de los residuos, para ello se dispone de la Figura 4.14 donde a través de una inspección visual se establece que los residuos se acercan a un comportamiento de distribución normal.

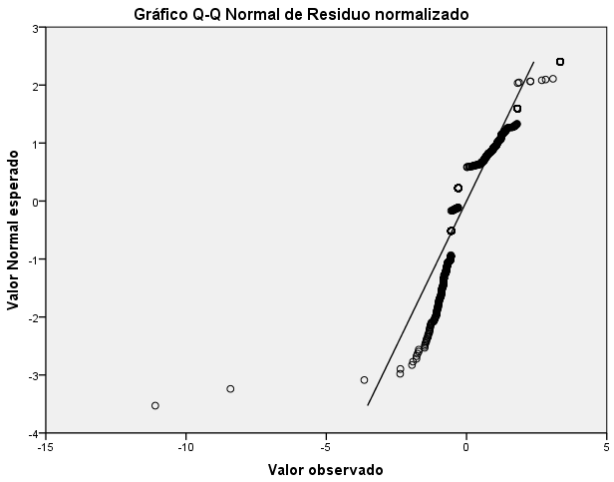


Figura 4.14. Residuos normalizados. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y mediante el software estadístico SPSS.

Para culminar, la regresión logística binaria multivariada ha dado evidencia de su utilidad para adelantar tareas de clasificación, está provista de métodos de análisis y criterios de decisiones útiles para la selección de variables, y posterior elaboración del modelo de clasificación a partir de un conjunto reducido de variables (rasgos) pero estadísticamente significativas. Se considera que los porcentajes globales de clasificación 72,7% son buenos.

4.5 Conclusiones generales del capítulo

Las cuatro técnicas que han sido seleccionadas son: 1. Análisis descriptivo de las variables, 2. La minería de datos mediante la selección de atributos, árboles de decisión, reglas de asociación y técnicas de agrupamiento, 3. El análisis factorial para la detección o no de patrones ocultos, y 4. El análisis de regresión logística para elaborar un modelo de clasificación de las empresas de alta tecnología.

Se concluye que estas técnicas son útiles para realizar el análisis de la actividad en I+D+I en las empresas industriales de alta y media-alta tecnología, dada su idoneidad para explorar, identificar y describir patrones y rasgos, e igualmente útiles para proveer un conjunto de variables relevantes y estadísticamente significativas para la clasificación de empresas de alta y media-alta tecnología.

En concreto, los rasgos y patrones explorados, identificados y descritos más relevantes a partir de las actividades en I+D+I, proveen evidencia en relación a las empresas de alta y media-alta tecnología por una orientación comercial hacia la importación, la colaboración con sus clientes, la disposición a mantener comités de tecnología, a prestar especial atención a los gastos en I+D y a realizar la contratación de ingenieros y licenciados graduados recientemente. Se identifica como aspecto complementario algún patrón que permiten identificar subconjuntos de empresas del sector de alta y media-alta tecnología que aplican incentivos fiscales en I+D. Y son estos mismos en su mayoría, los criterios que finalmente se seleccionan para clasificar de una manera adecuada las empresas de alta y media-alta tecnología y complementaria a la establecida de manera previa contemplando la Clasificación Nacional de Actividades Económicas CNAE-2009.

A partir de los hallazgos realizados con la aplicación de las técnicas utilizadas se presenta una discusión de resultados en el Capítulo 5.

5 CAPÍTULO 5. DISCUSIÓN DE RESULTADOS, CONCLUSIONES, LIMITACIONES Y LÍNEAS DE TRABAJO FUTURAS

El análisis de las actividades de Innovación y de Investigación y Desarrollo que se realizan en el contexto de una empresa resulta de enorme importancia, sobre todo en la búsqueda de la medición del grado en el que dichas actividades, pueden contribuir al logro de sus objetivos, y finalmente al aumento de sus beneficios. En ese sentido, los análisis realizados en esta tesis, y las metodologías utilizadas para ello, pueden resultar de gran interés no solo para empresas de sectores industriales sino para todo tipo de empresas que operan en un contexto como el actual marcado por la globalización y los altos niveles de competencia. Hoy en día, gran parte del éxito de cualquier organización está basado en su capacidad para adaptarse a ambientes cambiantes, para innovar y ofrecer servicios o productos novedosos, para innovar también en sus procesos organizativos internos, en las relaciones con otros agentes implicados. Queda clara, por tanto, la importancia de las actividades que en esta tesis han sido analizadas.

La información relativa a las actividades de innovación es útil por varias razones, permite entre otras, saber si las empresas innovadoras efectúan I+D, si compran conocimiento y tecnología en forma de I+D externa, y si las empresas están implicadas en actividades para modificar parte de su organización (OECD/Eurostat 2007).

En este capítulo se pretende realizar una discusión de los resultados obtenidos con el estudio analítico, extrayendo a partir de ahí las principales conclusiones alineadas con los objetivos planteados en el trabajo. Es importante también identificar aquellas posibles limitaciones del trabajo y en función de ellas establecer líneas futuras de trabajo que permitan complementar los análisis realizados.

5.1 DISCUSIÓN DE RESULTADOS

Para la realización de la discusión de resultados se aborda una síntesis de los hallazgos más relevantes de los estudios realizados y que, de alguna manera, han sido ya comentados cuando se expusieron los resultados de los mismos. En este caso es importante analizar si los resultados obtenidos cumplen con los objetivos planteados en esta tesis, dicho cumplimiento se analiza a continuación:

> *A. Explorar las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología a partir de un conjunto de variables de carácter tecnológico disponibles*.

A partir del análisis descriptivo se ha podido explorar un conjunto de actividades de especial interés para los objetivos de esta tesis, puesto que son realizadas por más del 45% de las empresas de alta y media-alta tecnología, llegando en algún caso a realizarse en el 90% de las empresas analizadas en esos sectores. Entre las actividades en I+D+I que son prácticamente realizadas por todas las empresas de alta y media-alta tecnología, estando presentes entre el 80 al 90% de las empresas de la muestra, se encuentran las actividades de importación y exportación. Parece por tanto clara una orientación internacional de estas empresas. Un segundo grupo de actividades también muy presentes en las empresas tecnológicas, siendo realizadas entre el 45 al 61% de las empresas, está compuesto, en orden descendente, por la realización y/o contratación de actividades de I+D, y el mantenimiento de una dirección o comité de tecnología. Estos hechos ponen de manifiesto la importancia de las actividades de I+D para las empresas de sectores industriales de alta y media-alta tecnología.

> *B. Identificar y describir los patrones y rasgos presentes en las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología mediante la aplicación de diferentes técnicas de minería de datos como selección de atributos, árboles de decisión, reglas de asociación, clustering y/o análisis factorial exploratorio*.

Para identificar y describir cuáles son los patrones presentes en las actividades de I+D+I en las empresas industriales, se resumen a continuación los principales hallazgos obtenidos en la ejecución de tareas de la minería de datos y

en la realización del análisis factorial exploratorio para la detección o no de patrones ocultos.

Producto de la elaboración de un ranking de las variables o atributos más importantes a la hora de poder discriminar la pertenencia o no de una empresa al segmento de empresas tecnológicas, y gracias también a la utilización de árboles de decisión, se identifican una serie de patrones dentro de estas empresas:

- Un primer patrón, al que responden 190 empresas de la muestra, está caracterizado por la realización de actividades de colaboración tecnológica con clientes y con el mantenimiento de una dirección o comité de tecnología.

- En un segundo patrón, presente en 57 empresas, se encuentran aquellas empresas que si bien, no realizan colaboración tecnológica con clientes, al menos sí que tienen tendencia a incorporar a su personal ingenieros y/o licenciados de graduación reciente, realizan exportaciones y mantienen dirección o comité de tecnología.

- En un tercer patrón, presente en 13 empresas, se encontrarían las empresas que no realizan colaboración tecnológica con clientes, ni mantienen dirección o comité de tecnología, pero al menos sí que incorporan ingenieros y/o licenciados de graduación reciente, realizan exportaciones, y aplican incentivos fiscales para I+D+I.

Es de destacar que con los tres patrones señalados se llega a clasificar correctamente el 75,65% del conjunto de empresas presentes en la muestra utilizada.

En relación a la utilización de técnicas de generación de reglas de asociación entre atributos, la verdad es que las reglas obtenidas más que caracterizar a las empresas de sectores tecnológicos lo que permiten es identificar a aquellas empresas que no pertenecen a sectores de alta y media-alta tecnología. Todas las reglas obtenidas tienen altos indicadores de calidad pero vienen a tener una estructura que relaciona las ausencias de determinadas

actividades con la caracterización de la empresa como de sectores de menor tecnología.

Una de las tareas típicas de la minería de datos es la clasificación no supervisada o *clustering*, que busca la identificación automática de grupos de instancias con alta afinidad o similitud entre ellas. En este caso, la aplicación sobre los datos de la muestra del algoritmo de las *K-medias* permite separar las empresas en dos grupos con un porcentaje de acierto del 74,80%. El grupo que el algoritmo clasifica como de empresas de alta y media-alta tecnología viene caracterizado por el acceso a incentivos fiscales, la colaboración tecnológica con clientes, la existencia de una dirección o comité tecnológico y tener unos gastos totales en I+D entre los valores más altos de la muestra.

Para establecer la existencia o no de patrones ocultos se recurrió a la utilización de técnicas de análisis factorial exploratorio. A partir de las pruebas de hipótesis utilizadas para el contraste de bondad de ajuste, se puede afirmar que no hay un patrón oculto asociado al conjunto de las variables de:

- Gastos externos en formación en idiomas (GEFI); gastos externos en formación en ingeniería y formación técnica (GEFIF); gastos externos en formación en informática y tecnologías de la información (GEFIT); gastos externos totales en formación (GEFT); gastos externos en formación de en ventas y marketing por trabajador (GEFVM).
- Financiación pública de I+D recibida de la administración central (FPIDES), gastos externos en formación en idiomas (GEFI); gastos externos en formación en ingeniería y formación técnica (GEFIF); gastos externos en formación en informática y tecnologías de la información (GEFIT); gastos externos totales en formación (GEFT); gastos externos en formación de en ventas y marketing por trabajador (GEFVM), gastos externos en I+D (GEID), gastos internos en I+D (GIID), valor de la exportaciones (VEXPORT) y valor de las importaciones (VIMPORT).

En los demás casos de análisis de patrones ocultos no es posible, a partir de los criterios de decisión utilizados en el análisis factorial, afirmar o rechazar la existencia de los mismos. La importancia de hallar estos patrones ocultos radica en que, de existir los patrones éstos resumen la información presente al interior del

conjunto de variables involucradas, permitiendo de esta manera contar con un conjunto reducido de variables manteniendo el máximo de información disponible, aspecto deseable cuando se cuenta con un gran número de variables. Tanto en el caso en el que se informa de la no existencia de patrones ocultos presentes en un grupo de variables, como los casos donde no se puede establecer la existencia o no de los mismos, se concluye que el análisis de las actividades en I+D+I a partir subgrupos de variables de tipo cuantitativo, se pueden describir a partir de las variables originales.

C. *Clasificar empresas industriales a partir de los rasgos característicos de la alta y media alta-tecnología y otros sectores de manufactura, mediante la obtención de un modelo de regresión logística binaria multivariada.*

Finalmente, para establecer cuáles son los rasgos relevantes para la clasificación de las empresas de manufactura en sectores de alta y media-alta tecnología frente a las de otros sectores de manufactura se contemplan igualmente los hallazgos obtenidos en la realización de un modelo de clasificación mediante la regresión logística binaria multivariada, el cual presenta un porcentaje global de clasificación de 72,7%.

De acuerdo al modelo, la probabilidad de que una empresa sea clasificada como de alta o media-alta tecnología viene dada por la siguiente expresión:

$$\Pr(VAR_CLAS_AMAT_OSMA) = \frac{1}{1 + e^{-(-2.410 + 1.22Import + 0.526CTCL + 0.385DCT + 0.764IILR + 0.198IDV)}}$$

Como puede observarse, en este modelo influye positivamente para clasificar la empresa como tecnológica el hecho de que incorporen ingenieros o licenciados recientes (IIRL), realicen importaciones (IMPORT), realicen colaboración tecnológica con clientes (CTCL), mantengan una dirección o comité de tecnología (DCT) y realicen mayores gastos totales en I+D en relación a su volumen de ventas (IDV). En tales casos la probabilidad de que la empresa sea catalogada como de alta tecnología y media-alta tecnología es de 66.44%.

A la vista de los resultados obtenidos por los diferentes algoritmos parece claro que existe un conjunto de rasgos (variables o atributos) que mejor describen a las empresas de alta y media alta tecnología, tal como queda evidenciado por

el ranking de atributos generado y por la participación de dichas variables en los modelos obtenidos con las diferentes técnicas estadísticas y de minería de datos. En general las tasas de error al clasificar las empresas haciendo uso de los valores de las variables seleccionadas son satisfactorias si se tiene en cuenta la dificultad a priori de la tarea de clasificación y la existencia de claras distorsiones al considerar como clasificación correcta la que se deduce de la inclusión de la empresa en un determinado subsector productivo. Debe tenerse en cuenta que empresas a priori clasificadas como de baja tecnología por estar en sectores poco intensivos en tecnología, pueden, sin embargo, tener características propias de empresas de alta tecnología. Obviamente, la situación inversa también puede producirse. Con estas salvedades, y volviendo a las tasas de acierto en la clasificación, éstas fueron similares en las diferentes técnicas exploradas: utilizando árboles de decisión se obtuvo un 75,65% de clasificación correcta de las empresas, mediante las técnicas de *clustering* un 74,80% y con el modelo de regresión logística binaria multivariada un 72,7%.

Una visión global de los resultados de las diferentes técnicas permite identificar una serie de características que parecen ser relevantes para clasificar a las empresas de sectores de alta y media-alta tecnología. Así, los resultados proveen evidencia de la presencia e importancia de las siguientes actividades en I+D+I en dichas empresas:

1. La realización de importaciones.
2. La incorporación de ingenieros o licenciados graduados recientemente.
3. La colaboración tecnológica con clientes.
4. Mantener una dirección o comité de tecnología.
5. El gasto en I+D, representado por el gasto total en I+D sobre el volumen de ventas.
6. Y en menor medida, pero igualmente de interés para un subgrupo de empresas, la aplicación de incentivos fiscales para I+D e Innovación tecnológica.

En resumen, los hallazgos que se destacan en la tesis son el resultado de identificar, explorar, describir patrones y elaborar un modelo que permita clasificar a las empresas a partir de variables o rasgos relevantes presentes en las actividades

en I+D+I en empresas industriales manufactureras de alta y media-alta tecnología. Son hallazgos de carácter exploratorio descriptivo, por lo tanto, su aporte complementa los hallazgos previos de autores que resaltan, en el contexto de la I+D e Innovación en la empresa, la importancia de la colaboración tecnológica con clientes, las importaciones, la contratación de personal, el gasto en I+D, el mantenimiento de comités de tecnología y la aplicación de incentivos fiscales.

Es así que, la colaboración tecnológica con clientes (CTCL) es un rasgo presente y relevante en las actividades en I+D+I en las empresas del sector manufacturero de alta y media-alta tecnología y útil para la clasificación de estas empresas. Este aspecto fue identificado por Tether (2002), quien examina en el Reino Unido la segunda encuesta de la comunidad europea sobre innovación (CIS-2) adelantada en 1997 e identifica como importante para la innovación la colaboración con clientes, también la colaboración con proveedores, competidores, consultores y universidades. De igual manera, Surroca & Sánchez (2006) argumentan que el establecimiento de acuerdos de cooperación tecnológica con clientes, proveedores, universidades y centros tecnológicos tiene un efecto positivo sobre la consecución de innovaciones en producto y/o proceso. En relación a la colaboración específica con los clientes, Belderbos et al. (2004) destacan también la importancia de los clientes líderes para ayudar a definir las innovaciones y, por lo tanto, reducir el riesgo asociado con su introducción en el mercado. Adicionalmente, Jin & Zhang (2009) identifican la necesidad de evaluar los principales factores de riesgo existentes en el proceso de innovación colaborativa entre empresas y clientes. De manera complementaria, (Ryzhkova 2015) plantea, en este caso para las empresas de servicios, que la colaboración a partir de la interacción con los clientes que utilizan métodos en línea tiene un efecto positivo en la producción de innovación de las empresas. Los análisis realizados en esta tesis vienen a reforzar esos planteamientos y de manera muy especial en las empresas industriales de alta y media-alta tecnología.

De igual manera, la contratación de ingenieros y licenciados de graduación reciente (IIRL) es un rasgo relevante en las activdadades en I+D+I en las empresas del sector manufacturero de alta y media-alta tecnología y ha resultado útil para la clasificación de las empresas de este sector. Esta evidencia está perfectamente alineada con estudios y afirmaciones de diferentes autores; por ejemplo, Shipton

et al. (2005) y Shipton et al. (2006) quienes sugieren que los sistemas de gestión de recursos humanos tienen el potencial de promover la innovación organizacional a través de enfoques sofisticados de reclutamiento y selección, inducción, evaluación y capacitación. Asimismo, de Saá Pérez & Díaz Díaz (2007) afirman que la contratación juega un papel importante en el proceso de construcción de conocimiento de las empresas contratantes, pero también puede suponer un freno al proceso innovador debido a que la integración del conocimiento de los recursos humanos internos y externos puede ser problemática. Por otra parte, Curran & Walsworth (2014) plantean que sus resultados sugieren que se puede pagar a los empleados para que innoven, siempre y cuando seleccione los incentivos de compensación adecuados.

Asimismo, la relación del gasto total en I+D sobre ventas (IDV) es otro de los rasgos relevantes en las activdadades en I+D+I en las empresas del sector manufacturero de alta y media-alta tecnología y de nuevo resulta útila para la clasificación de las empresas de alta y media-alta tecnología. Este aspecto está en síntonia con lo expuesto por la OECD/Eurostat (2007), la cual identifica que en los sectores de alta tecnología, la I+D desempeña un papel central en las actividades de innovación. Igualmente complementa a lo propuesto por Stam & Wennberg (2009), quienes identican que la I+D inicial estimula el desarrollo de nuevos productos en las empresas de alta tecnología, y a lo propuesto por Nunes, Serrasqueiro, & Leitão (2012) quienes resaltan la inversión en I+D como aspecto fundamental para el éxito de las actividades de las PYME de alta tecnología. Además, Daunfeldt, Elert, & Johansson (2016) proponen que las empresas de alta tecnología presentan alta intensidad de I+D y son consideradas más innovadoras. Por otra parte, Chen et al. (2016) observan que los gastos de I+D y comercialización tienen un impacto positivo en el rendimiento operativo de la empresa; más específicamente, Ernst, Hoyer, & Rübsaamen (2010) y Homburg et al. (2017) encuentran que la cooperación I+D y ventas es fundamental en etapas de desarrollo del producto.

Otro de los aspectos identificado como clave en la clasificación de las empresas de mayor componente tecnológico es el relacionado con las importaciones (IMPORT). De nuevo este hecho ya ha sido recogido por distintos autores y específicamente para el caso de la importación de tecnología. En este

sentido, García Vega & Huergo (2010) informan que las empresas tecnológicamente más intensivas tienen una mayor probabilidad de importar tecnología; o los hallazgos por Goldberg et al. (2010) que ponen de manifiesto, para el caso de importación de nuevos insumos, que éstos reportan ganancias sustanciales. Igualmente, en esta misma línea Feng, Li, & Swenson (2016) plantean para el caso de empresas manufactureras chinas, que la diseminación internacional de I+D está relacionada con las importaciones y que la mejora de los productos facilitada por la tecnología o la calidad incorporada en los insumos importados ayudó a las empresas chinas a aumentar la escala y la amplitud de su participación en los mercados de exportación.

Igualmente, el mantenimiento de una dirección o comité de tecnología (DCT) es un rasgo relevante en las actividadades en I+D+I en las empresas del sector manufacturero de alta y media-alta tecnología y útil para la clasificación de las empresas de alta y media-alta tecnología presentes en la muestra. Este aspecto está destacado en investigaciones previas asociado al desarrollo de nuevos producto. Así, Schmidt, Sarangee, & Montoya (2009) identifican que en la mayoría de las empresas que usan procesos de desarrollo de nuevos productos centrados en actividades y puntos de revisión, identifican como aspectos claves de estudio los puntos de revision, los criterios de decisión y los tomadores de decisión en los comités de revisión. Adicionalmente, Lechler & Thomas (2015) concluyen que terminar o cerrar un proyecto de desarrollo de nuevos productos es una decisión económica importante y un interesante dilema gerencial, confirman que la adopción de procesos de decisión para la terminación, establecimiento de criterios formales de decisión y la presencia de los comités de decisión tiene influencias positivas en la calidad de la decisión de terminación de los proyectos de desarrollo de nuevos productos.

Por último, la aplicación de inventivos fiscales para la I+D e innovación (APLFIN) es otro de los rasgos relevantes en las activdadades en I+D+I en un subconjunto menor de empresas del sector manufacturero de alta y media-alta tecnología y al que es deseable prestarle especial antencion. Está en sintonía con lo expuesto por OCDE[15]/FECYT (2003), entidad que destaca y meciona una

[15] En OCDE (2015a), en la última versión a la fecha del Manual de Frascati proporciona en el capítulo 13 ("Measurement of government tax relief for R&D") pautas

variedad de incentivos públicos para I+D útiles en el sector empresarial. Este argumento complementa también la importancia que destacan Blanes & Busom (2004) quienes a partir de una muestra de empresas españolas, proveen evidencia de que los programas nacionales y regionales terminan apoyando a diferentes tipos de empresas y que cada agencia puede usar subsidios de I+D con objetivos de política algo diferentes en cada industria y se puede esperar que estas subvenciones tengan diferentes efectos sobre el esfuerzo de I+D privado o sobre la productividad en diferentes industrias y para diferentes agencias. Asimismo, Busom, Corchuelo, & Martínez Ros (2010) destacan que los incentivos fiscales propician la realización de actividades de I+D+I, y las ayudas directas contribuyen a la cooperación con organismos públicos de investigación, con clientes y proveedores. Igualmente se podrían citar los hallazgos de Hud & Hussinger (2015), quienes para el caso de pequeñas y medianas empresas alemanas proponen que las empresas con mayor capacidad de I+D son más propensas a solicitar subsidios de I+D. Por otra parte, estas empresas pueden ser más propensas a recibir los fondos públicos si el gobierno quiere maximizar los beneficios del programa de subsidios.

Considerando la totalidad de los hallazgos y las técnicas de análisis utilizadas, este estudio puede ser de interés para el ámbito académico que esté interesado en identificar las características de las empresas manufactureras de alta y media-alta tecnología, igualmente puede ser de interés para el ámbito empresarial relacionado con la alta y media alta tecnología a la hora de analizar de manera particular sus decisiones y resultados en torno a las actividades de I+D+I de sus empresas.

5.2 CONCLUSIONES

Durante el desarrollo del presente trabajo se ha logrado el cumplimiento del objetivo general de establecer, por medio de técnicas estadísticas multivariadas junto con métodos complementarios de minería de datos, patrones y rasgos

sobre la presentación de informes sobre el apoyo gubernamental a la I+D mediante incentivos fiscales, con miras a ayudar a la producción de indicadores comparables internacionalmente de alivio fiscal para I+D (GTARD).

característicos de las actividades en Innovación e I+D, que permitan la construcción de modelos para pronosticar la pertenencia o clasificación de empresas, en un sector industrial de alta y media-alta tecnología.

Se ha cumplido satisfactoriamente el objetivo específico de explorar las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología a partir de un amplio conjunto de variables de diferentes tipos y relacionadas con diferentes aspectos clave en torno a las actividades evaluadas. Se han podido analizar variables tanto cuantitativas como cualitativas o categóricas, pudiendo descubrir las diferencias más significativas entre las empresas de alta y media-alta tecnología y de otros sectores de manufactura.

En síntesis, el estudio exploratorio descriptivo ha servido para recolectar evidencias preliminares que permiten identificar rasgos característicos de las empresas manufactureras de alta y media-alta tecnología, y que diferencien a estas empresas de las pertenecientes a otros sectores industriales de manufactura en España. La combinación de técnicas estadísticas y de minería de datos utilizadas ha resultado muy efectiva para el logro de los objetivos planteados.

A continuación, se destacan algunos de los aspectos más importantes observados en el estudio.

Se observa que el perfil de actividades de las empresas pertenecientes al segmento industrial de alta y media-alta tecnología y a otros sectores industriales de manufactura sigue un patrón similar; no obstante, las proporciones de empresas que realizan las actividades es sensiblemente mayor en el segmento de empresas de alta y media-alta tecnología. Básicamente el tipo de actividades realizadas en las empresas industriales es muy similar pero la principal diferencia está en la mayor intensidad de las mismas en los sectores de perfil más tecnológico.

Igualmente se cumple el objetivo específico de Identificar y describir los patrones y rasgos presentes en las actividades en I+D+I de las empresas manufactureras de alta y media-alta tecnología mediante la aplicación de diferentes técnicas de minería de datos como selección de atributos, árboles de decisión, reglas de asociación, *clustering* y/o análisis factorial exploratorio. En los resultados de las diferentes técnicas aparecen de manera recurrente una serie de actividades que vienen a reflejar la importancia de las mismas en las empresas de

mayor componente tecnológica. Es frecuente en estas empresas la colaboración tecnológica con sus clientes; son normalmente conscientes de la importancia de mantener una dirección o comité de tecnología, y en relación a su política de personal, parecen contratar con mayor frecuencia personal de alta cualificación como son los titulados superiores de reciente graduación. Otros dos rasgos característicos que han ido apareciendo en los diferentes análisis son la apuesta por mantener operaciones de comercio exterior, realizando importaciones con mayor frecuencia que en otras empresas, y la disposición a utilizar posibles incentivos fiscales a la I+D+I.

Las técnicas de clasificación utilizadas, tanto en los árboles de decisión como en la clasificación no supervisada o la regresión logística, han demostrado ser capaces de clasificar las empresas con márgenes de error aceptables teniendo en cuenta la dificultad de dicha tarea. Además, es de destacar que para la clasificación elaborada se ha requerido un número muy limitado de variables o indicadores relevantes, aspecto que se considera deseable y que facilita la utilización e interpretabilidad de los modelos.

Y para finalizar, se cumple el objetivo de clasificar empresas industriales a partir de los rasgos característicos de la alta y media alta-tecnología y otros sectores de manufactura, mediante la obtención de un modelo de regresión logística binaria multivariada, el cual provee una probabilidad de pertenencia de las empresas al sector de alta y media-alta tecnología.

El modelo de regresión logística binaria multivariante identificado para clasificar las empresas de alta y media alta tecnologías es:

$$\Pr(VAR_CLAS_AMAT_OSMA) = \frac{1}{1 + e^{-(-2.410 + 1.22Import + 0.526CTCL + 0.385DCT + 0.764IILR + 0.198IDV)}}$$

La interpretación del modelo permite identificar una serie de rasgos que inciden positivamente en la probabilidad de que la empresa corresponda a sectores de alto componente tecnológico; en concreto la realización de importaciones, la incorporación de ingenieros o licenciados recientemente, la realización de colaboración tecnológica con clientes, el mantenimiento una dirección o comité de tecnología y unos mayores de gastos totales en I+D en relación a su volumen de ventas.

A partir de los resultados obtenidos en el cumplimiento de los objetivos específicos y del uso de las diferentes técnicas seleccionadas para los análisis, se concluye que los resultados son consistentes, así, los rasgos anteriormente mencionados junto con la aplicación de incentivos fiscales para la innovación e I+D, son rasgos presentes en los patrones más relevantes en las actividades de I+D+I.

Todas las actividades requeridas para obtener innovación y fomentar la I+D son importantes, pero es interesante poder identificar cuáles son aquellas que como sector se realizan en mayor o menor medida, y a partir de ello, poder establecer cursos de acción para acercar a las empresas a los resultados en I+D+I del entorno empresarial y posterior mejora de los mismos, así como para diferenciar las acciones estratégicas en I+D+I de una empresa en particular.

Mediante la realización de esta investigación, se le provee a la comunidad académica evidencia empírica sobre los principales rasgos y patrones presentes en un conjunto de 1540 empresas que pertenecen a los sectores industriales de manufactura de alta y media-alta tecnología, catalogadas como tal a partir de la Clasificación Nacional de Actividades Económicas CNAE-2009. Igualmente se pone a disposición un modelo complementario de clasificación que ofrece como valor agregado el clasificar a las empresas de alta tecnología a partir de una probabilidad estimada, contemplando el conjunto de rasgos ya expuestos.

Además, permiten al empresario manufacturero y otros grupos de interés contar con información adicional en el momento de analizar de manera particular sus decisiones y resultados en torno a las actividades de I+D+I de sus empresas. Los resultados obtenidos vienen a aportar evidencia complementaria frente a los hallazgos reportados en la literatura especializada sobre las actividades en I+D+I.

El estudio se ha realizado tomando como referencia una amplia muestra de empresas manufactureras españolas y contemplando un variado conjunto de variables, el trabajo realizado puede extrapolarse a otros contextos temporales, geográficos, económicos, por subgrupos de sectores industriales de alta y media-alta tecnología, o para otros sectores industriales donde sea relevante analizar las actividades en I+D+I, pero en general es necesario obtener nuevos datos de empresas en cada caso.

5.3 LIMITACIONES DE LA INVESTIGACIÓN

En síntesis, la presente tesis ha servido para recolectar evidencias que permiten identificar rasgos relevantes y patrones característicos en las actividades en I+D+I de las empresas manufactureras españolas de alta y media-alta tecnología. Debe enfatizarse que el estudio se ha se realizado en forma de consolidado para cada uno de los dos segmentos establecidos, sin particularizarse por cada uno de los subsectores de empresas manufactureras que lo conforman. De igual manera para los análisis realizados, se agrupa a los subsectores de alta tecnología y los subsectores de media-alta tecnología, en un único sector denominado "de alta y media-alta tecnología". Es de destacar que se observa cierta disposición para realizar análisis conjuntos para los sectores de alta y media-alta tecnología (Delgado-Verde et al. 2011), uniendo estos dos sectores y consolidando datos estadísticos entre una y otra categoría (INE 2011; INE 2013a; INE 2016; OECD 2005b; OECD 2007b).

Un segundo aspecto que debería destacarse es la dificultad existente para definir que es la alta y la media-alta tecnología (Hatzichronoglou 1997) y la dificultad para establecer umbrales que las diferencien (INE 2005). Al respecto, Christensen (2013) proporciona una cuantificación de los defectos en las clasificaciones de las empresas. Además, explica por qué las clasificaciones de las empresas son imprecisas (la creciente complejidad de la producción, la inercia en los cambios en los sistemas estadísticos y la creciente integración de los productos y servicios de fabricación son algunas de las explicaciones primarias e interrelacionadas de esta falta de precisión). De manera complementaria Jericó & Marsili (2016) detectan la dificultades derivadas de los diferentes niveles de agregación de la actividad. Los resultados obtenidos en esta tesis deben ser analizados, por tanto, teniendo en cuenta esas dificultades.

Finalmente, se desea destacar algún grado de limitación referida a la naturaleza de los datos utilizados en cuanto, año y contexto geográfico específico. Es de resaltar que los datos (micro datos) utilizados para el desarrollo actual de la investigación se centran en periodo específico (2012) y ámbito geográfico particular (España), por tanto, los hallazgos se ven delimitados por estás

condiciones de partida de la investigación. Ampliar estas condiciones de partida se puede considerar para la realización de investigaciones complementarias. También sería importante destacar que para establecer relaciones entre variables que se consideran *inputs* de la innovación (por ejemplo, inversión en I+D, en capacitación de personal, en bienes de equipo,…) y *outputs* de la misma (por ejemplo, innovaciones generadas, patentes,…) debe tenerse en cuenta un cierto periodo entre ambas, es decir, los resultados de *inputs* durante un año se convierten en *outputs* en un periodo posterior y que en ocasiones abarca varios años. Sería por tanto necesario disponer de datos históricos que reflejen los valores de las diferentes variables durante un periodo temporal que permita realizar otro tipo de análisis más complejos.

5.4 LÍNEAS DE INVESTIGACIÓN FUTURAS.

Obviamente, cualquier trabajo de investigación es susceptible de mejora y ampliación para salvar algunas de las limitaciones identificadas. En ese sentido, con el ánimo de continuar con el propósito de identificar, explorar, describir y caracterizar con mayor precisión las empresas de alta y media-alta tecnología, se plantean una serie de líneas de investigación futuras.

En primer lugar, aunque la I+D desempeña un papel crucial en el proceso de innovación, y los gastos en I+D son la medida más utilizada de entrada para evaluar el nivel de innovación de las empresas, otras variables han sido identificados como relevantes en esta tesis, por ejemplo, gastos en formación de personal, la actividad de comercio exterior relacionada con las importaciones, acceso a programas de incentivos públicos, inversiones en bienes de equipo, etc. Como primer objetivo se puede abordar la utilización de la metodología multicriterio (Analytical Hierarchy Process – *AHP)* para realizar una organización jerárquica de dichos criterios y realizar una ponderación de los mismos a la hora de poder medir el nivel de innovación de las empresas. Gracias a esa metodología se podrá obtener un índice único a través del cual realizar análisis comparativos de empresas de diferentes sectores industriales, identificando aquellos sectores más innovadores y comparando los resultados con medidas habituales de la salida

de la innovación, como son número de patentes, el número de procesos y las innovaciones en los productos producidos.

En investigaciones futuras será de interés replicar el análisis en otros sectores industriales de manufactura que no son considerados o clasificados como alta y media-alta tecnología. El enfoque planteado en esta tesis para el análisis de patrones y rasgos relevantes en las actividades en I+D+I en la empresa de alta y media-alta tecnología puede ser aplicado para identificar las particularidades de la baja y media-baja tecnología. Por ejemplo, Heidenreich(2009) plantea el siguiente interrogante ¿Existe un patrón específico de innovación de las industrias de baja y media-baja tecnología ?.

La investigación en sectores tradicionales de baja tecnología puede demostrar que las industrias que no son intensivas en investigación son sorprendentemente innovadoras y juegan un papel esencial para el desarrollo de las economías modernas (Hirsch-Kreinsen 2015a). Los sectores de baja y media-baja tecnología son fundamentales para el bienestar económico si se mide en términos de producción, capital invertido o empleo (Sandven et al., 2005; Robertson, Smith, & von Tunzelmann, 2009). De manera particular para las industrias de baja tecnología en la UE, autores como Hansen & Winther (2011) establecen que estas industrias siguen teniendo una importancia económica considerable en Europa en términos de empleo y valor añadido, especialmente fuera de las principales regiones de crecimiento.

Igualmente, podría aplicarse el análisis de las actividades en I+D+I en empresas de alta y media-alta tecnología u otros sectores industriales de manufactura, considerando conjuntos de microdatos de empresas correspondientes a varios años. También es importante la realización de los análisis en un contexto de crisis económica; en este sentido, la OECD expresa especial interés y preocupación (OECD 2011c), y plantea, en relación a los indicadores de ciencia, tecnología e industria, cómo estos proporcionan información estadística en tiempos de recesión (OECD 2009b) y hace especial énfasis que la I+D en tiempos de crisis o recesión es objeto de recortes.

Igualmente se plantea como línea futura de trabajo la realización de análisis similares en otros contextos geográficos y establecer comparaciones entre

regiones o países. Teniendo en cuenta la estructura del tejido empresarial de cada contexto geográfico, podría resultar de interés restringir los análisis a empresas con determinadas características o particularidades. En concreto, teniendo en cuenta la importancia de las empresas familiares en el sector industrial de los respectivos países, se plantea la realización de estudios sobre la innovación como motor del cambio y del crecimiento en las empresas familiares y el análisis de las actividades de innovación en PYMEs familiares en comparación con otras empresas industriales.

6 ANEXOS

En esta sección se incluye la definición de variables disponibles para el estudio, anexos estadísticos obtenidos durante el análisis factorial exploratorio y la regresión logística binaria multivariada, el Real Decreto 475/2007, el listado de sectores y productos considerados de alta tecnología.

6.1 Definición de variables disponibles para el estudio

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
ACT	ACUERDOS DE COOPERACIÓN TECNOLÓGICA	Variable categorial que indica si la empresa mantuvo acuerdos de cooperación tecnológica (joint venture). Estados: - No - Si.	Categórica	Sin unidades /entero	1.SI 2. NO
ADBEM	ADQUISICIÓN DE BIENES DE EQUIPO PARA MEJORA DE PRODUCTOS	Variable categorial que indica si la empresa ha adquirido bienes de equipo para la mejora de productos. Estados de la variable: - No - Si.	Categórica	Sin unidades /decimal	1. SI 2. NO
AID	ACTIVIDADES DE I+D	Variable categorial que indica si la empresa realizó o contrató actividades de I+D durante el	Categórica	Sin unidades /entero	1. NO REALIZ. NI CONTR. 2. REALIZ NO CONTR. 3. CONTR. NO REALIZ.

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		ejercicio. Estados de la variable: - No realiza ni contrata - Realiza, pero no contrata al exterior - Contrata pero no realiza en la empresa - Realiza y contrata.			4. CONTR. REALIZ.
APLIFN	APLICAR INCENTIVOS FISCALES I+D	Variable categorial que indica si la empresa aplica los incentivos fiscales para I+D e innovación tecnológica. Estados de la variable: - Si - No.	Categórica	Sin unidades /entero	1. SI 2. NO
AUTEC	AUTONOMÍA TECNOLÓGICA	Porcentaje que los gastos totales en I+D realizados por la empresa representan sobre el total de gastos en I+D y las importaciones de tecnología. Toma valor 0 si los gastos en I+D son nulos ó si éstos y las importaciones	Escala	Porcentaje/ decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		en tecnología son 0, en el resto de los casos toma valores positivos.			
AUTEC1	AUTONOMÍA TECNOLÓGICA	Variable categorial que indica los intervalos en que se sitúa el porcentaje que los gastos totales en I+D y las importaciones de tecnología representan sobre el total de ventas. Estados de la variable: - Cero - De 0 a 50% - De 50 a 100% - 100%.	Categórica	Sin unidades /entero	1. CERO % 2. DE 0 A 50 % 3. DE 50 A 100 % 4. 100%
BSFE	BUSCAR SIN ÉXITO FINANCIAC. EXTERNA DE LA INNOVAC.	Variable categorial que indica si la empresa ha buscado sin éxito financiación externa de la innovación. Estados: - No-Si.	Categórica	Sin unidades /entero	1. SI 2. NO
CCLIE	CONCENTRACIÓN DE CLIENTES	Porcentaje que representan sobre el total de Ventas finales de la empresa las realizadas a sus tres	Escala	Porcentaje/entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		principales clientes.			
COMPE	COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	Porcentaje que representa sobre las ventas totales de la empresa la comercialización de productos no fabricados por ella, procedentes del extranjero.	Escala	Porcentaje/entero	Ninguna
COMPE1	COMERCIALIZACIÓN DE PRODUCTOS EXTRANJEROS	Variable categorial que indica los intervalos en que se sitúa el porcentaje que representa sobre las ventas totales de la empresa la comercialización de productos no Fabricados por ella, procedentes del extranjero. Estados de la variable: - Cero - De 0 a 10% - De 10% a 20% - Más de 20%.	Categórica	Sin unidades /entero	1. CERO 2. DE 0 A 10% 3. DE 10 A 20% 4. MÁS DE 20%
COMPN	COMERCIALIZACIÓN PRODUCTOS NACIONALES	Porcentaje que representa sobre las ventas totales de la empresa la comercialización	Escala	Porcentaje/entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		n de productos no fabricados por ella, procedentes del mercado interior.			
COMPN1	COMERCIALIZACIÓN DE PRODUCTOS NACIONALES	Variable categorial que indica los intervalos en que se sitúa el porcentaje que representa sobre las ventas totales de la empresa la comercialización de productos no fabricados por ella, procedentes del mercado interior. Estados de la variable: - Cero - De 0 a 10% - De 10% a 20% - Más de 20%.	Categórica	Sin unidades /entero	1. CERO 2. DE 0 A 10% 3. DE 10 A 20% 4. MÁS DE 20%
COMSV	COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	Porcentaje que representa sobre las ventas totales de la empresa la comercialización de productos no fabricados por ella, procedentes del mercado interior y del extranjero.	Escala	Porcentaje/entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
CTCL	COLABORACIÓN TECNOLÓGICA CON CLIENTES	Variable categorial que indica si la empresa tuvo colaboración tecnológica con clientes. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
CTCO	COLABORACIÓN TECNOLÓGICA CON COMPETIDORES	Variable categorial que indica si la empresa tuvo colaboración tecnológica con competidores. Estados: - No- Si.	Categórica	Sin unidades /entero	1. SI 2. NO
CTPR	COLABORACIÓN TECNOLÓGICA CON PROVEEDORES	Variable categorial que indica si la empresa tuvo colaboración tecnológica con proveedores Estados: - No -Si.	Categórica	Sin unidades /entero	1. SI 2. NO
CUCT	COLABORACIÓN CON UNIVERSIDADES Y/O C. TECNOLÓGICOS	Variable categorial que indica si la empresa colaboró con Universidades y/o Centros Tecnológicos. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
DCT	DIRECCIÓN O COMITÉ DE TECNOLOGÍA	Variable categorial que indica si la empresa mantuvo una dirección o comité de	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		tecnología o I+D. Estados: - No - Si.			
DEDID	DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES	Valor de las deducciones por I+D que ha aplicado en el impuesto de sociedades del año. Expresado en euros.	Escala	Unidades /entero	Ninguna
DEDIDG	DEDUCCIONES DE I+D SOBRE GASTOS I+D	Porcentaje que las deducciones por I+D representan sobre el total de gastos de I+D.	Escala	Porcentaje/decimal	Ninguna
DEDIT	DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES	Valor de las deducciones por innovación tecnológica que ha aplicado en el impuesto de sociedades del año. Expresado en euros.	Escala	Unidades /entero	Ninguna
DEDITG	DEDUCCIONES DE INNOVACIÓN TECNOLÓGICA SOBRE GASTOS I+D	Porcentaje que las deducciones por innovación tecnológica representan sobre el total de gastos de I+D.	Escala	Porcentaje/decimal	Ninguna
DEDTOGN	DEDUC. TOTALES I+D E INNOV. SOBRE GASTOS TOTALES I+D	Porcentaje del total de las deducciones por I+D e innovación tecnológica	Escala	Porcentaje/decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		representan sobre el total de gastos de I+D.			
DEDTON	DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES	Valor total de las deducciones que ha aplicado en el impuesto de sociedades del año por I+D e innovación tecnológica. Expresado en euros.	Escala	Unidades /entero	Ninguna
DMERPN	DINAMISMO DE LOS MERCADOS	Variable categorial que clasifica la empresa según el valor del Índice de dinamismo de los mercados durante el año. Estados de la variable: - Expansivos (índice entre 65-100) - Estables (índice entre 35-65) - Recesivos (índice entre 0-35).	Categórica	Sin unidades /entero	1.EXPANSIVOS 2.ESTABLES 3. RECESIVOS
EPCT	EVALUACIÓN PERSPECTIVAS CAMBIO TECNOLÓGICO	Variable categorial que indica si la empresa evaluó perspectivas de cambio tecnológico. Estados: - No -Si.	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
ETAE	EVALUACIÓN DE TECNOL. ALTERNATIVAS	Variable categorial que indica si la empresa ha evaluado alternativas tecnológicas para la empresa. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
EXPORT	EXPORTACIÓN	Variable categorial que indica si la empresa realizó exportaciones. Estados de la variable: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
EXPTEC	EXPORTACIÓN DE TECNOLOGÍA	Ingresos por licencias y asistencia técnica del extranjero, expresado en miles de euros.	Escala	Miles/ decimal	Ninguna
FICS	FINANCIACIÓN DE LA INNOVACIÓN CON CRED. SUBVENC.	Variable categorial que indica si la empresa financió la innovación con créditos subvencionados . Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
FPIDCA	FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS	Recoge los recursos financieros recibidos de las comunidades	Escala	Miles/ decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		autónomas para la investigación y el desarrollo tecnológico, expresados en miles de euros.			
FPIDES	FINANCIACIÓN PÚBLICA I+D ESTADO	Recoge los recursos financieros recibidos de la administración central para la investigación y el desarrollo tecnológico, expresados en miles de euros.	Escala	Miles/ decimal	Ninguna
FPIDOT	FINANCIACIÓN PÚBLICA I+D OTROS	Recoge los recursos financieros recibidos de otros organismos para la investigación y el desarrollo tecnológico, expresado en miles de euros.	Escala	Miles/ decimal	Ninguna
FPIDTO	FINANCIACIÓN PÚBLICA I+D TOTAL	Recoge el total de los recursos financieros públicos recibidos para la investigación y el desarrollo tecnológico, expresado en miles de euros.	Escala	Miles/ decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
GEFI	GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS	Valor de los gastos externos en la formación de los trabajadores que se realizó en el año en idiomas. Expresado en euros.	Escala	Unidades /entero	Ninguna
GEFIF	GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA	Valor de los gastos externos en la formación de los trabajadores que se realizó en el año en ingeniería y formación técnica. Expresados en euros.	Escala	Unidades /entero	Ninguna
GEFIFN	GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR	Gastos externos en la formación de los trabajadores en ingeniería y formación técnica, sobre la aproximación del personal total medio. Expresado en euros por ocupado.	Escala	Unidades / Decimal	Ninguna
GEFIN	GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	Gastos externos en la formación de los trabajadores en idiomas sobre la	Escala	Unidades / Decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		aproximación del personal total medio. Expresado en euros por ocupado.			
GEFIT	GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN	Valor de los gastos externos en la formación de los trabajadores que se realizó en el año en informática y tecnologías de la información. Expresado en euros.	Escala	Unidades /entero	Ninguna
GEFITC	GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL	Porcentaje que los gastos externos en la formación de los trabajadores en informática y tecnología de la información representan sobre el coste de personal.	Escala	Porcentaje/decimal	Ninguna
GEFITN	GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR	Gastos externos en la formación de los trabajadores en informática y tecnología de la información sobre la aproximación del personal total medio.	Escala	Unidades /decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		Expresado en euros por ocupado.			
GEFT	GASTOS EXTERNOS TOTALES EN FORMACIÓN	Valor total de los gastos externos en la formación de los trabajadores que se realizó en el año. Expresado en euros.	Escala	Unidades /entero	Ninguna
GEFTC	GASTOS TOTALES EXTERNOS EN FORMACIÓN SOBRE COSTES DE PERSONAL	Porcentaje que los gastos totales externos en la formación de los trabajadores representan sobre el coste de personal.	Escala	Porcentaje/decimal	Ninguna
GEFTN	GASTOS TOTALES EXTERNOS EN FORMACIÓN POR TRABAJADOR	Gastos totales externos en la formación de los trabajadores, sobre la aproximación del personal total medio. Expresado en euros por ocupado.	Escala	Unidades /decimal	Ninguna
GEFVM	GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MÁRKETING	Valor de los gastos externos en la formación de los trabajadores que se realizó en el año en ventas y marketing.	Escala	Unidades /entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		Expresados en euros.			
GEFVMN	GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MÁRKETING POR TRABAJADOR	Gastos externos en la formación de los trabajadores en ventas y marketing, sobre la aproximación del personal total medio. Expresado en euros por ocupado.	Escala	Unidades /decimal	Ninguna
GEID	GASTOS EXTERNOS I+D	Total de gastos externos en actividades de I+D, expresados en euros.	Escala	Unidades /decimal	Ninguna
GEIDE	GASTOS EXTERNOS DE I+D A EMPRESAS	Cuenta 620 (PGC). Recoge los gastos de investigación y desarrollo por servicios encargados a otras empresas. En miles de euros.	Escala	Miles/ decimal	Ninguna
GIID	GASTOS INTERNOS I+D	Total de gastos internos en actividades de I+D, expresados en euros.	Escala	Unidades /decimal	Ninguna
GPIDSV	GASTOS DE PUBLICIDAD E I+D SOBRE VENTAS	Porcentaje que el total de los gastos en I+D y los gastos de publicidad,	Escala	Porcentaj e/decim al	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		propaganda y relaciones públicas representan sobre el total de ventas.			
GPV	GASTOS DE PUBLICIDAD SOBRE VENTAS	Porcentaje que los gastos en publicidad, propaganda y relaciones públicas (cuenta 627 (PGC)) representan sobre las ventas.	Escala	Porcentaje/decimal	Ninguna
GTID	GASTOS TOTALES I+D	Total de gastos en actividades de I+D durante el ejercicio, expresados en euros.	Escala	Unidades /decimal	Ninguna
IBESV	INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	Porcentaje que representan las compras y grandes reparaciones de equipos para procesos de información, maquinaria industrial, utillaje e instalaciones técnicas, elementos de transporte y mobiliario, equipo de oficina y otro inmovilizado	Escala	Porcentaje /decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		material, sobre el total de las ventas.			
IBESV1	INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	Variable categorial que indica los intervalos en que se sitúa el porcentaje que representan las compras y grandes reparaciones de equipos para procesos de información, maquinaria industrial, utillaje e instalaciones técnicas, elementos de transporte y mobiliario, equipo de oficina y otro inmovilizado material, sobre el total de las ventas. Estados de la variable: - De 0 a 25% - De 25 a 50% - De 50 a 75% - Más de 75%.	Categórica	Sin unidades /entero	1. DE 0 A 25% 2. DE 25 A 50% 3. DE 50 A 75% 4. MÁS DE 75%
ICO	INNOVACIONES DE COMERCIALIZACIÓN	Variable categorial que indica si la empresa introdujo innovaciones de	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		comercialización referentes a modificaciones significativas en el diseño o envasado de los productos, nuevos métodos en la utilización de canales de venta, nuevos métodos en la promoción de los productos o nuevos métodos en la fijación de los precios de los productos.			
ICOCAN	INNOVACIONES EN CANALES DE VENTA	Variable categorial que indica si la empresa introdujo innovaciones de comercialización referentes a nuevos métodos en la utilización de canales de venta.	Categórica	Sin unidades /entero	1. SI 2. NO
ICODIS	INNOVACIONES DE COMERCIALIZACIÓN POR DISEÑO	Variable categorial que indica si la empresa introdujo innovaciones de comercialización referentes a modificaciones significativas en	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		el diseño o envasado de los productos.			
ICOFIJ	INNOVACIONES EN LA FIJACIÓN DE LOS PRECIOS	Variable categorial que indica si la empresa introdujo innovaciones de comercialización referentes a nuevos métodos en la fijación de los precios de los productos.	Categórica	Sin unidades /entero	1. SI 2. NO
ICOPRO	INNOVACIONES DE COMERCIALIZACIÓN POR PROMOCIÓN	Variable categorial que indica si la empresa introdujo innovaciones de comercialización referentes a nuevos métodos en la promoción de los productos.	Categórica	Sin unidades /entero	1. SI 2. NO
IDMERP1N	ÍNDICE DE DINAMISMO DE LOS MERCADOS	Variable categorial que agrupa en tres intervalos los valores del índice de dinamismo de los mercados durante el año. El índice se obtiene agregando los productos de las	Categórica	Sin unidades /entero	1. DE 0 A 35 % 2. DE 35 A 65 % 3. DE 65 A 100 %

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		variables: ponderación mercado y Situación de dinamismo del mercado. Estados de la variable: - De 0 a 35% - De 35 a 65% - De 65 a 100%.			
IDMERPN	ÍNDICE DE DINAMISMO DE LOS MERCADOS	Índice agregado de dinamismo de todos los mercados servidos por la empresa durante el año. El índice se obtiene agregando los productos de las variables: Ponderación Mercado y Situación de Dinamismo del Mercado.	Escala	Porcentaje/decimal	Ninguna
IDV	GASTOS DE I+D SOBRE VENTAS	Porcentaje que representan los gastos totales en I+D sobre el volumen de ventas.	Escala	Porcentaje/decimal	Ninguna
IGM	ÍNDICE DE GLOBALIZACIÓN MERCADOS	Porcentaje de los mercados servidos por la empresa cuyo ámbito	Escala	Porcentaje/entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		geográfico es simultáneamente interior y exterior.			
IILR	INCORPOR. DE INGENIEROS Y/O LICENCIADOS RECIENTES	Variable categorial que indica si la empresa ha incorporado ingenieros y/o licenciados de graduación reciente. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
IMO	INNOVACIONES DE MÉTODOS ORGANIZATIVOS	Variable categorial que indica si la empresa introdujo nuevos métodos organizativos, referentes a nuevas prácticas empresariales en la organización del trabajo o nuevos métodos de gestión de las relaciones externas con otras empresas o instituciones públicas.	Categórica	Sin unidades /entero	1. SI 2. NO
IMOGE	INNOVACIONES EN LA GESTIÓN DE LAS RELACIONES EXTERNAS	Variable categorial que indica si la empresa introdujo nuevos	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		métodos organizativos, referentes a nuevos métodos de gestión de las relaciones externas con otras empresas o instituciones públicas.			
IMOPE	INNOVACIONES EN LA ORGANIZACIÓN DEL TRABAJO	Variable categorial que indica si la empresa introdujo nuevos métodos organizativos, referentes a nuevas prácticas empresariales en la organización del trabajo.	Categórica	Sin unidades /entero	1. SI 2. NO
IMPORT	IMPORTACIÓN	Variable categorial que indica si la empresa realizó importaciones. Estados de la variable: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
IMPTEC	IMPORTACIÓN DE TECNOLOGÍA	Pagos por licencias y asistencia técnica del extranjero, expresados en miles de euros.	Escala	Miles/ decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
IP	OBTENCIÓN INNOVACIONES DE PRODUCTO	Variable categorial que indica si la empresa ha obtenido innovaciones de producto durante el ejercicio. Estados de la variable: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
IPNC	INNOVACIÓN PRODUCTO POR NUEVOS COMPONENTES	Variable categorial que indica si la innovación de producto se debe a la incorporación de nuevos componentes o productos intermedios. Estados de la variable: - No innova - No - Si.	Categórica	Sin unidades /entero	1. NO INNOVA 2. NO 3. SI
IPNF	INNOVACIÓN PRODUCTO POR NUEVAS FUNCIONES	Variable categorial que indica si la innovación de producto se debe a que el producto cumple nuevas funciones. Estados de la variable: - No innova - No - Si.	Categórica	Sin unidades /entero	1. NO INNOVA 2. NO 3. SI

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
IPNM	INNOVACIÓN PRODUCTO POR NUEVOS MATERIALES	Variable categorial que indica si la innovación de producto se debe a la incorporación de nuevos materiales. Estados de la variable: - No innova - No - Si.	Categórica	Sin unidades /entero	1. NO INNOVA 2. NO 3. SI
IPR	OBTENCIÓN INNOVACIONES DE PROCESO	Variable categorial que indica si la empresa ha obtenido innovaciones de proceso en el ejercicio. Estados de la variable: - Si - No.	Categórica	Sin unidades /entero	1. SI 2. NO
IPRME	INNOVACIONES DE PROCESO POR NUEVOS EQUIPOS	Variable categorial que indica si la empresa introdujo innovaciones de proceso consistentes en la introducción de nuevas máquinas y equipos.	Categórica	Sin unidades /entero	1. SI 2. NO
IPRPI	INNOVACIONES DE PROCESO POR PROGRAMAS INFORMÁTICOS	Variable categorial que indica si la empresa introdujo	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		innovaciones de proceso consistentes en la introducción de nuevos programas informáticos ligados a los procesos industriales.			
IPRTM	INNOVACIONES DE PROCESO POR NUEVAS TÉCNICAS	Variable categorial que indica si la empresa introdujo innovaciones de proceso consistentes en la introducción de nuevas técnicas y métodos.	Categórica	Sin unidades /entero	1. SI 2. NO
IRI	INDICADORES DE RESULTADOS DE LA INNOVACIÓN	Variable categorial que indica si en la empresa se elaboraron indicadores de resultados de la innovación. Estados: - NO - SI.	Categórica	Sin unidades /entero	1. SI 2. NO
MODUT	MODELOS DE UTILIDAD	Número de modelos de utilidad registrados por la empresa durante el ejercicio.	Categórica	Unidades /entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
NACECLIO	ACTIVIDAD	Código representativo de la actividad principal de la empresa, según una agregación de los códigos 3 dígitos CNAE-09 a 20 sectores manufactureros.	Categórica	Sin unidades /entero	1. Industria cárnica 2. Productos alimenticios y tabaco 3. Bebidas 4. Textiles y confección 5. Cuero y calzado 6. Industria de la madera 7. Industria del papel 8. Artes gráficas 9. Industria química y productos farmaceutic. 10. Productos de caucho y plástico 11. Productos minerales no metálicos 12. Metales férreos y no férreos 13. Productos metálicos 14. Máquinas agrícolas e industriales 15. Productos informáticos, electrónicos y ópticos 16. Maquinaria y

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
					material eléctrico 17. Vehículos de motor 18. Otro material de transporte 19. Industria del mueble 20. Otras industrias manufacture-ras
NIP	NÚMERO DE INNOVACIONES DE PRODUCTO	Número de innovaciones de producto que la empresa ha obtenido durante el ejercicio.	Escala	Unidades / Entero	Ninguna
NTES	NÚMERO DE ESTABLECIMIENTOS	Número total de establecimientos , industriales y no industriales, que posee la empresa.	Escala	Unidades /entero	Ninguna
NUMER	NÚMERO DE MERCADOS	Número de mercados definidos por la empresa.	Escala	Unidades /entero	Ninguna
PAI	PLAN DE ACTIVIDADES DE INNOVACIÓN	Variable categorial que indica si la empresa contó con un plan de actividades de innovación. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
PATESP	PATENTES REGISTRADAS EN ESPAÑA	Número de patentes registradas en España por la empresa durante el ejercicio.	Escala	Unidades /entero	Ninguna
PATEXT	PATENTES REGISTRADAS EN EL EXTRANJERO	Número de patentes registradas en el extranjero por la empresa durante el ejercicio.	Escala	Unidades /entero	Ninguna
PEIT	PARTICIPACIÓN EN EMPRESAS CON INNOVACIÓN TECNOL.	Variable categorial que indica si la empresa participó en empresas que desarrollan innovación tecnológica. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
PIUE	PARTICIPACIÓN PROGRAMA INVESTIGACIÓN DE UE	Variable categorial que indica si la empresa ha participado en algún programa de investigación de la UE. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
PM	INTENSIDAD IMPORTADORA	Porcentaje que las importaciones que realiza la empresa representan	Escala	Porcentaje/decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		sobre el total de ventas.			
PX	PROPENSIÓN EXPORTADORA	Porcentaje que las exportaciones que realiza la empresa representan sobre el total de ventas.	Escala	Porcentaje/decimal	Ninguna
REEID	RECLUTAR PERS. EXPERIENCIA EMPRESARIAL I+D	Variable categorial que indica si la empresa reclutó personal con experiencia empresarial en I+D. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
REPID	RECLUTAR PERS. EXPERIENCIA PROF. S.P. I+D	Variable categorial que indica si la empresa reclutó personal con experiencia profesional en el sistema público de I+D. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
RXM	RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN	Cociente entre las exportaciones y las importaciones que realiza la empresa, siempre que éstas sean positivas. Si no lo son, este	Escala	Sin unidades /decimal	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		indicador toma el valor cero.			
SPR	SUBCONTRATACIÓN PRODUCTOS O COMPONENTES	Variable categorial que indica si la empresa contrató la "fabricación de productos terminados componentes ""a medidas" "para su empresa""." Estados de la variable: - Si -No.	Categórica	Sin unidades /entero	1. SI REALIZA 2. NO REALIZA
STEREL	SALDO TECNOLÓGICO RELATIVO	Cociente de la diferencia entre las Exportaciones y las Importaciones de Tecnología que realiza la empresa y el total de ventas.	Escala	Sin unidades /decimal	Ninguna
TEMPR2	TAMAÑO DE LA EMPRESA (INTERVALOS DE TAMAÑO)	Variable categorial que indica el intervalo de tamaño de la empresa según el personal total ocupado a 31 de diciembre. Estados de la variable: - 200 y menos trabajadores -	Categórica	Sin unidades /entero	1. 200 Y MENOS TRABAJADORES 2. MÁS DE 200 TRABAJADORES

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		Más de 200 trabajadores.			
TEMPRE	TAMAÑO DE LA EMPRESA (INTERVALOS DE TAMAÑO)	Variable categorial que indica el intervalo de tamaño de la empresa según el personal total ocupado a 31 de diciembre. Estados de la variable: - Menos de 20 - De 21 a 50 - De 51 a 100 - De 101 a 200 - De 201 a 500 - Más de 500	Categórica	Sin unidades /entero	1. MENOS DE 20 2. DE 21 A 50 3. DE 51 A 100 4. DE 101 A 200 5. DE 201 A 500 6. MÁS DE 500
UAIT	UTILIZACIÓN DE ASESORES PARA INFORMARSE SOBRE TECNOL.	Variable categorial que indica si la empresa ha utilizado asesores y/o expertos para informarse sobre tecnología. Estados: - No - Si.	Categórica	Sin unidades /entero	1. SI 2. NO
VENTAS	VENTAS	Cuenta 70 (PGC). Recoge las ventas de mercaderías, las ventas de productos transformados (terminados y semi-terminados), la prestación de servicios y otras	Escala	Unidades /entero	Ninguna

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		ventas (envases, embalajes, subproductos y residuos), excluidos rappels y devoluciones de ventas. Unidades: euros.			
VEXPOR	VALOR DE LAS EXPORTACIONES	Variable que recoge el valor de las exportaciones en euros.	Escala	Unidades /entero	Ninguna
VIMPOR	VALOR DE LAS IMPORTACIONES	Variable que recoge el valor de las importaciones en euros.	Escala	Unidades /entero	Ninguna
WEBB2B	VENTAS A EMPRESAS POR INTERNET	Variable categorial que indica la disposición de un sistema de ventas a otras empresas por Internet. Estados de la variable: - No procede - No - Si	Categórica	Sin unidades /entero	1. No procede 2. No 3. Si
WEBB2C	VENTAS A CONSUMIDORES FINALES POR INTERNET	Variable categorial que indica la disposición de un sistema de ventas a consumidores finales por Internet. Estados de la variable: -	Categórica	Sin unidades /entero	1. No procede 2. No 3. Si

Variable	Nombre de la variable	Definición de la variable	Tipo de variable	Unidades /Formato	Estados de la variable
		No procede - No - Si			
WEBVEN	INCIDENCIA DE INTERNET SOBRE LAS VENTAS	Variable categorial que indica la incidencia (directa o indirecta) que la presencia de Internet ha tenido sobre las ventas de la empresa. Estados de la variable: - No procede - Ninguna - Ligera - Fuerte - No es evaluable	Categórica	Sin unidades /entero	0. No procede 1. Ninguna 2. Ligera 3. Fuerte 4. No es evaluable

Tabla 6.1. Definición de variables disponibles para todo el estudio. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS.

6.2 Resumen de los estadísticos descriptivos con las variables tipificadas y disponibles para el análisis factorial exploratorio.

Para adelantar el análisis factorial exploratorio se dispone de las variables tipificadas, en la Tabla 6.2 se presenta el número de casos, los valores mínimos, máximos, la media y la desviación típica.

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. típ.
AUTONOMÍA TECNOLÓGICA	439	-,72	1,41	,53	1,03
CONCENTRACIÓN DE CLIENTES	439	-1,51	1,80	,037	,98
COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	439	-,32	7,85	,27	1,33
COMERCIALIZACIÓN PRODUCTOS NACIONALES	439	-,39	5,94	-,07	,82
COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	439	-,48	5,41	,14	1,12
DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES	439	-,09	22,73	,16	1,73
DEDUCCIONES DE I+D SOBRE GASTOS I+D	270	-,17	2,39	-,03	,29
DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES	439	-,06	39,25	,12	1,98
DEDUCCIONES DE INNOVACIÓN TECNOLÓGICA SOBRE GASTOS I+D	270	-,13	7,02	-,03	,53
DEDUC. TOTALES I+D E INNOV. SOBRE GASTOS TOTALES I+D	270	-,20	2,31	-,044	,32
DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES	439	-,09	23,07	,1749	1,740
EXPORTACIÓN DE TECNOLOGÍA	439	-,03	10,23	,00	,52
FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS	439	-,10	23,88	,16	1,57
FINANCIACIÓN PÚBLICA I+D ESTADO	439	-,08	35,74	,15	1,93
FINANCIACIÓN PÚBLICA I+D OTROS	439	-,09	20,94	,049	1,16
FINANCIACIÓN PÚBLICA I+D TOTAL	439	-,10	34,24	,17	1,88
GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS	439	-,15	14,03	,20	1,19
GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA	439	-,12	33,11	,15	1,730
GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR	439	-,32	6,51	,29	1,14
GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	439	-,38	9,52	,44	1,35
GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN	439	-,068	40,33	,093	1,95
GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL	439	-,15	6,13	-,019	,49
GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR	439	-,14	4,89	,014	,47
GASTOS EXTERNOS TOTALES EN FORMACIÓN	439	-,14	25,85	,14	1,39

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. típ.
GASTOS TOTALES EXTERNOS EN FORMACIÓN SOBRE COSTES DE PERSONAL	439	-,52	5,30	,24	,87
GASTOS TOTALES EXTERNOS EN FORMACIÓN POR TRABAJADOR	439	-,54	8,10	,40	1,16
GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MÁRKETING	439	-,10	29,08	,13	1,94
GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MÁRKETING POR TRABAJADOR	439	-,20	15,78	,18	1,46
GASTOS EXTERNOS I+D	439	-,06	36,98	,15	1,96
GASTOS EXTERNOS DE I+D A EMPRESAS	439	-,04	42,12	,12	2,05
GASTOS INTERNOS I+D	439	-,09	23,19	,15	1,42
GASTOS DE PUBLICIDAD E I+D SOBRE VENTAS	439	-,45	11,34	,33	1,41
GASTOS DE PUBLICIDAD SOBRE VENTAS	439	-,36	11,51	-,022	,88
GASTOS TOTALES I+D	439	-,083	32,55	,16	1,78
INVERSIÓN EN BIENES DE EQUIPO SOBRE VENTAS	439	-,31	29,71	,06	1,53
ÍNDICE DE DINAMISMO DE LOS MERCADOS	439	-1,05	2,19	,09	1,01
GASTOS DE I+D SOBRE VENTAS	439	-,30	15,13	,48	1,67
ÍNDICE DE GLOBALIZACIÓN MERCADOS	439	-,61	1,80	,27	1,11
IMPORTACIÓN DE TECNOLOGÍA	439	-,05	27,55	,13	1,94
MODELOS DE UTILIDAD	439	-,08	33,96	,13	1,87
NÚMERO DE INNOVACIONES DE PRODUCTO	434	-,12	8,01	,05	,62
NÚMERO DE ESTABLECIMIENTOS	439	-,14	4,0	,01	,45
NÚMERO DE MERCADOS	439	-,82	2,89	,19	1,10
PATENTES REGISTRADAS EN ESPAÑA	438	-,11	25,15	,17	1,81
PATENTES REGISTRADAS EN EL EXTRANJERO	439	-,04	41,38	,11	2,05
INTENSIDAD IMPORTADORA	439	-,70	3,96	,48	1,16
PROPENSIÓN EXPORTADORA	439	-,83	2,40	,56	1,06
RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN	439	-,055	9,03	-,007	,44
SALDO TECNOLOGICO RELATIVO	439	-22,23	5,52	-,15	1,49
VENTAS	439	-,21	20,31	,22	1,77
VALOR DE LAS EXPORTACIONES	439	-,14	23,46	,24	1,90
VALOR DE LAS IMPORTACIONES	439	-,14	20,30	,21	1,79

Tabla 6.2. Los estadísticos descriptivos con las variables tipificadas y disponibles para el análisis factorial exploratorio. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS.

6.3 Resumen gráficos Q-Q para el análisis visual de normalidad de las variables

En primera instancia se recurre a un análisis visual basado en gráficos Q-Q para identificar el comportamiento de la variable en relación a una distribución normal, ver Tabla 6.3.

Dado que si se desea utilizar la matriz de correlaciones en el Análisis Factorial Exploratorio se debe utilizar valores tipificados de las variables, se procede entonces, a utilizar las variables ya tipificadas N (0,1) para construir el gráfico Q-Q.

Gráfico Q-Q

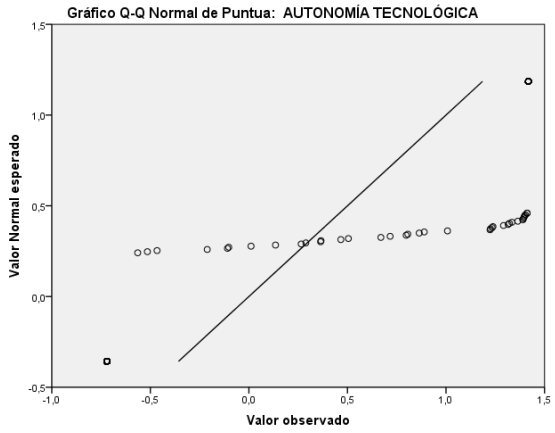
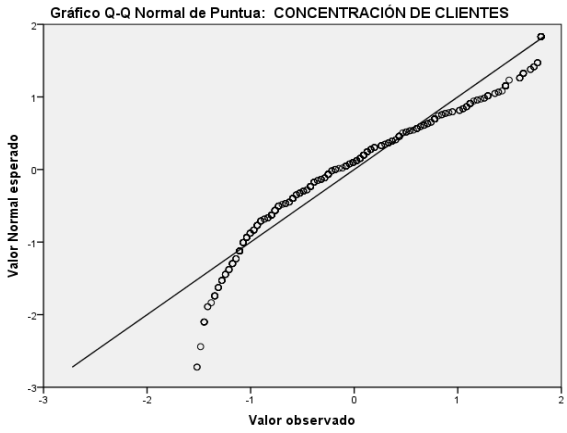
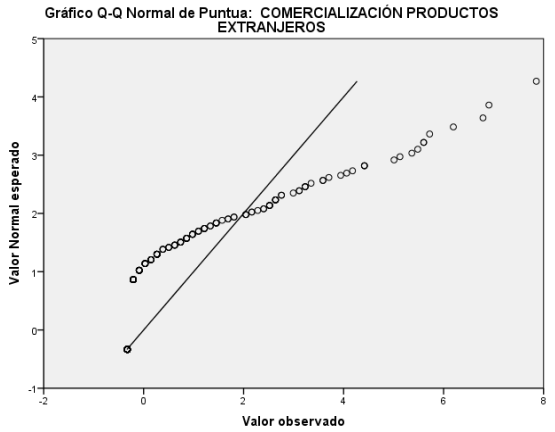
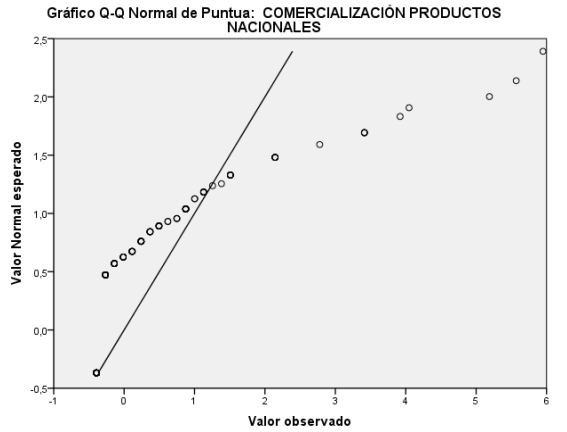
<table>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Se excluye: AUTEC</td><td></td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td></td><td></td></tr>
</table>

Gráfico Q-Q Normal de Puntua: COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS

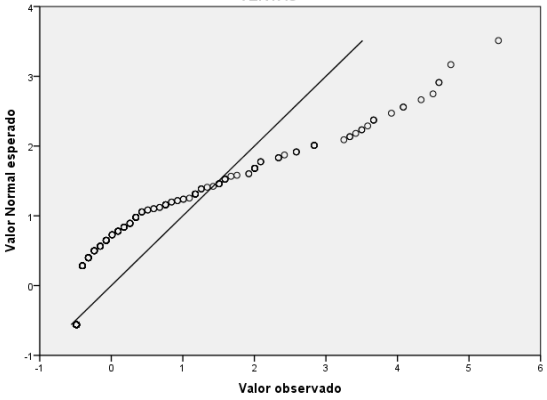


Gráfico Q-Q Normal de Puntua: DEDUCCIONES POR I+D EN EL IMPUESTO DE SOCIEDADES

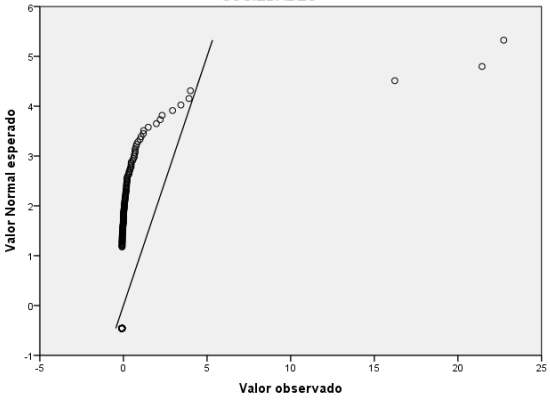


Gráfico Q-Q Normal de Puntua: DEDUCCIONES DE I+D SOBRE GASTOS I+D

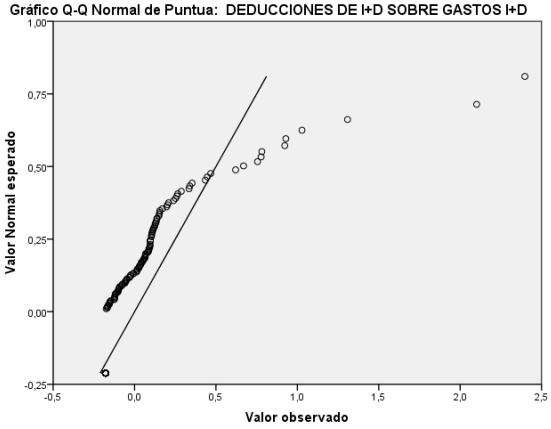


Gráfico Q-Q Normal de Puntua: DEDUCCIONES POR INNOVACIÓN TECNOLÓGICA EN EL IMPUESTO DE SOCIEDADES

Gráfico Q-Q Normal de Puntua: DEDUC. TOTALES I+D E INNOV. SOBRE GASTOS TOTALES I+D

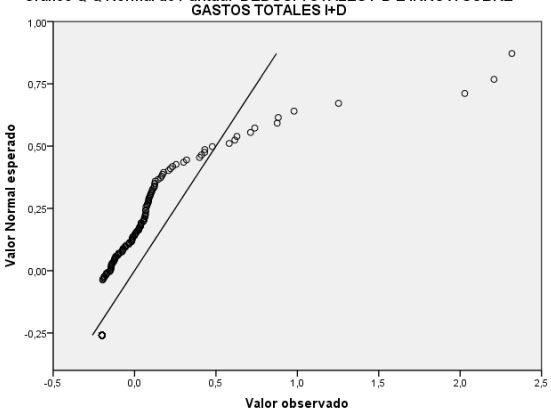
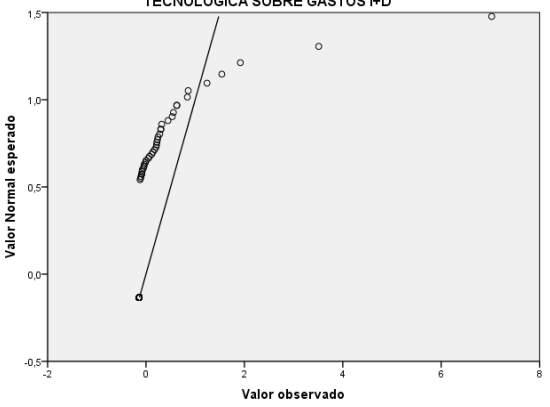


Gráfico Q-Q Normal de Puntua: DEDUCCIONES DE INNOVACIÓN TECNOLÓGICA SOBRE GASTOS I+D



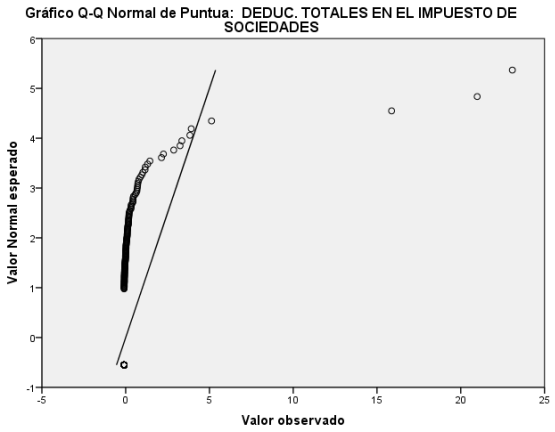


Gráfico Q-Q Normal de Puntua: DEDUC. TOTALES EN EL IMPUESTO DE SOCIEDADES
Valor Normal esperado
Valor observado

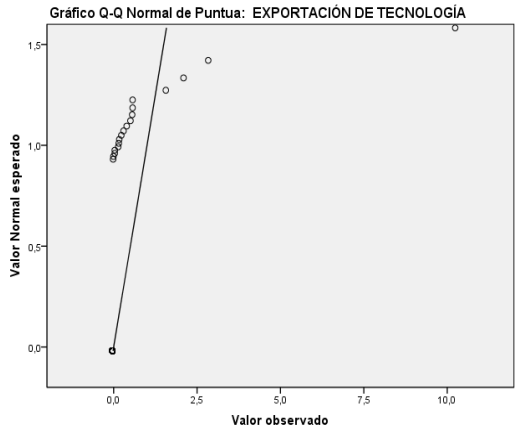


Gráfico Q-Q Normal de Puntua: EXPORTACIÓN DE TECNOLOGÍA
Valor Normal esperado
Valor observado

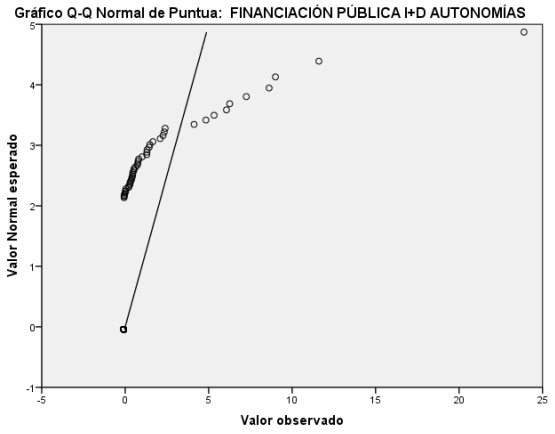


Gráfico Q-Q Normal de Puntua: FINANCIACIÓN PÚBLICA I+D AUTONOMÍAS
Valor Normal esperado
Valor observado

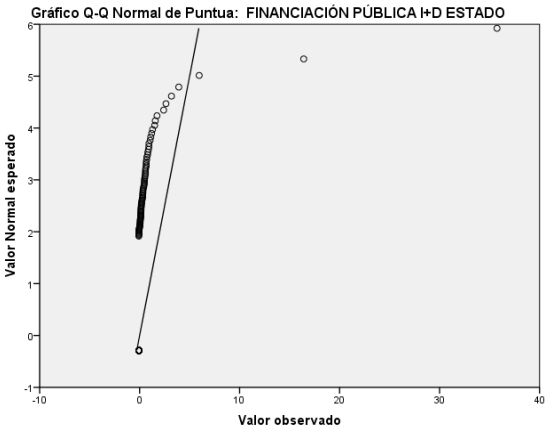


Gráfico Q-Q Normal de Puntua: FINANCIACIÓN PÚBLICA I+D ESTADO
Valor Normal esperado
Valor observado

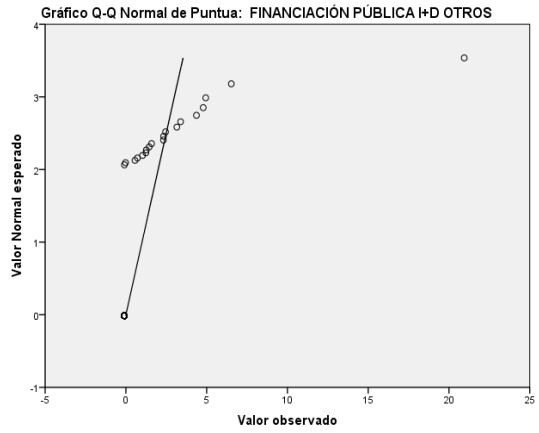


Gráfico Q-Q Normal de Puntua: FINANCIACIÓN PÚBLICA I+D OTROS
Valor Normal esperado
Valor observado

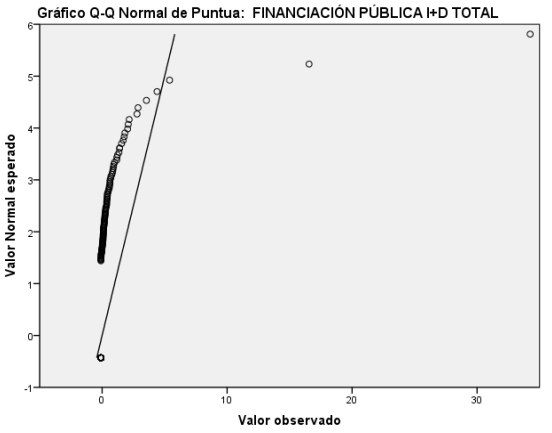


Gráfico Q-Q Normal de Puntua: FINANCIACIÓN PÚBLICA I+D TOTAL
Valor Normal esperado
Valor observado

Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN IDIOMAS

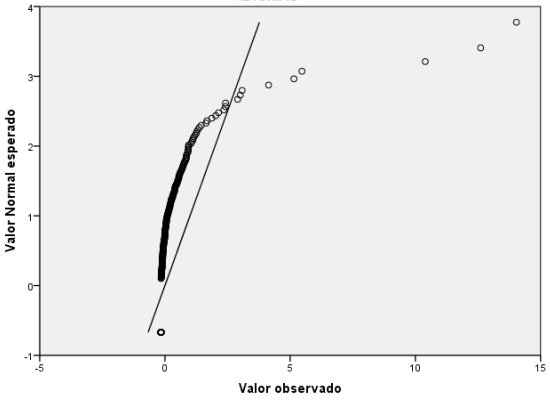


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA

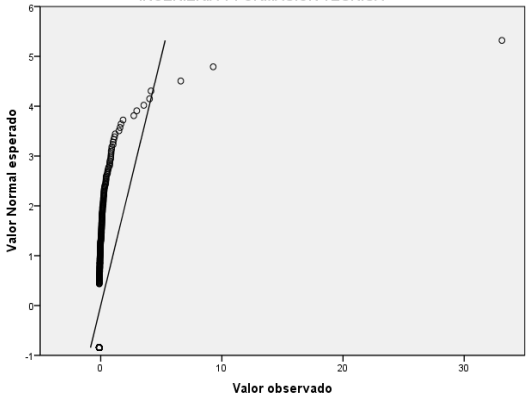


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN INGENIERÍA Y FORMACIÓN TÉCNICA POR

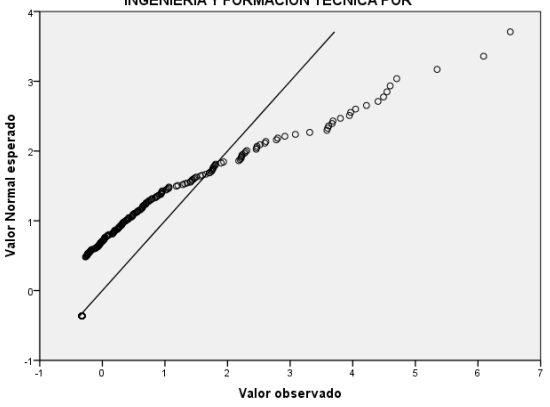


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR

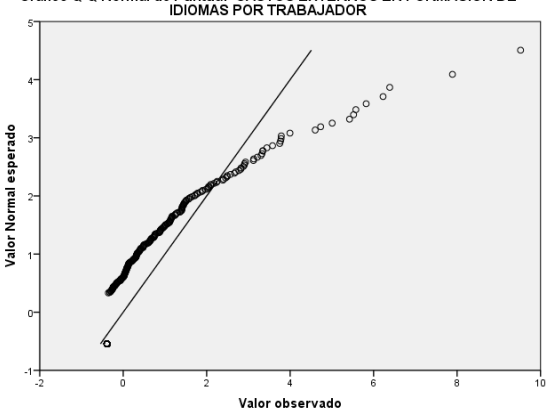


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA Y TECNOLOGÍAS DE LA INFORMACIÓN

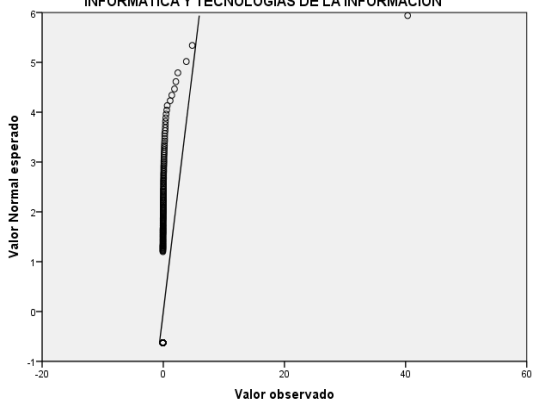


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN INFORMÁTICA SOBRE COSTE PERSONAL

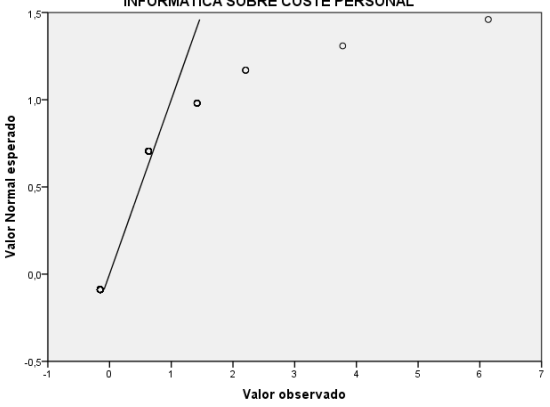


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN DE INFORMÁTICA POR TRABAJADOR

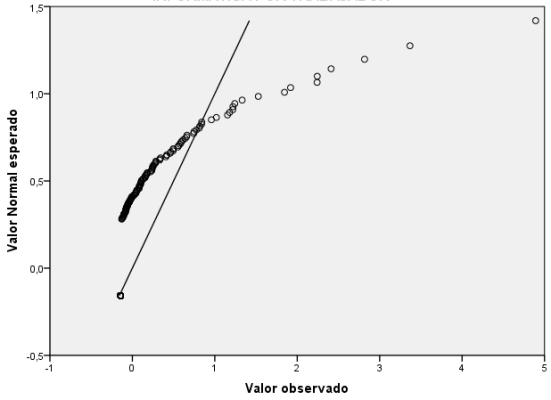


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS TOTALES EN FORMACIÓN

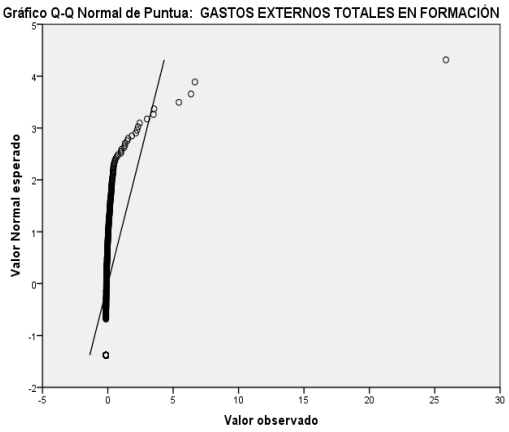


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN EN VENTAS Y MARKETING

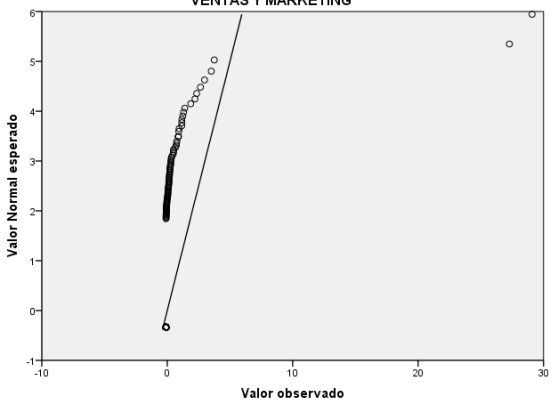


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS EN FORMACIÓN DE EN VENTAS Y MARKETING POR TRABAJADOR

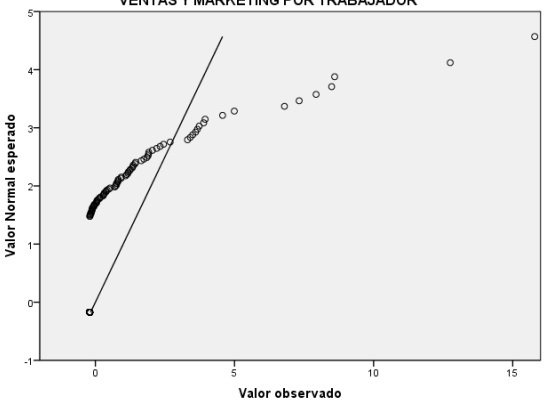


Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS I+D

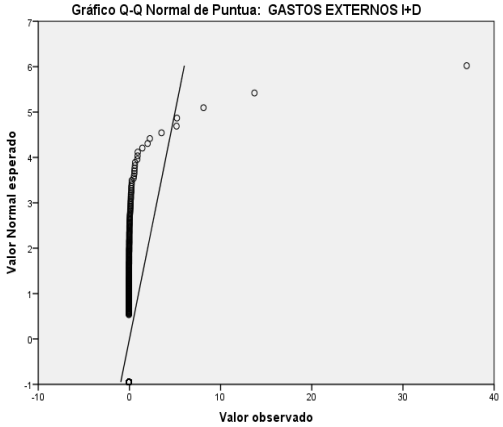
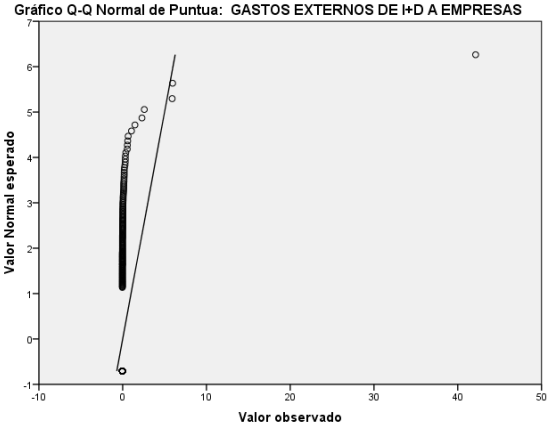
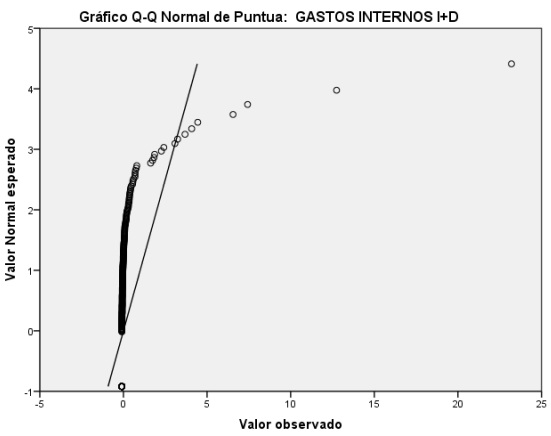
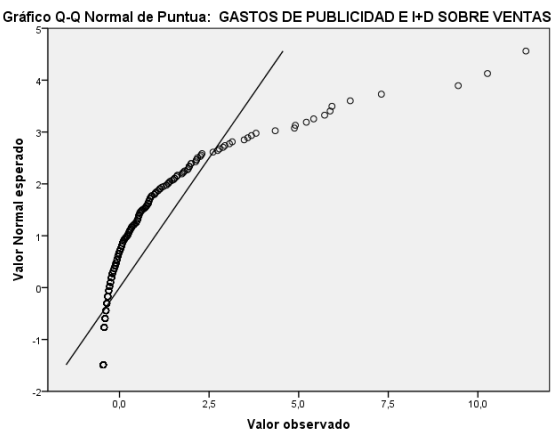
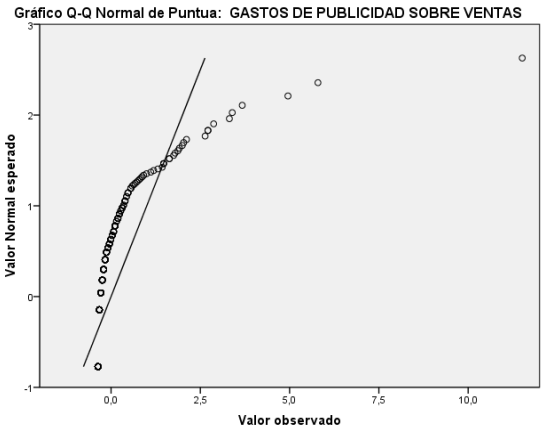


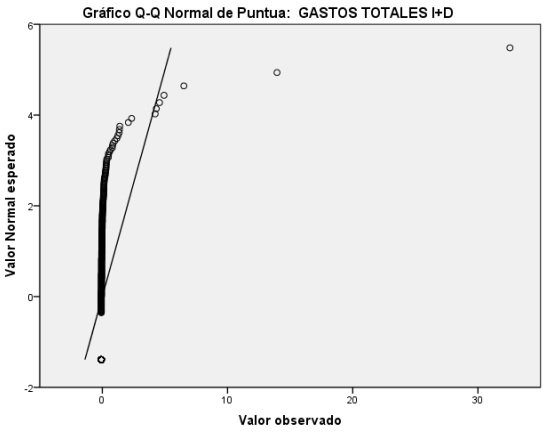
Gráfico Q-Q Normal de Puntua: GASTOS EXTERNOS DE I+D A EMPRESAS

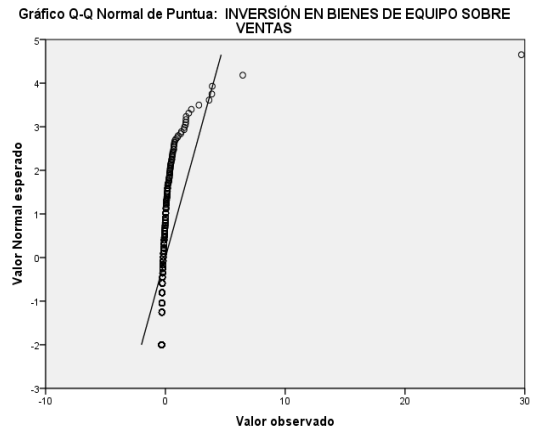


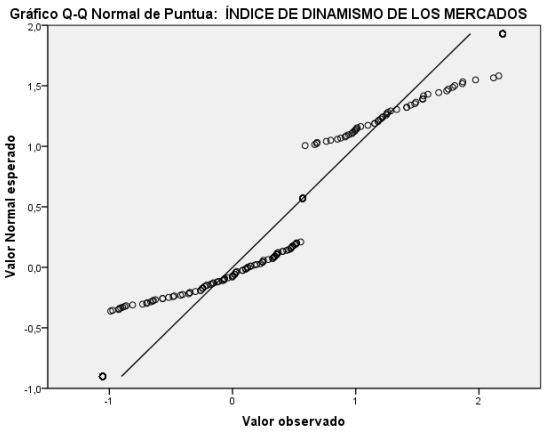




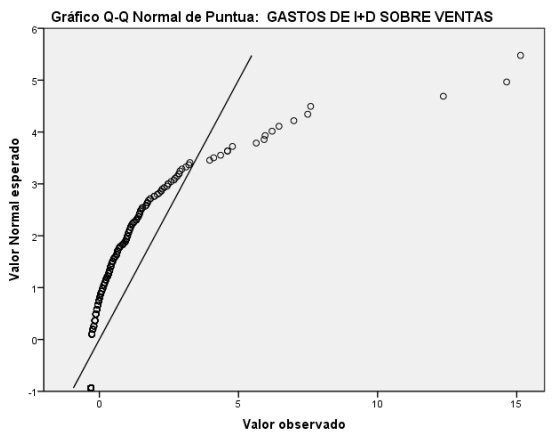


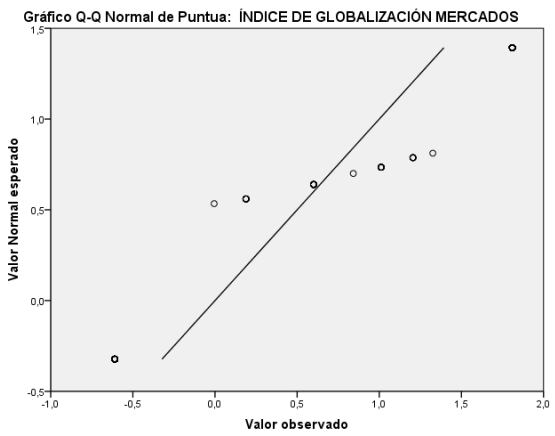




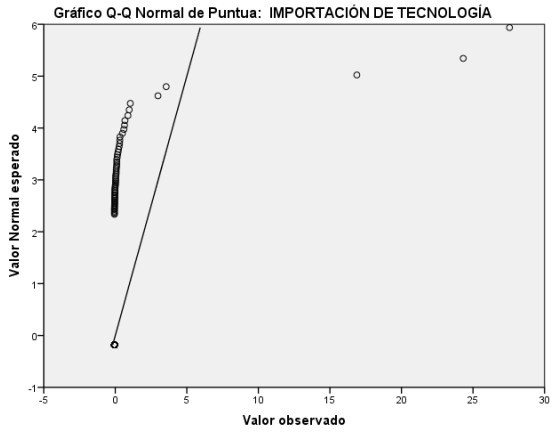


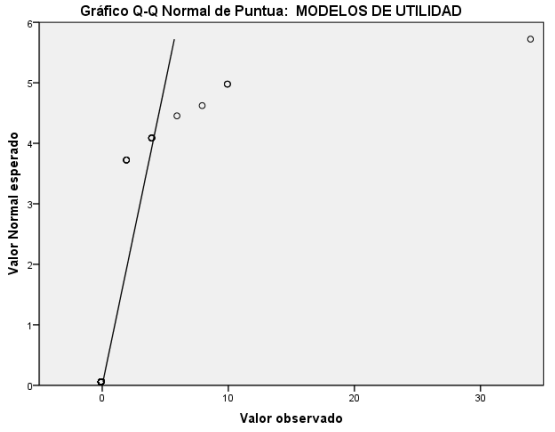
Se excluye: IDMERPN

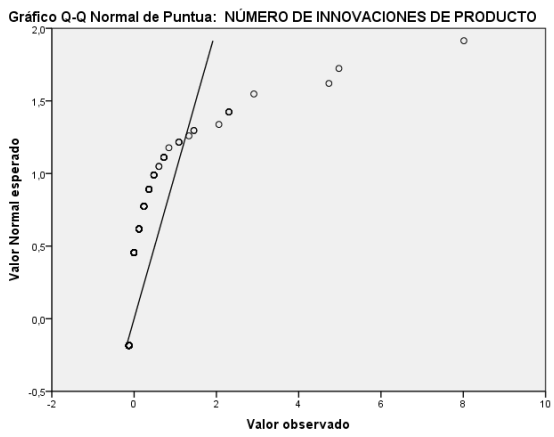


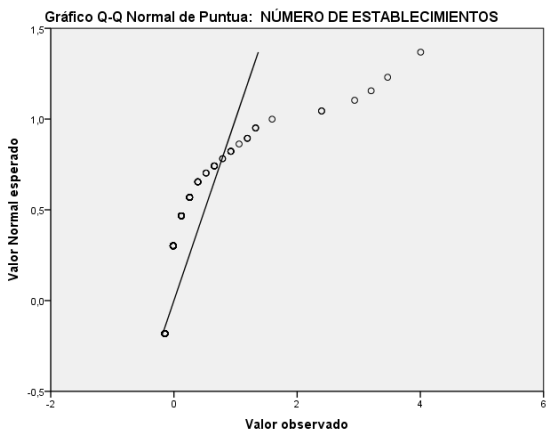


Se excluye: IGM









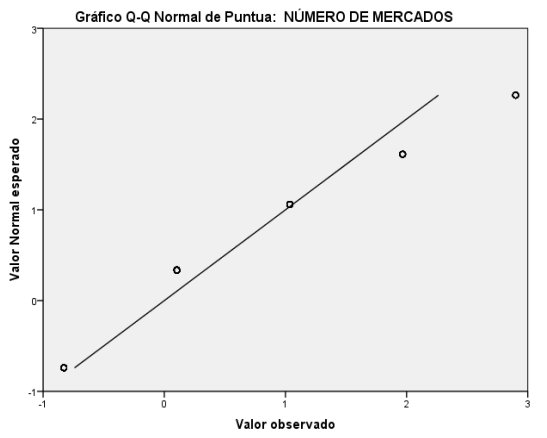


Gráfico Q-Q Normal de Puntua: NÚMERO DE MERCADOS
Valor Normal esperado
Valor observado

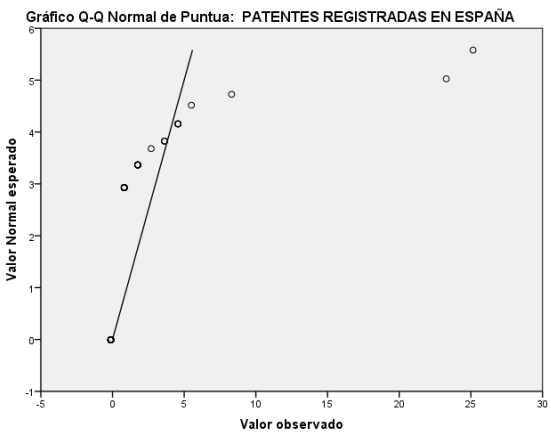


Gráfico Q-Q Normal de Puntua: PATENTES REGISTRADAS EN ESPAÑA
Valor Normal esperado
Valor observado

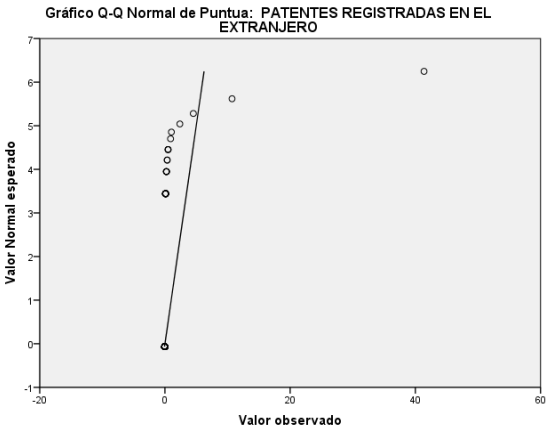


Gráfico Q-Q Normal de Puntua: PATENTES REGISTRADAS EN EL EXTRANJERO
Valor Normal esperado
Valor observado

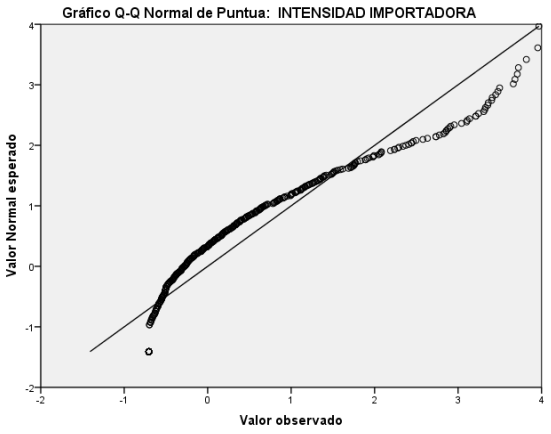


Gráfico Q-Q Normal de Puntua: INTENSIDAD IMPORTADORA
Valor Normal esperado
Valor observado

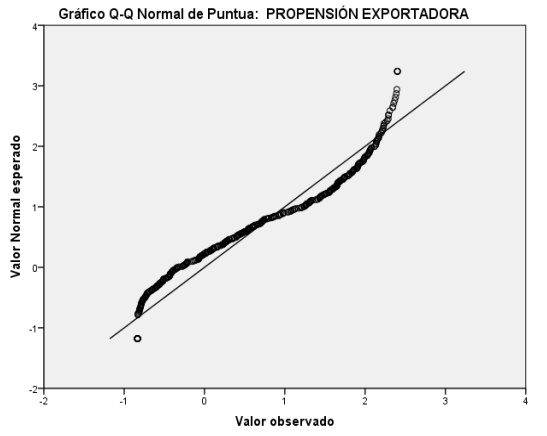


Gráfico Q-Q Normal de Puntua: PROPENSIÓN EXPORTADORA
Valor Normal esperado
Valor observado

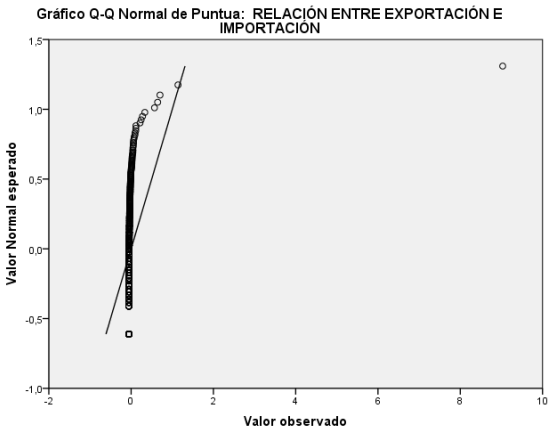


Gráfico Q-Q Normal de Puntua: RELACIÓN ENTRE EXPORTACIÓN E IMPORTACIÓN
Valor Normal esperado
Valor observado

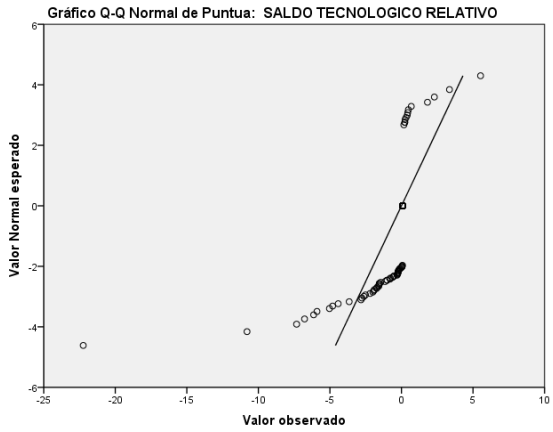
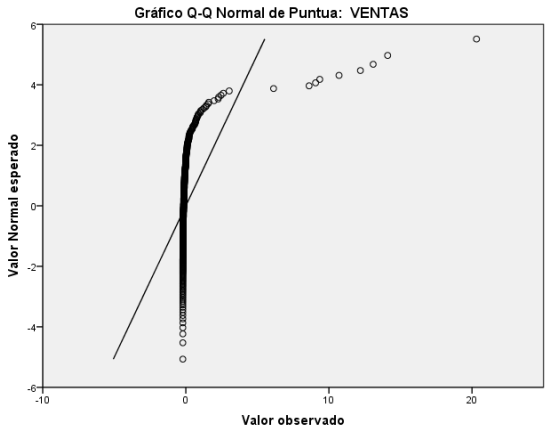
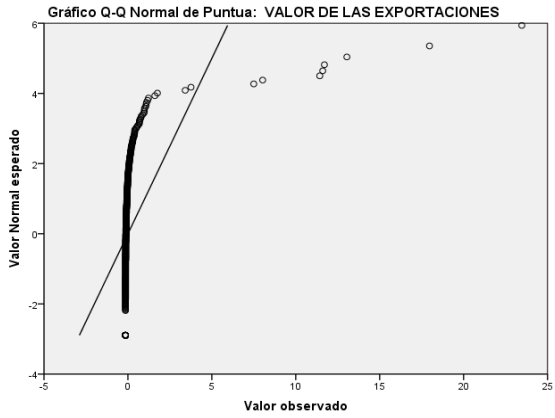
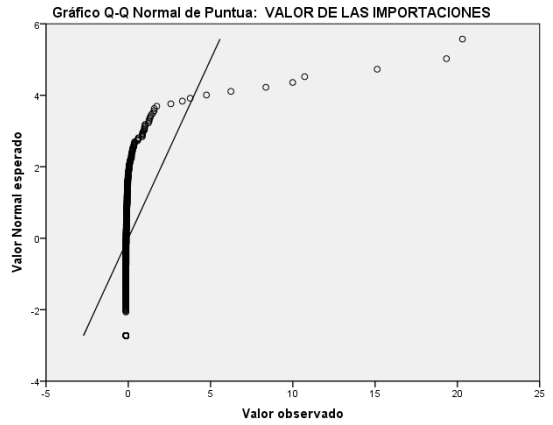
<table>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Se excluye: STEREL</td><td></td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td></td><td></td></tr>
</table>

Tabla 6.3. Gráficos Q-Q de las variables tipificadas en el AFE. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS.

Se excluyen del análisis de Análisis Factorial por notorios problemas de normalidad a las variables: Saldo tecnológico relativo (STEREL), Autonomía tecnológica (AUTEC), Índice de globalización de mercados (IGM) e Índice de globalización de mercados (IDMERPN). La totalidad de las variables restantes deben ser observadas con detenimiento y especial precaución en el AFE, no obstante, en esta etapa del análisis se mantienen dado la relevancia teórica de las mismas.

6.4 Resumen de datos perdidos disponible para el Análisis de Regresión Logística Binaria Multivariada

Se presenta en la Tabla 6.4 el resumen de los casos perdidos de las variables de partida para la elaboración del modelo final de regresión logística binaria multitvariada.

	Casos					
	Incluidos		Excluidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
APLICAR INCENTIVOS FISCALES I+D	1869	100,0%	0	,0%	1869	100,0%
COMERCIALIZACIÓN PRODUCTOS EXTRANJEROS	1869	100,0%	0	,0%	1869	100,0%
COMERCIALIZACIÓN PRODUCTOS SOBRE VENTAS	1869	100,0%	0	,0%	1869	100,0%
COLABORACIÓN TECNOLÓGICA CON CLIENTES	1869	100,0%	0	,0%	1869	100,0%
COLABORACIÓN CON UNIVERSIDADES Y/O C. TECNOLÓGICOS	1869	100,0%	0	,0%	1869	100,0%
DIRECCIÓN O COMITÉ DE TECNOLOGÍA	1869	100,0%	0	,0%	1869	100,0%
FINANCIACIÓN DE LA INNOVACIÓN CON CRED. SUBVENC.	1869	100,0%	0	,0%	1869	100,0%
GASTOS EXTERNOS EN FORMACIÓN DE IDIOMAS POR TRABAJADOR	1869	100,0%	0	,0%	1869	100,0%
INNOVACIONES DE COMERCIALIZACIÓN	1869	100,0%	0	,0%	1869	100,0%
GASTOS DE I+D SOBRE VENTAS	1863	99,7%	6	,3%	1869	100,0%
INCORPOR. DE INGENIEROS Y/O LICENCIADOS RECIENTES	1869	100,0%	0	,0%	1869	100,0%
INNOVACIONES DE MÉTODOS ORGANIZATIVOS	1869	100,0%	0	,0%	1869	100,0%
INNOVACIONES DE PROCESO POR NUEVOS EQUIPOS	1869	100,0%	0	,0%	1869	100,0%
INNOVACIONES DE PROCESO POR NUEVAS TÉCNICAS	1869	100,0%	0	,0%	1869	100,0%
INTENSIDAD IMPORTADORA	1866	99,8%	3	,2%	1869	100,0%
PROPENSIÓN EXPORTADORA	1868	99,9%	1	,1%	1869	100,0%
RECLUTAR PERS. EXPERIENCIA EMPRESARIAL I+D	1869	100,0%	0	,0%	1869	100,0%

Tabla 6.4. Resumen de datos perdidos disponible para el Análisis de Regresión Logística Binaria Multivariada. Fuente: Elaboración propia a partir de datos de la Fundación SEPI (2012) y uso del software estadístico SPSS.

6.5 REAL DECRETO 475/2007, por el cual se aprueba la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009)

REAL DECRETO 475/2007[16], de 13 de abril, por el que se aprueba la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009)[17].

El Real Decreto 1560/1992, de 18 de diciembre, estableció una Clasificación Nacional de Actividades Económicas, denominada CNAE-93, que sufrió una actualización de orden menor en el año 2003, denominada CNAE-93 Rev.1.

Con el objeto de reflejar los cambios estructurales de la economía, y en especial el desarrollo tecnológico habido desde la última revisión de la clasificación, debe establecerse una Clasificación Nacional de Actividades Económicas actualizada, denominada en lo sucesivo CNAE-2009.

Esta clasificación supone un impulso en los esfuerzos por modernizar la producción de las estadísticas nacionales al adecuarse a la realidad actual, y permite que las empresas, las entidades financieras, los gobiernos y los demás operadores del mercado dispongan de datos fiables y comparables.

La comparabilidad internacional de las estadísticas requiere que los países utilicen clasificaciones de actividades económicas que sigan las recomendaciones internacionales.

España tiene que cumplir los requerimientos del Reglamento (CE) n.º 1893/2006 del Parlamento Europeo y del Consejo, de 20 de diciembre de 2006, que establece la clasificación europea de actividades económicas (NACE Rev.2) y la relación que deben tener las versiones nacionales con esta clasificación. Además,

---

[16] Real Decreto 475/ 2007, para mayor información consultar en INE:
http://www.ine.es/daco/daco42/clasificaciones/cnae09/cnae_2009_rd.pdf
[Recuperado: 09-02-2017]

[17] CNAE 2009, Para mayor información consultar en INE:
http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft40%2Fclasrev%2F&file=inebase&L=0
[Recuperado: 09-02-2017]

esta clasificación europea, y por ende las versiones nacionales, siguen las recomendaciones adoptadas por la Comisión de Estadística de Naciones Unidas materializadas en la vigente Clasificación Internacional Industrial Uniforme (CIIU Rev.4).

Para garantizar la comparabilidad de la información a lo largo del periodo de vigencia de la clasificación es necesario interpretar de un modo uniforme las distintas categorías de la CNAE-2009. Asimismo, la realidad económica cambia gradualmente, pudiendo hacer necesarios ligeros cambios en la estructura de la CNAE-2009, en especial para adaptarse a las modificaciones que puedan sufrir las clasificaciones internacionales de referencia.

La transición de la CNAE-93 Rev.1 a la CNAE-2009 requiere una plena coordinación del sistema estadístico. Para ello es preciso que los productores de estadísticas oficiales adapten sus sistemas estadísticos antes de implantar la nueva clasificación. De ahí que sea conveniente un periodo entre la aprobación de este Real Decreto y la aplicación de la clasificación en las operaciones estadísticas.

El Instituto Nacional de Estadística, conforme a la facultad que le confiere el artículo 26 de la Ley 12/1989, de 9 de mayo, de la Función Estadística Pública, eleva la propuesta de aprobación de la CNAE-2009, con el objeto de proceder a su implantación efectiva, una vez concluidos los trabajos de adaptación de la CNAE-2009 a la citada NACE Rev.2, y disponiendo de la conformidad de la Comisión Europea, según consta en el artículo 4.3 del citado Reglamento, y del preceptivo dictamen del Consejo Superior de Estadística.

En su virtud, a propuesta del Ministro de Economía y Hacienda, previa deliberación del Consejo de Ministros en su reunión del día 13 de abril de 2007.

BOE núm. 102 Sábado 28 abril 2007 18573

D I S P O N G O:

Artículo 1. Objeto. El presente Real Decreto establece la Clasificación Nacional de Actividades Económicas 2009, denominada en lo sucesivo CNAE-2009, que figura en el Anexo.

Artículo 2. Fecha de aplicación de la CNAE-2009. La CNAE-2009 se aplicará a partir del 1 de enero del 2009, salvo que haya alguna normativa en la legislación comunitaria que estipule otra fecha en algún ámbito estadístico.

Artículo 3. Estructura de la CNAE-2009. La CNAE-2009 incluye:

a) un primer nivel consistente en rúbricas identificadas mediante un código alfabético (secciones), b) un segundo nivel consistente en rúbricas identificadas mediante un código numérico de dos cifras (divisiones), c) un tercer nivel consistente en rúbricas identificadas mediante un código numérico de tres cifras (grupos), y d) un cuarto nivel consistente en rúbricas identificadas mediante un código numérico de cuatro cifras (clases).

Artículo 4. Actividades del Instituto Nacional de Estadística para facilitar el uso de la CNAE-2009. El Instituto Nacional de Estadística, en función de sus competencias, garantizará la difusión y el mantenimiento de la CNAE-2009:

a) elaborando, actualizando y publicando notas explicativas y normas de clasificación de la CNAE-2009; b) revisando periódicamente el contenido de las rúbricas comprendidas en la CNAE-2009, siempre de acuerdo a lo dispuesto en el artículo 6.1 del Reglamento (CE) n.º 1893/2006 del Parlamento Europeo y del Consejo; y c) publicando cuadros de equivalencias entre la CNAE-93 Rev.1 y la CNAE-2009 y entre la CNAE-2009 y la NACE Rev.2.

Artículo 5. Cambios en la estructura de la CNAE-2009. El Instituto Nacional de Estadística, previo dictamen del Consejo Superior de Estadística, podrá:

a) modificar la CNAE-2009 para incorporar modificaciones introducidas en la NACE Rev.2 en virtud del artículo 6.2 del Reglamento (CE) n.º 1893/2006 del Parlamento Europeo y del Consejo; y b) incorporar a la CNAE-2009 un nivel adicional consistente en rúbricas identificadas mediante un código numérico de cinco cifras (subclases), que se corresponda con una subdivisión exacta de las partidas del cuarto nivel (clases), con el objeto de dar respuesta a las necesidades que sobre esta clasificación se puedan plantear en el futuro.

Artículo 6. Aplicación de la CNAE-2009 en el ámbito estadístico. La CNAE-2009 será de uso obligatorio en el ámbito de aplicación de la Ley 12/1989, de 9 de

mayo, de la Función Estadística Pública para todas aquellas operaciones estadísticas recogidas en el Plan Estadístico Nacional.

Disposición final única. Entrada en vigor. El presente real decreto entrará en vigor el día siguiente al de su publicación en el «Boletín Oficial del Estado».

Dado en Madrid, el 13 de abril de 2007.

JUAN CARLOS R.

El Vicepresidente Segundo del Gobierno

y Ministro de Economía y Hacienda,

PEDRO SOLBES MIR

6.6 Listado de sectores según CNAE 2009: alta y media-alta tecnología, servicios de alta tecnología o de punta

Los Sectores Alta y Media-alta tecnología compuestos por los sectores manufactureros y servicios de alta tecnología o punta, se presentan a continuación:

Cuadro 4. Sectores de Alta y Media-Alta Tecnología.	
Lista utilizada por el INE	
CNAE 2009	**SECTORES**
	Sectores manufactureros de tecnología alta
21	Fabricación de productos farmacéuticos
26	Fabricación de productos informáticos, electrónicos y ópticos
30.3	Construcción aeronáutica y espacial y su maquinaria
	Sectores manufactureros de tecnología media-alta
20	Industria química
25.4	Fabricación de armas y municiones
27 a 29	Fabricación de material y equipo eléctrico; Fabricación de maquinaria y equipo n.c.o.p.; Fabricación de vehículos de motor, remolques y semirremolques
30- 30.1 - 30.3	Fabricación de otro material de transporte excepto: construcción naval; construcción aeronáutica y espacial y su maquinaria.
32.5	Fabricación de instrumentos y suministros médicos y odontológicos
	Servicios de alta tecnología o de punta
59 a 63	Actividades cinematográficas, de video y de programas de televisión, grabación de sonido y edición musical; Actividades de programación y emisión de radio y televisión; Telecomunicaciones; Programación, consultoria y otras actividades relacionadas con la informática; Servicios de información.
72	Investigación y desarrollo

Tabla 6.5. Listado de sectores según CNAE 2009: alta y media-alta tecnología, servicios de alta tecnología o de punta. Fuente: INE, en: www.ine.es/daco/daco42/daco4217/lstsectcnae.xls [Recuperado: 09-02-2017]

Para determinar la lista de actividades y de productos de alta tecnología, se ha partido de los trabajos metodológicos llevados a cabo por la OCDE[18] en esta materia, así como de las adaptaciones de los mismos al ámbito europeo realizadas por parte de la Eurostat (INE 2005).

[18] OCDE (1995). Classification des secteurs et des produits haute technologíe.

6.7 Lista de productos de alta tecnología según códigos SITC.REV 4 y correspondencias con CPA-2008 y PRODCOM 2008

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
1.	Construcción aeronáutica y espacial		
1.1	Motores para aeronaves	71441+ 71449+ 71481+ 71491	30301200+ 30301300+ 30301600
1.2	Helicópteros	79211+ 79215	30303100
1.3	Aviones	7922+ 7923+ 7924	30303200+ 30303300+ 30303400
1.4	Naves espaciales	7925	30304000
1.5	Hélices, rotores y sus partes	79291	30305030
1.6	Trenes de aterrizaje y sus partes	79293	30305050
1.7	Brújulas; otros instrumentos y aparatos para la navegación aérea	87411	26511120+ 26511150+ 26511180
2.	Maquinaria de oficina y equipo informático		
2.1	Máquinas de escribir automáticas y máquinas para el tratamiento de textos	75194	26201800
2.2	Fotocopiadoras	75195	26201640
2.3	Ordenadores	7522+ 7523+ 7526+ 7527+ 7528+ 7529	26201100+ 26201300+ 26201400+ 26201500+ 26201650+ 26201660+ 26202100+ 26990OZ0
2.4	Partes y accesorios de ordenadores	75997	26204000
3.	Material electrónico; equipos y aparatos de radio, televisión y comunicaciones		
3.1	Aparatos de grabación y reproducción de sonido	76331	26403100
3.2	Aparatos de vídeo	76381+ 76384	26403300

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
3.3	Equipo de telecomunicación	76411+ 76412+ 76418+ 76421+ 76422+ 76423+ 76424+ 76425+ 76426+ 76431+ 76432+ 76483+ 76484+ 76492	26302100+ 26302200+ 26302330+ 26302310+ 26302320+ 26302370+ 26302340+ 26404400+ 26304040+ 26303000+ 26404100+ 26404235+ 26404237+ 26404239+ 26404270+ 26404355+ 26404359+ 26404370+ 26301200+ 26301100+ 26512020+ 26512050+ 26512080+ 26301300+ 26701300+ 26405180
3.4	Circuitos impresos	7722	26121020+ 26121050+ 26121080
3.5	Cuadros y otros soportes con aparatos eléctricos para una tensión < 1.000 V	77261	27123130+ 27123150+ 27123170
3.6	Cables de fibra óptica	77318	27311100
3.7	Tubos de microondas	77625	26111200
3.8	Otras válvulas y tubos	77627	
3.9	Semiconductores	77631+ 77632+ 77633+ 77635+ 77637+ 77639	26112120+ 26112150+ 26112180+ 26112220+ 26112240+ 26112260
3.10	Circuitos electrónicos integrados	77642+ 77644+ 77646+ 77649	26113003+ 26113006+ 26113023+ 26113027+ 26113034+ 26113054+ 26113065+ 26113067+ 26113080+ 26113091+ 26113094
3.11	Cristales piezoeléctricos	77681+ 77688+ 77689	26112280+ 26114070+ 26114090
3.12	Soportes ópticos	89844	26801200
3.13	Partes semiconductoras	89846	26202200+ 26123000+ 26801300
4.	Productos farmacéuticos		
4.1	Antibióticos	54131+ 54132+ 54133+ 54139	21105400

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
4.2	Hormonas y sus derivados	54153+ 54154+ 54155+ 54156+ 54159	21105200
4.3	Glucósidos, glándulas, sueros, vacunas	54161+ 54162+ 54163+ 54164	21105300+ 21106020+ 21106040+ 21202120+ 21202140+ 21202160+ 21106050
4.4	Medicamentos que contienen penicilinas u otros antibióticos	54211+ 54212+ 54213+ 54219	21201130+ 21201150+ 21201160+ 21201180
4.5	Medicamentos que contienen hormonas, pero no antibióticos	54221+ 54222+ 54223+ 54224+ 54229	21201230+ 21201250+ 21201260+ 21201270
5.	Instrumentos científicos		
5.1	Aparatos de electrodiagnóstico para medicina, cirugía y radiología	77411+ 77412+ 77413+ 77421+ 77422+ 77423+ 77429	26601230+ 26601280+ 26601300+ 26601115+ 26601119+ 26601130+ 26601150+ 26601170
5.2	Instrumentos y aparatos ópticos	87111+ 87115+ 87119+ 87131+ 87139+ 87141+ 87143+ 87145+ 87149+ 87191+ 87192+ 87193+ 87199	26702230+ 26702250+ 26702410+ 26516100+ 26518300+ 26702270+ 26702430+ 26702310+ 26702330+ 26702390+ 26702500
5.3	Tornos dentales	87211	32501130
5.4	Instrumentos y aparatos de medida	87412+ 87413+ 87414+ 87431+ 87435+ 87437+ 87439+ 87441+ 87442+ 87443+ 87445+ 87446+ 87449+ 87451+ 87452+ 87453+ 87454+ 87455+ 87456+ 87461+ 87463+ 87465+ 87469+ 87471+ 87473+ 87475+ 87477+ 87478+ 87479+ 8749	26518600+ 26511215+ 26511270+ 28293960+ 26511235+ 26511239+ 26518200+ 26515235+ 26515239+ 26515255+ 26515259+ 26515271+ 26515274+ 26515279+ 26515283+ 26515289+ 26515313+ 26515319+ 26515320+ 26515330+ 26515350+ 26515390+ 26515381+ 26515383+ 26513100+ 32995300+ 26516210+ 26516230+ 26516255+ 26516259+

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
			32501340+ 26515110+ 26515135+ 26515139+ 26515150+ 26515175+ 26515179+ 26517015+ 26517019+ 26517030+ 26516500+ 26517090+ 26518550+ 26514100+ 26514200+ 26514530+ 26514555+ 26514559+ 26514310+ 26514330+ 26514355+ 26514359+ 26514400+ 26514520
5.5	Aparatos fotográficos	88111	26701250+ 26701400
5.6	Cámaras cinematográficas	88121	26701530+ 26701550
5.7	Lentes de contacto	88411	32504130
5.8	Otras fibras ópticas	88419	27311200+ 26702180+ 26702153
5.9	Aparatos de ortopedia	89961+ 89963+ 89966+ 89967	26601433+ 32502239+ 32502235+ 32502290+ 26601450
6.	Maquinaria y material eléctrico		
6.1	Condensadores eléctricos fijos, variables o regulables	77862+ 77863+ 77864+ 77865+ 77867+ 77868	27905220+ 27905240+ 27905300
6.2	Máquinas y aparatos eléctricos con función propia	77871+ 77878+ 77879	27904010+ 27904030+ 28491283+ 27901150+ 28993960+ 27903370
6.3	Aparatos eléctricos de señalización sonora o visual	77884	26305080+ 26305020+ 27902050+ 27902020+ 27902080
7.	Productos químicos		
7.1	Selenio, teluro, fósforo, arsenio y boro	52222	20132140+ 20132180
7.2	Silicio	52223	20132150
7.3	Calcio, estroncio y bario	52229	20132300

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
7.4	Otros productos químicos inorgánicos	52269	20121950+ 20121200+ 20121973+ 20121975+ 20121990
7.5	Materias radioactivas	52511+ 52513+ 52515+ 52517+ 52519+ 52591+ 52595	20136100+ 20136500
7.6	Materias colorantes orgánicas sintéticas y lacas colorantes	53111+ 53112+ 53113+ 53114+ 53115+ 53116+ 53117+ 53119+ 53121+ 53122	20122110+ 20122120+ 20122130+ 20122140+ 20122150+ 20122160+ 20122170
7.7	Tereftalato de polietileno	57433	20164062+ 20164064
7.8	Insecticidas y desinfectantes	5911+ 5912+ 5913+ 5914+ 5919	20201160+ 20201130+ 20201140+ 20201150+ 20201190+ 20201515+ 20201530+ 20201545+ 20201560+ 20201575+ 20201590+ 20201220+ 20201230+ 20201240+ 20201250+ 20201260+ 20201270+ 20201290+ 20201350+ 20201370+ 20201430+ 20201450+ 20201490+ 20201930+ 20201980
8.	Maquinaria y equipo mecánico		
8.1	Turbinas de gas	71489	28112300
8.2	Partes de turbinas de gas	71499	28113300
8.3	Reactores nucleares	71871+ 71877+ 71878	25302100+ 25302200
8.4	Maquinaria y equipos para la separación de isótopos	72847	28993910
8.5	Máquinas-herramienta operadas por láser u otro tipo de luz o haz de luz, etc.	73111+ 73112+ 73113+ 73114	28411110+ 28411130+ 28411150+ 28411170
8.6	Tornos horizontales de control numérico	73131	28412123+ 28412127+ 28412129

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
8.7	Los demás tornos de control numérico	73135	28412160
8.8	Las demás taladradoras de control numérico	73142	28412213
8.9	Las demás escariadoras-fresadoras de control numérico	73144	28412240
8.10	Fresadoras knee-type de control numérico	73151	28412217
8.11	Las demás fresadoras de control numérico	73153	28412223+ 28412225
8.12	Máquinas de desbarbar, afilar, amolar, rectificar, lapear (bruñir), pulir o hacer otras operaciones de acabado, de control numérico	73161	28412305
8.13	Las demás máquinas de desbarbar, afilar, amolar, rectificar, lapear (bruñir), pulir o hacer otras operaciones de acabado, de control numérico	73163	28412315+ 28412325
8.14	Afiladoras de control numérico	73165	28412365
8.15	Máquinas de enrollar, curvar, plegar, enderezar o aplanar, de control numérico	73312	28413120+ 28413140
8.16	Fundidoras de control numérico	73314	28413220
8.17	Perforadoras de control numérico	73316	28413240
8.18	Partes y accesorios destinados a las máquinas de los subgrupos 8.5 a 8.17	73591+ 73595	28414030+ 28414050
8.19	Máquinas y equipos para la soldadura por	73733	27903145

Grupo	Descripción	Código SITC rev.4	Código CPA-2008/Código PRODCOM2008
	resistencia de metales, parcial o completamente automatizados		
8.20	Máquinas y equipos para la soldadura por arco de metales, parcial o completamente automatizados	73735	27903154
9.	Armas y municiones		
9.1	Armas y municiones	89111+ 89112+ 89113+ 89114+ 89122+ 89123+ 89124+ 89129+ 89131+ 89139+ 89191+ 89193+ 89195+ 89199	25711500+ 25401230+ 25401300+ 25401250+ 25401270+ 25401290+ 25401400

Tabla 6.6. Lista de productos de alta tecnología según códigos SITC.REV 4 y correspondencias con CPA-2008 y PRODCOM 2008. Fuente: Elaboración propia a partir de INE. http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176979&menu=metodologia&idp=1254735576669

7 BIBLIOGRAFÍA

Acosta, M. & Coronado, D., 2003. Science–technology flows in Spanish regions. *Research Policy*, 32(10), pp.1783–1803. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733303000647.

Acosta, M., Coronado, D. & Toribio, M.R., 2011. The use of scientific knowledge by Spanish agrifood firms. *Food Policy*, 36(4), pp.507–516. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0306919211000625.

Agrawal, R. & Srikant, R.A., 1994. Fast Algorithms for Mining Association Rules in Large Databases. *20th International Conference on Very Large Data Bases*, p.478–499. Available at: http://cs.stanford.edu/people/chrismre/cs345/rl/ar-mining.pdf.

Aldieri, L. & Cincera, M., 2009. Geographic and technological R&D spillovers within the triad: micro evidence from US patents. *The Journal of Technology Transfer*, 34(2), pp.196–211. Available at: http://link.springer.com/10.1007/s10961-007-9065-8.

Alegre, J., 2004. *La gestión del conocimiento como motor de la innovación : lecciones de la industria de alta tecnología para la empresa*, Castelló de la Plana: Publicacions de la Universitat Jaume I.

Archibugi, D., 1992. Patenting as an indicator of technological innovation: a review. *Science and Public Policy*. Available at: https://academic.oup.com/spp/article/19/6/357/1678742/Patenting-as-an-indicator-of-technological.

Baden-Fuller, C. & Haefliger, S., 2013. Business Models and Technological Innovation. *Long Range Planning*, 46(6), pp.419–426. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0024630113000691.

Barbic, F. et al., 2016. Governance dynamics in multi-partner R&D alliances: Integrating the control and coordination perspectives. *Baltic Journal of*

Management, 11(4), pp.405–429. Available at: http://www.emeraldinsight.com/doi/10.1108/BJM-09-2015-0173.

Barbosa, N., Faria, A.P. & Eiriz, V., 2014. Industry- and firm-specific factors of innovation novelty. *Industrial and Corporate Change*, 23(3), pp.865–902. Available at: http://icc.oxfordjournals.org/cgi/doi/10.1093/icc/dtt029.

Becker, B. & Hall, S.G., 2013. Do R&D strategies in high-tech sectors differ from those in low-tech sectors? An alternative approach to testing the pooling assumption. *Economic Change and Restructuring*, 46(2), pp.183–202. Available at: http://link.springer.com/10.1007/s10644-012-9122-7.

Belderbos, R. et al., 2004. Heterogeneity in R&D cooperation strategies. *International Journal of Industrial Organization*, 22(8–9), pp.1237–1263. Available at: http://linkinghub.elsevier.com/retrieve/pii/S016771870400102X.

Di Benedetto, A., 2010. Comment on "Is open innovation a field of study or a communication barrier to theory development?" *Technovation*, 30(11–12), p.557. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497210000957.

Berchicci, L., 2013. Towards an open R&D system: Internal R&D investment, external knowledge acquisition and innovative performance. *Research Policy*, 42(1), pp.117–127. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733312001266.

Bergfors, M. & Lager, T., 2011. Innovation of process technology: exploring determinants for organizational design. *International Journal of Innovation Management*, 15(5), pp.1113–1140. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-80055093876&partnerID=40&md5=4fc197baa4bdf62df8a84bc1b6bdd7e2.

Bianchi, M. et al., 2011. Organisational modes for Open Innovation in the bio-pharmaceutical industry: An exploratory analysis. *Technovation*, 31(1), pp.22–33. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497210000337.

Biscotti, A.M. & D'Amico, E., 2016. Theoretical foundation of IC disclosure strategies

in high-tech industries. *International Journal of Disclosure and Governance*, 13(1), pp.1–25. Available at: http://link.springer.com/10.1057/jdg.2015.8.

Blanes, J.V. & Busom, I., 2004. Who participates in R&D subsidy programs? *Research Policy*, 33(10), pp.1459–1476. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733304001027.

Block, J.H., 2012. R&D investments in family and founder firms: An agency perspective. *Journal of Business Venturing*, 27(2), pp.248–265. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902610000868.

Bonacci, I. & Tamburis, O., 2016. Empowering openness: the case of CROs-related innovation networks in the Italian bio-pharmaceutical sector. *International Journal of Learning and Intellectual Capital*, 13(2/3), p.184. Available at: http://www.inderscience.com/link.php?id=75697.

Buesa, M. et al., 2006. Regional systems of innovation and the knowledge production function: the Spanish case. *Technovation*, 26(4), pp.463–472. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497204002214.

Busom, I., Corchuelo, B. & Martínez Ros, E., 2010. Efectividad de los incentivos públicos a la innvetigación e innovación empresarial. In L. Sanz Menéndez & L. Cruz Castro, eds. *Análisis sobre ciencia e innovación en España*. Consejo Superior de Investigaciones Científicas – CSIC y FECYT, pp. 670–707.

Bychkova, O., 2016. Innovation by coercion: Emerging institutionalization of university-industry collaborations in Russia. *Social Studies of Science*, 46(4), pp.511–535. Available at: http://sss.sagepub.com/cgi/doi/10.1177/0306312716654768.

Cardamone, P., 2012. A micro-econometric analysis of the role of R&D spillovers using a nonlinear translog specification. *Journal of Productivity Analysis*, 37(1), pp.41–58. Available at: http://link.springer.com/10.1007/s11123-011-0225-3.

Cassou, S.P. & Xavier de Oliveira, E., 2011. Barriers to technological adoption in Spain and Portugal. *Portuguese Economic Journal*, 10(3), pp.189–209. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-81255188895&doi=10.1007%2Fs10258-010-0069-

1&partnerID=40&md5=75d1656df03961d7a0571dd7ad31a767.

Cea, M.Á., 2002. *Análsis Multivarible. Teoría y práctica en la investigación social.*, Madrid: Editorial Síntesis, S.A.

Chen, P.-C. et al., 2016. Do R&D expenditures matter more than those of marketing to company performance? The moderating role of industry characteristics and investment density. *Technology Analysis & Strategic Management*, 28(2), pp.205–216. Available at: http://www.tandfonline.com/doi/full/10.1080/09537325.2015.1093107.

Chen, P.-C. & Hung, S.-W., 2016. An actor-network perspective on evaluating the R&D linking efficiency of innovation ecosystems. *Technological Forecasting and Social Change*, 112, pp.303–312. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0040162516303067.

Chen, S.-H., 2014. The confluence of academia and industry: A case study of the Taiwanese Biopharmaceutical Innovation System. *Review of Policy Research*, 31(5), pp.408–429. Available at: http://doi.wiley.com/10.1111/ropr.12089.

Chen, W.-C., 2012. Innovation and duration of exports. *Economics Letters*, 115(2), pp.305–308. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0165176511005866.

Chesbrough, H., 2010. Business model innovation: Opportunities and barriers. *Long Range Planning*, 43(2–3), pp.354–363. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0024630109000569.

Chesbrough, H. & Crowther, A.K., 2006. Beyond high tech: early adopters of open innovation in other industries. *R and D Management*, 36(3), pp.229–236. Available at: http://doi.wiley.com/10.1111/j.1467-9310.2006.00428.x.

Christensen, J.L., 2013. The ability of current statistical classifications to separate services and manufacturing. *Structural Change and Economic Dynamics*, 26, pp.47–60. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0954349X13000416.

Ciftci, M. & Cready, W.M., 2011. Scale effects of R&D as reflected in earnings and returns. *Journal of Accounting and Economics*, 52(1), pp.62–80. Available at:

http://linkinghub.elsevier.com/retrieve/pii/S0165410111000206.

Comrey, A.L., 1985. *Manual de análisis factorial*, Madrid: Ediciones Cátedra, S.A.

Cooper, R.G., 1987. New products: What separates winners from losers? *Journal of Product Innovation Management*, 4(3), pp.169–184. Available at: http://doi.wiley.com/10.1016/0737-6782(87)90002-6.

Cordeiro, P.V.M., Dergint, D.E.A. & Hatakeyama, K., 2014. Proposal of method for an automatic complementarities search between companies' R&D. *International Journal of Innovation and Technology Management*, 11(2), p.1450001. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219877014500011.

Cordón, E., Rubio, E. & Agote, Á., 2005. La interfaz I+D/marketing en las empresas españolas de alta tecnología tendencias actuales e implicaciones para el éxito de los nuevos productos. *Investigaciones europeas de dirección y economía de la empresa*, Vol. 11, N, pp.179–201. Available at: http://dialnet.unirioja.es/servlet/articulo?codigo=1172972.

COTEC, 2014. *Informe COTEC 2014.Tecnología e Innovación en España.* J. E. Barranco, ed., Madrid: FUNDACIÓN COTEC PARA LA INNOVACIÓN TECNOLÓGICA-COTEC.

COTEC, 2012. *Tecnología e Innovación en España INFORME COTEC 2012* J. E. Barranco, ed., Madrid: FUNDACIÓN COTEC PARA LA INNOVACIÓN TECNOLÓGICA-COTEC. Available at: http://www.cotec.es.

Cruz-González, J. et al., 2014. Directions of external knowledge search: investigating their different impact on firm performance in high-technology industries P. Manlio Del Giudice & P. Vincenzo Maggioni, eds. *Journal of Knowledge Management*, 18(5), pp.847–866. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84914099131&doi=10.1108%2FJKM-06-2014-0243&partnerID=40&md5=9158055f6f07ef9b45411799fccacc7a.

Cuadras, C.M., 1991. *Métodos de Análisis Multivariante* 2a ed. PPU, ed., Barcelona: CMC Editions.

Curran, B. & Walsworth, S., 2014. Can you pay employees to innovate? Evidence from the Canadian private sector. *Human Resource Management Journal*, 24(3), pp.290–306. Available at: http://doi.wiley.com/10.1111/1748-8583.12036.

Czarnitzki, D. & Delanote, J., 2012. Size, age and innovativeness: Key determinants of growth? In *2012 IEEE International Conference on Management of Innovation & Technology (ICMIT)*. IEEE, pp. 746–749. Available at: http://ieeexplore.ieee.org/document/6225899/.

Daunfeldt, S.-O., Elert, N. & Johansson, D., 2016. Are high-growth firms overrepresented in high-tech industries? *Industrial and Corporate Change*, 25(1), pp.1–21. Available at: http://icc.oxfordjournals.org/lookup/doi/10.1093/icc/dtv035.

Delgado-Verde, M. et al., 2011. B.Capital social, capital relacional e innovación tecnológica. Una aplicación al sector manufacturero español de alta y media-alta tecnología. *Cuadernos de Economía y Dirección de la Empresa*, 14(4), pp.207–221. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866439101&doi=10.1016%2Fj.cede.2011.04.001&partnerID=40&md5=8496fcb181313e70ae7451fdc2e2a4a4.

Dirección General de Industria y de la PYME, 2014. *Retrato de las PYME 2014*, Madrid: Dirección General de Industria y de la PYME. Available at: http://www.ipyme.org/Publicaciones/Retrato_PYME_2014.pdf.

Dirección General de Política de la Pyme, 2010. Retrato de las PYME 2010. , pp.1–7. Available at: www.ipyme.org.

Dirección General de Política de La Pyme, 2010. *Competitividad, innovación y empresas de alto crecimiento en España. Identificación de las necesidades de estudio*. S. Martínez Rodriguez, ed., Madrid: Ministerio de Industria, Turismo y Comercio. Available at: http://www.ipyme.org/Publicaciones/CompetitividadInnovacionEmpresasCrecimiento.pdf.

Do, T.H., 2014. Determinants of innovation commercialization management and anticipated returns: An exploratory typology of SMEs. *International Journal of*

Innovation and Technology Management, 11(6), p.1450042. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219877014500424.

Dosi, G., Faillo, M. & Marengo, L., 2008. Organizational capabilities, patterns of knowledge accumulation and governance structures in business firms: An introduction. *Organization Studies*, 29(8–9), pp.1165–1185. Available at: http://oss.sagepub.com/cgi/doi/10.1177/0170840608094775.

Duarte Castillo, J., 2005. *Factores determinantes y críticos en empresas de servicios, para la obtención de ventajas competitivas sostenibles y transferibles a estrategias de globalización un análisis de la industria del software*. Universitat Autònoma de Barcelona. Available at: http://dialnet.unirioja.es/servlet/tesis?codigo=5174%0A%0A.

Einola, S. et al., 2016. Retrospective relational sensemaking in R&D offshoring. *Industrial Marketing Management*, pp.1–12. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0019850116301833.

Ernst, H., Hoyer, W.D. & Rübsaamen, C., 2010. Sales, marketing, and research-and-development cooperation across new product development stages: Implications for success. *Journal of Marketing*, 74(5), pp.80–92. Available at: http://journals.ama.org/doi/abs/10.1509/jmkg.74.5.80.

European Union, 2016. *The 2016 EU Industrial R&D Investment Scoreboard*, Sevilla. Available at: http://iri.jrc.ec.europa.eu/scoreboard16.html.

Eurostat, 2007. *Eurostat regional yearbook 2007*, Luxemburgo: Eurostat. Available at: http://ec.europa.eu/eurostat/documents/3217494/5615357/KS-AF-07-001-07-EN.PDF/40b31d3e-99b1-4864-bd7a-6887abab8a61.

Eurostat, 2008. *NACE Rev. 2 – Statistical classification of economic activities in the European Community*, Luxembourg: Luxembourg: Office for Official Publications of the European Communities. Available at: http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

Faems, D., Van Looy, B. & Debackere, K., 2005. Interorganizational collaboration and innovation: Toward a portfolio approach*. *Journal of Product Innovation*

Management, 22(3), pp.238–250. Available at: http://doi.wiley.com/10.1111/j.0737-6782.2005.00120.x.

FECYT, 2006. *Carencias y necesidades del Sistema Español de Ciencia y Tecnología. Recomendaciones para mejorar los procesos de transferencia de conocimiento y tecnología a las empresas. Informe 2005.*, Madrid: FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA – FECYT.

FECYT, 2012. *Estudio sobre los efectos de la I+D en los resultados empresariales para España*, Fundación Española para la Ciencia y la Tecnología (FECYT). Available at: https://icono.fecyt.es/informesypublicaciones/Documents/Informe_Efectos_innovacion.pdf.

FECYT/COTEC, 2010. *Análisis de patrones en el proceso de transferencia de conocimiento y tecnología*, Available at: https://icono.fecyt.es/informesypublicaciones/Paginas/Transferencia-Tecnologica.aspx.

Feng, L., Li, Z. & Swenson, D.L., 2016. The connection between imported intermediate inputs and exports: Evidence from Chinese firms. *Journal of International Economics*, 101, pp.86–101. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0022199616300393.

Fernández Ribas, A. & Catalán, P., 2010. Apoyo público a la innovación desde diferentes niveles de gobierno. In L. Sanz Menéndez & L. Cruz Castro, eds. *Análisis sobre ciencia e innovación en España*. Consejo Superior de Investigaciones Científicas – CSIC y FECYT, pp. 735–761.

Feser, E.J., 2001. A flexible test for agglomeration economies in two US manufacturing industries. *Regional Science and Urban Economics*, 31(1), pp.1–19. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166046200000570.

Fontana, R., Geuna, A. & Matt, M., 2006. Factors affecting university–industry R&D projects: The importance of searching, screening and signalling. *Research Policy*, 35(2), pp.309–323. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733305002210.

Fossas-Olalla, M. et al., 2015. Product innovation: When should suppliers begin to collaborate? *Journal of Business Research*, 68(7), pp.1404–1406. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296315000363.

Fundación SEPI, 2017. FUNDACIÓN SOCIEDAD ESTATAL DE PARTICIPACIONES INDUSTRIALES (SEPI). Available at: http://www.fundacionsepi.es/indice.asp.

García-Manjón, J.V. & Romero-Merino, M.E., 2010. Efectos de la inversión en I+D sobre el crecimiento empresarial. Revista de Globalización, Competitividad y Gobernabilidad. , 4, (2), pp.16–27. Available at: http://dialnet.unirioja.es/servlet/articulo?codigo=3719450.

García, E., Gil, J. & Rodríguez, G., 2000. *Análisis Factorial*, Madrid: Edit. LA MURALLA, S.A.

García Vega, M. & Huergo, E., 2010. Transferencia de tecnología de las multinacionales. In L. Sanz Menéndez & L. Cruz Castro, eds. *Análisis sobre ciencia e innovación en España*. Consejo Superior de Investigaciones Científicas – CSIC y FECYT, pp. 533–564.

Gilsing, V. et al., 2008. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy*, 37(10), pp.1717–1731. Available at: http://linkinghub.elsevier.com/retrieve/pii/S004873330800190X.

Goldberg, P.K. et al., 2010. Imported intermediate inputs and domestic product growth: Evidence from India. *The Quarterly Journal of Economics*, 125(4), pp.1727–1767. Available at: https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2010.125.4.1727.

González Bañales, D., 2007. *La influencia de la innovación tecnológica, la orientación al mercado y el capital relacional en los resultados de las empresas de un sector de alta tecnología. Aplicación a la industria del software de México*. Universitat Politècnica de València. Available at: http://dialnet.unirioja.es/servlet/tesis?codigo=17936.

Grimpe, C. & Sofka, W., 2009. Search patterns and absorptive capacity: Low- and high-technology sectors in European countries. *Research Policy*, 38(3), pp.495–

506. Available at: http://www.sciencedirect.com/science/article/pii/S0048733308002254.

Gumbau-Albert, M. & Maudos, J., 2006. Technological activity and productivity in the Spanish regions. *The Annals of Regional Science*, 40(1), pp.55–80. Available at: http://link.springer.com/10.1007/s00168-005-0027-5.

Guo, B., Gao, J. & Chen, X., 2013. Technology strategy, technological context and technological catch-up in emerging economies: Industry-level findings from Chinese manufacturing. *Technology Analysis and Strategic Management*, 25(2), pp.219–234. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84873323381&partnerID=40&md5=9847240682ce43471887435d26c75bac.

Gupta, S. et al., 2016. Marketing innovation: A consequence of competitiveness. *Journal of Business Research*, 69(12), pp.5671–5681. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296316302776.

Hair, J.F. et al., 1999. *Análisis Multivariante* 5a ed. PRENTICE HALL IBERIA, ed., Madrid.

Hall, M.A., 1999. *Correlation-based feature selection for machine learning*. The University of Waikato. Available at: http://www.cs.waikato.ac.nz/~mhall/thesis.pdf.

Hansen, T. & Winther, L., 2011. Innovation, regional development and relations between high- and low-tech industries. *European Urban and Regional Studies*, 18(3), pp.321–339. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-79961068095&partnerID=40&md5=82d1b7b4000ebfb485c620b2b60152ef.

Hatzichronoglou, T., 1997. *Revision of the HighTechnology Sector and Product Classification*, Paris. Available at: http://dx.doi.org/10.1787/134337307632.

Hecker, A. & Ganter, A., 2016. Organisational and technological innovation and the moderating effect of open innovation strategies. *International Journal of Innovation Management*, 20(2), p.1650019. Available at: http://www.worldscientific.com/doi/abs/10.1142/S1363919616500195.

Heidenreich, M., 2009. Innovation patterns and location of European low- and

medium-technology industries. *Research Policy*, 38(3), pp.483–494. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733308002242.

Hernández Orallo, J., Ramírez Quintana, M.J. & Ferri Ramírez, C., 2004. *Introducción a la minería de datos*, Madrid: Pearson Prentice Hall. Available at: http://users.dsic.upv.es/~jorallo/escrits/escritsa.htm.

Hernández Sampieri, R., Fernández-Collado, C. & Baptista Lucio, P., 2006. *Metodología de la investigación* 4a Edición., México, D.F.: McGraw-Hill.

Herrera, L. & Nieto, M., 2010. Efectos regionales de las ayudas a la I+D empresarial. In L. Sanz Menéndez & L. Cruz Castro, eds. *Análisis sobre ciencia e innovación en España*. Consejo Superior de Investigaciones Científicas – CSIC y FECYT, pp. 709–732.

Hervas-Oliver, J.-L., Ripoll-Sempere, F. & Moll, C.B., 2016. Does management innovation pay-off in SMEs? Empirical evidence for Spanish SMEs. *Small Business Economics*, 47(2), pp.507–533. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84964253185&doi=10.1007%2Fs11187-016-9733-x&partnerID=40&md5=e3d25bebb3b830bf774b89c4d6442ad5.

Hirsch-Kreinsen, H., 2015a. Innovation in Low-Tech Industries: Current conditions and future prospects. In *Low-tech Innovation*. Cham: Springer International Publishing, pp. 17–32. Available at: http://link.springer.com/10.1007/978-3-319-09973-6_2.

Hirsch-Kreinsen, H., 2015b. Patterns of knowledge use in "low-tech" industries. *Prometheus*, 33(1), pp.67–82. Available at: http://www.tandfonline.com/doi/full/10.1080/08109028.2015.1062237.

Homburg, C. et al., 2017. The contingent roles of R&D–sales versus R&D–marketing cooperation in new-product development of business-to-business firms. *International Journal of Research in Marketing*, 34(1), pp.212–230. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167811616300751.

Hopkins, M.M. & Siepel, J., 2013. Just how difficult can it be counting up R&D funding for emerging technologies (and is tech mining with proxy measures going to

be any better)? *Technology Analysis & Strategic Management*, 25(6), pp.655–685. Available at: http://www.tandfonline.com/doi/abs/10.1080/09537325.2013.801950.

Hud, M. & Hussinger, K., 2015. The impact of R&D subsidies during the crisis. *Research Policy*, 44(10), pp.1844–1855. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733315001006.

Huizingh, E.K.R.E., 2011. Open innovation: State of the art and future perspectives. *Technovation*, 31(1), pp.2–9. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497210001100.

Hullova, D., Trott, P. & Simms, C.D., 2016. Uncovering the reciprocal complementarity between product and process innovation. *Research Policy*, 45(5), pp.929–940. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733316300051.

Hurley, R.F. & Hult, G.T.M., 1998. Innovation, market orientation, and organizational learning: An integration and empirical examination. *Journal of Marketing*, 62(3).

INE, 2012a. *CNAE-2009*,

INE, 2014. *Indicadores de alta tecnología 2012. Resultados definitivos. Boletín Notas de prensa*, Available at: http://www.ine.es/prensa/np831.pdf.

INE, 2005. *Indicadores de ciencia y tecnología 2003*, Madrid: INE. Servicio Editorial.

INE, 2016. Indicadores del sector de alta tecnología, Años 2014. Resultados definitivos. , p.8. Available at: http://www.ine.es/prensa/np959.pdf.

INE, 2013a. *Indicadores del sector de alta tecnología Año 2011. Resultados definitivos*, Madrid. Available at: http://www.ine.es/prensa/np768.pd.

INE, 2011. Instituto Nacional de Estadística -INE 2011. Indicadores del sector de alta tecnología. Metodología General. , p.17. Available at: http://www.ine.es/daco/daco43/notaiat.pdf.

INE, 2009. *Introducción a la CNAE-2009*,

INE, 2013b. *Notas de prensa-NP794*, Available at: http://www.ine.es/prensa/np794.pdf.

INE, 2012b. *Notas de prensa -NP732*, Available at: http://www.ine.es/prensa/np732.pdf.

Instituto de Estudios Fiscales (IEF), 2014. *La evaluación ex ante del Programa Operativo de Crecimiento Inteligente (POCInt)*, Available at: http://www.dgfc.sgpg.meh.es/sitios/dgfc/es-ES/ipr/fcp1420/p/Prog_Op_Plurirregionales/Documents/Eval_ex_ante_POCINT.pdf.

Jamasb, T. & Pollitt, M., 2008. Liberalisation and R&D in network industries: The case of the electricity industry. *Research Policy*, 37(6–7), pp.995–1008. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733308000851.

Jeannerat, H. & Crevoisier, O., 2016. Editorial: From "Territorial Innovation Models" to "Territorial Knowledge Dynamics": On the Learning Value of a New Concept in Regional Studies. *Regional Studies*, 50(2), pp.185–188. Available at: http://www.tandfonline.com/doi/full/10.1080/00343404.2015.1105653.

Jericó, J.P. & Marsili, M., 2016. Towards a statistical mechanics of economies. *The European Physical Journal Special Topics*, 225(17–18), pp.3211–3224. Available at: http://link.springer.com/10.1140/epjst/e2016-60237-7.

Jin, M. & Zhang, X., 2009. Analysis and assessment on risks of enterprise-customer collaborative innovation. In *2009 International Conference on Management and Service Science*. IEEE, pp. 1–4. Available at: http://ieeexplore.ieee.org/document/5300926/.

De Jong, J.P.J. & Freel, M., 2010. Absorptive capacity and the reach of collaboration in high technology small firms. *Research Policy*, 39(1), pp.47–54. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733309002005.

Jovell, A.J., 1995. *Análisis de regresión logística* Cuadernos., Madrid: CENTRO DE INVESTIGACIONES SOCIOLÓGICAS.

Kleinbaum, D.G., 1998. *Logistic regression : a self-learning text*, New York: Springer.

Lagendijk, A., 2006. Learning from conceptual flow in regional studies: Framing present debates, unbracketing past debates. *Regional Studies*, 40(4), pp.385–399. Available at: http://www.tandfonline.com/doi/abs/10.1080/00343400600725202.

Lakshmi, B.N. & Raghunandhan, G.H., 2011. A conceptual overview of data mining. In *2011 National Conference on Innovations in Emerging Technology*. IEEE, pp. 27–32. Available at: http://ieeexplore.ieee.org/document/5738828/.

Lawson, C. & Lorenz, E., 1999. Collective learning, tacit knowledge and regional innovative capacity. *Regional Studies*, 33(4), pp.305–317. Available at: http://www.tandfonline.com/doi/abs/10.1080/713693555.

Lechler, T.G. & Thomas, J.L., 2015. Examining new product development project termination decision quality at the portfolio level: Consequences of dysfunctional executive advocacy. *International Journal of Project Management*, 33(7), pp.1452–1463. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0263786315000708.

Lee, C.-Y., 2010. A theory of firm growth: Learning capability, knowledge threshold, and patterns of growth. *Research Policy*, 39(2), pp.278–289. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733309002236.

Leydesdorff, L., Rotolo, D. & de Nooy, W., 2013. Innovation as a nonlinear process, the scientometric perspective, and the specification of an "innovation opportunities explorer." *Technology Analysis & Strategic Management*, 25(6), pp.641–653. Available at: http://www.tandfonline.com/doi/abs/10.1080/09537325.2013.801948.

Limonge, W., Bergamasco, A. & de Souza Damiani, J.H., 2007. A case study of the application of `Strategic Management of Technology and Innovation' concepts at Atech - Tecnologias Críticas, a firm of the Brazilian Aero-Spatial cluster. In *PICMET '07 - 2007 Portland International Conference on Management of Engineering & Technology*. IEEE, pp. 669–679. Available at: http://ieeexplore.ieee.org/document/4349383/.

Liu, B., Hsu, W. & Ma, Y., 1998. Integrating classification and association rule mining. *KDD-98 Proceedings*, pp.80–86.

Lloret-Segura, S. et al., 2014. El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3). Available at: http://revistas.um.es/analesps/article/view/199361.

LÖFQVIST, L., 2017. Product innovation in small companies: managing resource scarcity through financial bootstrapping. *International Journal of Innovation Management*, 21(2), p.1750020. Available at: http://www.worldscientific.com/doi/abs/10.1142/S1363919617500207.

Lööf, H. & Broström, A., 2006. Does knowledge diffusion between university and industry increase innovativeness? *The Journal of Technology Transfer*, 33(1), pp.73–90. Available at: http://link.springer.com/10.1007/s10961-006-9001-3.

Lucheng, H., Yanhua, Y. & Zhihua, Z., 2010. A study on the application of Data Mining in the Patent information analysis for company. In *2010 Second International Workshop on Education Technology and Computer Science*. IEEE, pp. 618–622. Available at: http://ieeexplore.ieee.org/document/5458763/.

Macneill, S. & Jeannerat, H., 2016. Beyond production and standards: Toward a status market approach to territorial innovation and knowledge policy. *Regional Studies*, 50(2), pp.245–259. Available at: http://www.tandfonline.com/doi/full/10.1080/00343404.2015.1019847.

Maranzato, F.P. et al., 2016. Evaluation of integration models in intra-firm technology transfer between research and development. In *IAMOT 2016 - 25th International Association for Management of Technology Conference, Proceedings: Technology - Future Thinking*. pp. 850–869. Available at: https://www.scopus.com/record/display.uri?eid=2-s2.0-84988391866&origin=resultslist&sort=plf-f&src=s&st1=%22Evaluation+of+integration+models+in+intra-firm+technology+transfer+between+research+and+development%22&st2=&sid=F69087D383393E28D6E09C962FD0594F.ws.

Marek, P. et al., 2016. R&D collaborations and the role of proximity. *Regional Studies*, pp.1–13. Available at: https://www.tandfonline.com/doi/full/10.1080/00343404.2016.1242718.

Martín, Q., Cabero, M.T. & De Paz, Y. del R., 2008. *Tratamiento estadístico de datos*

con SPSS, Madrid: International Thomson.

Mas-Verdú, F., Ortiz-Miranda, D. & García-Álvarez-Coque, J.M., 2016. Examining organizational innovations in different regional settings. *Journal of Business Research*, 69(11), pp.5324–5329. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296316303368.

Mastrocinque, E., Mondragon, A.E.C. & Hogg, P.J., 2016. Manufacturing technology selection in the supply chain context by means of fuzzy-AHP: A case in the high performance textile industry. *ARPN Journal of Engineering and Applied Sciences*, 11(1).

Ministerio De Economía Y Hacienda, 2007. Real Decreto 475/2007, de 13 de abril, por el que se aprueba la Clasificación Nacional de Actividades Económicas 2009 (CNAE-2009). , p.22. Available at: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-8824.

Mishra, P., 2010. R&D efforts by Indian Pharmaceutical Firms in the new patent regime. *South East European Journal of Economics and Business*, 5(2). Available at: http://www.degruyter.com/view/j/jeb.2010.5.issue-2/v10033-010-0018-z/v10033-010-0018-z.xml.

Mortara, L. & Minshall, T., 2011. How do large multinational companies implement open innovation? *Technovation*, 31(10–11), pp.586–597. Available at: http://linkinghub.elsevier.com/retrieve/pii/S016649721100071X.

Moulaert, F. & Sekia, F., 2003. Territorial Innovation Models: A critical survey. *Regional Studies*, 37(3), pp.289–302. Available at: http://www.tandfonline.com/doi/abs/10.1080/0034340032000065442.

Na-Allah, A. & Muchie, M., 2012. Social absorption capability, systems of innovation and manufactured export response to preferential trade incentives. *Research Policy*, 41(1), pp.93–101. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733311001156.

Nørskov, S., Chrysochou, P. & Milenkova, M., 2015. The impact of product innovation attributes on brand equity. *Journal of Consumer Marketing*, 32(4), pp.245–254. Available at: http://www.emeraldinsight.com/doi/10.1108/JCM-10-2014-1198.

Nunes, P.M., Serrasqueiro, Z. & Leitão, J., 2012. Is there a linear relationship between R&D intensity and growth? Empirical evidence of non-high-tech vs. high-tech SMEs. *Research Policy*, 41(1), pp.36–53. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-81855199779&partnerID=40&md5=272aa2ee45467816381466314 4cc7ba4.

Observatorio Español de I+D+I (ICONO), 2013. *Indicadores del sistema español de ciencia, tecnología e innovación: 2013*, Madrid: Fundación Española para la Ciencia y la Tecnología (FECYT).

Observatorio Español de I+D+I (ICONO), 2014. *Indicadores del Sistema Español de Ciencia, Tecnología e Innovación 2014* FECYT, ed., Available at: https://icono.fecyt.es/informesypublicaciones/Documents/Indicadores SECTI-Web.pdf.

Observatorio Español de I+D+I (ICONO), 2012. *Indicadores del Sistema Español de Ciencia, Tecnología e Innovación 2014* 2016th ed. FECYT, ed., Madrid: Fundación Española para la Ciencia y la Tecnología (FECYT). Available at: https://icono.fecyt.es/informesypublicaciones/Documents/Libro-de-Indicadores_2012.pdf.

Observatorio Español de I+D+I (ICONO), 2015. *Indicadores del Sistema Español de Ciencia, Tecnología e Innovación 2015* FECYT, ed., Available at: https://icono.fecyt.es/informesypublicaciones/Documents/Indicadores SECTI 2015.pdf.

Observatorio Español de I+D+I (ICONO), 2016. *Indicadores del Sistema Español de Ciencia, Tecnología e Innovación 2016* 2016th ed. FECYT, ed., Available at: https://icono.fecyt.es/informesypublicaciones/Documents/Indicadores SECYT2016.pdf.

OCDE, 2009. Manual de estadisticas de patentes de la OCDE. , p.179.

OCDE/ FECYT, 2007. *I+D e Innovación en España: mejorando los instrumentos.*, Madrid.

OCDE / FECYT, 2003. *Manual de Frascati 2002. Medición de las actividades científicas y tecnológicas : propuesta de una norma práctica para encuestas*

de investigación y desarrollo experimental, Madrid: Fundación Española Ciencia y Tecnología. Available at: http://www.idi.mineco.gob.es/stfls/MICINN/Investigacion/FICHEROS/ManuaFrascati-2002_sp.pdf.

OECD, 2015a. *Frascati Manual 2015*, OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en.

OECD, 2011a. *Main Science and Technology Indicators. Principaux indicateurs de la science et de la technologie*, Available at: http://www.oecd.org/science/inno/49501885.pdf.

OECD, 2015b. *OECD Digital Economy Outlook 2015*, OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-outlook-2015_9789264232440-en.

OECD, 1999. *OECD Science, Technology and Industry Scoreboard 1999: Benchmarking Knowledge-based Economies* OECD Publishing, ed., Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/sti_scoreboard-1999-en.

OECD, 2001. *OECD Science, Technology and Industry Scoreboard 2001: Towards a Knowledge-based Economy*, Paris: OECD Publishing.

OECD, 2003. *OECD Science, Technology and Industry Scoreboard 2003*, Paris: OECD Publishing.

OECD, 2005a. *OECD Science, Technology and Industry Scoreboard 2003 (Summary in Spanish)*, Paris. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2003-sum-es.

OECD, 2005b. *OECD Science, Technology and Industry Scoreboard 2005*, Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2005-en.

OECD, 2005c. *OECD Science, Technology and Industry Scoreboard 2005 (Summary in Spanish)*, Paris. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2005-sum-es.

OECD, 2007a. *OECD Science, Technology and Industry Scoreboard 2007*, Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2007-en.

OECD, 2007b. *OECD Science, Technology and Industry Scoreboard 2007 (Summary in Spanish)*, Paris. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2007-sum-es.

OECD, 2009a. *OECD Science, Technology and Industry Scoreboard 2009*, Paris: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2009_sti_scoreboard-2009-en.

OECD, 2009b. *OECD Science, Technology and Industry Scoreboard 2009 (Summary in Spanish)*, Paris. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2009-sum-es.

OECD, 2011b. *OECD Science, Technology and Industry Scoreboard 2011*, Paris: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2011_sti_scoreboard-2011-en.

OECD, 2011c. *OECD Science, Technology and Industry Scoreboard 2011 (Summary in Spanish)*, Paris. Available at: http://dx.doi.org/10.1787/sti_scoreboard-2011-sum-es.

OECD, 2013. *OECD Science, Technology and Industry Scoreboard 2013: Innovation for Growth (Summary in Spanish)*, Paris: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2013_sti_scoreboard-2013-en.

OECD, 2015c. *OECD Science, Technology and Industry Scoreboard 2015*, Paris: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2015_sti_scoreboard-2015-en.

OECD, 2015d. *OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society (Summary in Spanish)*, Paris: OECD Publishing.

Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-industry-scoreboard-2015_sti_scoreboard-2015-en.

OECD, 2005d. *Oslo Manual* 3rd Editio., Paris: OECD Publishing. Available at: http://www.oecd-ilibrary.org/science-and-technology/oslo-manual_9789264013100-en.

OECD/Eurostat, 2007. *Manual de Oslo: Guía para la recogida e interpretación de datos sobre innovación. Tercera edición* 3a Edición. GRUPO TRAGSA, ed., Madrid: Tragsa. Available at: http://www.oecd-ilibrary.org/science-and-technology/manual-de-oslo_9789264065659-es.

Pardo Merino, A. & Ruíz Díaz, M.Á., 2002. *SPSS 11 : guía para el análisis de datos*, Madrid: McGraw-Hill Interamericana de España.

Paunov, C., 2013. Innovation and Inclusive development: A discussion of the main policy issues. *OECD Science, Technology and Industry Working Papers 2013/01*, p.62. Available at: http://dx.doi.org/10.1787/5k4dd1rvsnjj-en.

Pavitt, K., 1984. Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13(6), pp.343–373. Available at: http://linkinghub.elsevier.com/retrieve/pii/0048733384900180.

Pérez Hoyos, S., 1996. *Introducción a la regresión logística* Quaderns d., Valencia: Institut Valencià d'Estudis en Salut Pública.

Pérez López, C. & Santín González, D., 2007. *Minería de datos : técnicas y herramientas.*, Madrid: Thomson.

PITEC, 2011. PITEC 2009 : La innovación en la empresa española en el marco de la Estrategia Estatal de Innovación.

Plotnikova, M., Romero, I. & Martínez-Román, J.A., 2016. Process innovation in small businesses: the self-employed as entrepreneurs. *Small Business Economics*, 47(4), pp.939–954. Available at: http://link.springer.com/10.1007/s11187-016-9743-8.

Rio, R. et al., 2011. Working in open innovation: How to determine networks and their relationship using techmining. In *2011 Atlanta Conference on Science and*

Innovation Policy. IEEE, pp. 1–4. Available at: http://ieeexplore.ieee.org/document/6064484/.

Ritala, P. & Sainio, L.-M., 2014. Coopetition for radical innovation: technology, market and business-model perspectives. *Technology Analysis & Strategic Management*, 26(2), pp.155–169. Available at: http://www.tandfonline.com/doi/full/10.1080/09537325.2013.850476.

Robertson, P., Smith, K. & von Tunzelmann, N., 2009. Innovation in low- and medium-technology industries. *Research Policy*, 38(3), pp.441–446. Available at: http://www.sciencedirect.com/science/article/pii/S0048733308002199.

Roelofs, M.R., Østbye, S.E. & Heen, E.E., 2016. Asymmetric firms, technology sharing and R&D investment. *Experimental Economics*. Available at: http://link.springer.com/10.1007/s10683-016-9500-5.

Rua, O.L. & Correia, M.J., 2015. Innovation in healthcare organizations: Empirical evidence from Portugal. In *Handbook of Research on Internationalization of Entrepreneurial Innovation in the Global Economy*. IGI Global, pp. 134–157. Available at: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8216-0.ch007.

Ryu, H.-S. & Lee, J.-N., 2016. Innovation patterns and their effects on firm performance. *The Service Industries Journal*, 36(3–4), pp.81–101. Available at: http://www.tandfonline.com/doi/full/10.1080/02642069.2016.1155114.

Ryzhkova, N., 2015. Does online collaboration with customers drive innovation performance? *Journal of Service Theory and Practice*, 25(3), pp.327–347. Available at: http://www.emeraldinsight.com/doi/10.1108/JSTP-02-2014-0028.

de Saá Pérez, P. & Díaz Díaz, N.L., 2007. Incidencia de los Recursos Humanos de I+D internos y contratados en la innovación. *Cuadernos de Economía y Dirección de la Empresa*, 33, pp.7–30. Available at: http://www.redalyc.org/pdf/807/80703301.pdf.

Sáenz, J., Aramburu, N. & Rivera, O., 2009. Knowledge sharing and innovation performance N. Bontis, ed. *Journal of Intellectual Capital*, 10(1), pp.22–36. Available at:

http://www.emeraldinsight.com/doi/10.1108/14691930910922879.

Sáez, N.M. & Pascual, M.C.O., 2013. Retail marketing innovation in Spain. *International Journal of Innovation and Learning*, 14(3/4), p.453. Available at: http://www.inderscience.com/link.php?id=56238.

Salzberg, S.L., 1994. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning*, 16(3), pp.235–240. Available at: http://link.springer.com/10.1007/BF00993309.

Santamaría, L., Nieto, M.J. & Barge-Gil, A., 2009. Beyond formal R&D: Taking advantage of other sources of innovation in low- and medium-technology industries. *Research Policy*, 38(3), pp.507–517. Available at: http://www.sciencedirect.com/science/article/pii/S0048733308002229.

Dos Santos, I.C. & Neto, J.A., 2007. Knowledge management for High Technology industries. In *PICMET '07 - 2007 Portland International Conference on Management of Engineering & Technology*. IEEE, pp. 1019–1031. Available at: http://ieeexplore.ieee.org/document/4349423/.

Santos, L. & Vázquez, R., 1997. *La innovación en las empresas de alta tecnología: Factores Condicionantes Del Resultado Comercial*, Oviedo: Universidad de Oviedo. Facultad de Ciencias Económicas y Empresariales.

Schmidt, J.B., Sarangee, K.R. & Montoya, M.M., 2009. Exploring new product development project review practices. *Journal of Product Innovation Management*, 26(5), pp.520–535. Available at: http://doi.wiley.com/10.1111/j.1540-5885.2009.00678.x.

Schmiele, A., 2012. Drivers for international innovation activities in developed and emerging countries. *The Journal of Technology Transfer*, 37(1), pp.98–123. Available at: http://link.springer.com/10.1007/s10961-011-9221-z.

Schreiber, D. et al., 2016. Analysis of innovation and its environmental impacts on the Chemical Industry. *BAR - Brazilian Administration Review*, 13(1), pp.56–75. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-76922016000100005&lng=en&nrm=iso&tlng=en.

Segarra-Blasco, A., Garcia-Quevedo, J. & Teruel-Carrizosa, M., 2008. Barriers to

innovation and public policy in Catalonia. *International Entrepreneurship and Management Journal*, 4(4), pp.431–451. Available at: http://link.springer.com/10.1007/s11365-008-0086-z.

Seng, J.-L. & Chen, T.C., 2010. An analytic approach to select data mining for business decision. *Expert Systems with Applications*, 37(12), pp.8042–8057. Available at: http://linkinghub.elsevier.com/retrieve/pii/S095741741000494X.

Shipton, H. et al., 2006. HRM as a predictor of innovation. *Human Resource Management Journal*, 16(1), pp.3–27. Available at: http://doi.wiley.com/10.1111/j.1748-8583.2006.00002.x.

Shipton, H. et al., 2005. Managing people to promote innovation. *Creativity and Innovation Management*, 14(2), pp.118–128. Available at: http://doi.wiley.com/10.1111/j.1467-8691.2005.00332.x.

Sociedad Asturiana de Estudios Económicos - SADEI, 2008. La clasificación nacional de actividades económicas (CNAE-2009) en las cuentas regionales de Asturias. *Revista Asturiana de Economía -RAE*, No 42, pp.163–172. Available at: https://dialnet.unirioja.es/servlet/articulo?codigo=3680515.

Stam, E. & Wennberg, K., 2009. The roles of R&D in new firm growth. *Small Business Economics*, 33(1), pp.77–89. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-67349156210&partnerID=40&md5=5b981f85658484c13a634fb89cd5b844.

Sterlacchini, A. & Venturini, F., 2014. R&D and Productivity in High-Tech Manufacturing: A Comparison between Italy and Spain. *Industry and Innovation*, 21(5), pp.359–379. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84911808450&doi=10.1080%2F13662716.2014.959315&partnerID=40&md5=90a1e79f4e6091323eaa0931953ce9a3.

Su, C.-T., Chen, Y.-H. & Sha, D.Y., 2006. Linking innovative product development with customer knowledge: a data-mining approach. *Technovation*, 26(7), pp.784–795. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497205000799.

Surroca, J. & Sánchez, L.S., 2006. *La cooperación tecnológica como determinante de los resultados empresariales*, Madrid. Available at: http://e-archivo.uc3m.es/bitstream/handle/10016/32/db060101.pdf;jsessionid=4F1E6E2BDC0876F2F24ED5A80E53314C?sequence=1.

Tether, B.S., 2002. Who co-operates for innovation, and why. *Research Policy*, 31(6), pp.947–967. Available at: http://linkinghub.elsevier.com/retrieve/pii/S004873330100172X.

Tirole, J., 1990. *La teoría de la organización industrial*, Barcelona: Ariel.

Uriel, E., 1995. *Análisis de datos : series temporales y análisis multivariante*, Madrid: Editorial AC.

Vega Jurado, J., Gutierrez Gracia, A. & Fernández De Lucio, I., 2010. Generación y adquisición de conocimiento como estrategias de innovación. En: Análisis sobre ciencia e innovación en España. In L. Sanz Méndez & L. Cruz Castro, eds. *Análisis sobre ciencia e innovación en España*. Madrid: Consejo Superior de Investigaciones Científicas – CSIC y FECYT, pp. 502–530. Available at: http://icono.fecyt.es/informesypublicaciones/Paginas/analisis_ciencia_innovacion_espana.aspx.

Wang, C.-H. et al., 2013. R&D, productivity, and market value: An empirical study from high-technology firms. *Omega*, 41(1), pp.143–155. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84866034118&partnerID=40&md5=38f7b8674141a78f25da9483be7003fd.

Wang, C.H. et al., 2014. Innovating brand strategy development using integration of data mining and the CBBE model. *International Journal of Innovation and Learning*, 16(4), p.353. Available at: http://www.inderscience.com/link.php?id=65543.

Wang, H.-W. & Wu, M.-C., 2012. Business type, industry value chain, and R&D performance: Evidence from high-tech firms in an emerging market. *Technological Forecasting and Social Change*, 79(2), pp.326–340. Available at: http://www.sciencedirect.com/science/article/pii/S0040162511001211.

Wang, Y. & Yuan, R., 2011. Data Mining application in the innovative activities of

the science and technology. In *2011 International Conference on Management and Service Science*. IEEE, pp. 1–3. Available at: http://ieeexplore.ieee.org/document/5999195/.

Witten, I.H. & Frank, E., 2005. *Data Mining: Practical machine learning tools and techniques* Second Edi., San Francisco: Morgan Kaufmann Publishers.

Wöhrl, R., Hüsig, S. & Dowling, M., 2009. The interaction of R&D intensity and firm age: Empirical evidence from technology-based growth companies in the German "Neuer Markt." *The Journal of High Technology Management Research*, 20(1), pp.19–30. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1047831009000030.

Yang, Y. et al., 2012. What are the effective strategic orientations for new product success under different environments? An empirical study of Chinese businesses. *Journal of Product Innovation Management*, 29(2), pp.166–179. Available at: http://doi.wiley.com/10.1111/j.1540-5885.2011.00900.x.

Zanello, G. et al., 2016. The creation and diffusion of innovation in developing countries: a systematic literature review. *Journal of Economic Surveys*, 30(5), pp.884–912. Available at: http://doi.wiley.com/10.1111/joes.12126.

Zheng, G., Guo, Y. & Wang, Y., 2016. Non-R&D Innovation Patterns in Chinese SMEs: An Empirical Study from Zhejiang Province. *International Journal of Innovation and Technology Management*, 13(6), p.1640014. Available at: http://www.worldscientific.com/doi/abs/10.1142/S0219877016400149.

Zobel, A.-K., Lokshin, B. & Hagedoorn, J., 2017. Formal and informal appropriation mechanisms: The role of openness and innovativeness. *Technovation*, 59, pp.44–54. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497216303339.

Zuniga, P. & Crespi, G., 2013. Innovation strategies and employment in Latin American firms. *Structural Change and Economic Dynamics*, 24(1), pp.1–17. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0954349X12000872.

GLOSARIO DE SIGLAS

CNAE: Clasificación Nacional de Actividades Económicas (España).

DIRCE: Directorio Central de Empresas.

EESE: Encuesta de Estrategias Empresariales.

EUROSTAT: Oficina Estadística de las Comunidades Europeas.

FECYT: Fundación Española para la Ciencia y Tecnología.

INE: Instituto Nacional de Estadística (España).

I+D+I: Investigación y Desarrollo e Innovación.

KKD: Proceso de extracción de conocimiento a partir de datos.

NACE: Nomenclatura Estadística de Actividades Económicas de la Unión Económica Europea.

OCDE: Organización para la Cooperación y el Desarrollo Económicos.

OECD: Organisation for Economic Co-operation and Development.

OPIs: Organismos Públicos de Investigación.

SEPI: Sociedad Estatal de Participaciones Industriales.