

# **PITEC**

## **Panel de innovación tecnológica**

### **Metodología**

**Madrid, Junio 2017**

# Índice

---

1	Introducción	3
2	Datos, actualización y accesibilidad	3
3	Muestras	4
4	Ficheros	5
4.1	Diseño	5
4.2	Identificadores	6
4.3	Variables	6
	ANEXO 1. Método de anonimización	8
	ANEXO 2. Caso práctico del método de anonimización	10
	ANEXO 3. Muestra de empresas	26

---

---

## 1 Introducción

El panel de innovación tecnológica (PITEC) es un instrumento estadístico para el seguimiento de las actividades de innovación tecnológica de las empresas residentes en España. Iniciado en 2004, el objetivo final de este proyecto es contribuir a mejorar la información estadística disponible sobre las actividades tecnológicas de las empresas y las condiciones para la realización de investigaciones científicas sobre las mismas.

PITEC está compuesto por datos de panel. Las estadísticas de panel consisten en observaciones repetidas a lo largo del tiempo de las unidades económicas incluidas en las muestras. Son capaces de producir estimaciones mucho más precisas de los cambios temporales (como, por ejemplo, la importancia del inicio de actividades de innovación, la evolución de la composición de las mismas y del propio gasto en innovación) así como, apreciar la heterogeneidad en las decisiones adoptadas por las empresas (como por ejemplo, las distintas composiciones del gasto total en gastos en I+D interna y externa) o sus efectos (como por ejemplo, los distintos impactos en la productividad).

---

## 2 Datos, actualización y accesibilidad

La base de datos generada se encuentra a disposición de los investigadores en la página web del INE. El investigador que desee tener acceso a los datos debe firmar un documento de aceptación de condiciones en las que se compromete a llevar a cabo un uso adecuado del fichero de microdatos. Una vez aceptadas las condiciones de acceso podrá descargarse la base de datos de la página web.

Los datos se encuentran recogidos en un sistema de ficheros coordinados, a razón de un fichero por año, cubierto por el panel (por el momento, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014 y 2015). Las variables disponibles en este panel son seleccionadas a partir de la Encuesta sobre Innovación en las Empresas y Estadística sobre Actividades en I+D. Cada año se añade el fichero correspondiente a la última recogida de datos. Los ficheros de datos de cada año están disponibles en cinco formatos: (i) Ficheros texto de ancho fijo sin delimitadores, (ii) Ficheros texto con delimitadores (iii) Ficheros en formato SAS, (iv) Ficheros en formato STATA, (v) Ficheros en formato SPSS, y (vi) Ficheros en formato R. En el caso del fichero con delimitador el delimitador utilizado es “;”.

Los ficheros accesibles en la web se encuentran anonimizados, de manera que no permitan la identificación de las empresas que componen el panel. El proceso de anonimización introduce las siguientes modificaciones con respecto a los valores originales de las variables: a) Reemplaza las observaciones individuales de 5 variables cuantitativas (Cifra de negocios, Inversión Bruta en bienes materiales, Número de empleados, Gastos totales en innovación y Personal total en I+D) por datos generados mediante un proceso de perturbación de los valores originales; b) Reemplaza las observaciones individuales del resto de variables cuantitativas por porcentajes referidos al valor agregado (por ejemplo, la Cifra de gastos internos en I+D se reemplaza por el porcentaje que los gastos de I+D representan como

proporción de los Gastos totales en innovación); c) Reemplaza las actividades originales por una agrupación en 56 actividades hasta 2008 y en 44 actividades a empresas partir de 2008 (para el año 2008 se proporcionan ambas agrupaciones); d) Censura el contenido de las variables correspondientes a un número dado de empresas. Los detalles de la anonimización" de los datos se encuentran descritos en el Anexo 1, de este documento.

*Nota: Acceso a los datos no anonimizados*

El uso de los datos accesibles a través de la web puede ser suficiente para muchos trabajos estadísticos y econométricos. Dadas las técnicas empleadas en el proceso de anonimización, los sesgos esperables son en cualquier caso pequeños. Sin embargo, el investigador puede desear aplicar los programas desarrollados, a los datos no perturbados ni censurados por la anonimización. Dada la igualdad de los ficheros, esto podrá ser realizado con facilidad en el lugar y bajo las condiciones señaladas por el INE, siempre y cuando su uso se justifique en proyectos de investigación científica. Para ello, habrá que realizar una solicitud al INE y rellenar la documentación en la que se explique el proyecto de investigación que se quiere llevar a cabo. Dicha solicitud estará sujeta a un estudio de viabilidad, y en caso de ser aprobada, requerirá la firma tanto de la entidad investigadora, como de todos los miembros del equipo de investigación, comprometiéndoles a salvaguardar el secreto estadístico.

---

### **3 Muestras**

Por razones de oportunidad y viabilidad, el panel se inició con datos de dos muestras correspondientes al año 2003, una compuesta por las empresas de 200 o más trabajadores (muestra de empresas grandes, MEG, cuya representatividad se evaluó, con el DIRCE, en un 73% del total) y otra, por empresas con gasto en I+D interna (muestra MID). Esta muestra experimentó ampliaciones en los años 2004 y 2005 debidas a los progresos informativos realizados por el INE sobre las empresas con actividades de I+D interna.

En el año 2004 se incluyeron también una muestra de las empresas con menos de 200 trabajadores que tienen gastos por compra de servicios de I+D (I+D externa) pero que no realizan I+D interna (muestra MIDE) y una submuestra representativa de empresas con menos de 200 trabajadores sin gastos en innovación (muestra MEP).

Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel, y dicha caracterización no variará, en los siguientes años. El investigador podrá obtener las muestras de los ficheros, codificadas del 1 al 5 de la siguiente manera:

1. Empresas de más de 200 trabajadores.
2. Empresas que realizan I+D interna.
3. Empresas de más de 200 trabajadores y que realizan I+D interna.
4. Empresas de menos de 200 trabajadores que compran I+D externa pero no realizan I+D interna.

5. Empresas de menos de 200 trabajadores, que no realizan ningún gasto en Innovación.

Cada año se añadirá información a la tabla que detalla la evolución de la muestra del panel en todos los años.

La Tabla Abis (Ver anexo 3.), resume la composición del panel en cada año. Esta tabla proporciona los números de empresas que el investigador hallará en los ficheros y ayuda a evaluar el grado de representatividad que mantiene la muestra sobre las poblaciones correspondientes.

La correspondencia entre las 4 muestras y los 5 valores codificados es la siguiente:

- MEG=1
- MID =2
- MEG Y MID=3
- MIDE=4
- MEP=5

En 2014, con el objetivo de reducir carga al informante, manteniendo siempre la utilidad del PITEC para los usuarios, se estableció un procedimiento que permite rotar ciertas unidades manteniéndolas inactivas y, por tanto, no encuestadas en ciertos años. Estas unidades se denominan a efectos de muestra como "durmientes", y se reconocen en función de los valores de una nueva variable denominada CP (Código de Permanencia de las empresas durmientes).

Las empresas que forman parte de dicho subpanel no han sido entrevistadas en el año de referencia y por tanto, no se dispone de la información sobre las mismas. La variable CP toma valor 1 para todas ellas.

*Nota: Utilización de la muestra de empresas con I+D interna*

Los cambios en la composición de la muestra MID como consecuencia de las ampliaciones que tuvieron lugar en los años 2004 y 2005 obligan a interpretar con cautela los análisis temporales de dicha muestra de empresas. En este sentido, la caracterización de la persistencia en la actividad de I+D interna podría verse afectada por una menor probabilidad de persistir haciendo I+D interna de las empresas incorporadas a la muestra.

---

## **4 Ficheros**

---

### **4.1 DISEÑO**

Hay un fichero de datos para cada año  $t$  y para cada formato de datos. El número de filas (o registros) del fichero del año  $t$ , es igual al número de empresas que PITEC ha abarcado en total hasta este momento. La razón es que no se suprime el registro correspondiente a ninguna empresa, aunque haya desaparecido. Esto facilita el uso conjunto de los ficheros. La posición de las empresas se repite los

distintos años y las empresas incorporadas en t se añaden como filas (o registros) al fichero del año t.

El número de columnas (o campos) del fichero del año t es igual al número total de variables incluidas en el PITEC hasta este momento. La posición de las variables se repite en los distintos años y las variables incorporadas se añaden en columnas (campos) al final del fichero.

---

## 4.2 IDENTIFICADORES

Los cinco primeros campos del fichero contienen los siguientes identificadores:

**IDENT:** Código identificador de la empresa. Es un código de orden que va desde el número 1 hasta el número de empresas ya presente en el fichero t.

**INCINE:** Indicador de incidencia. Clasifica a las empresas en las categorías según su respuesta en t o afectación por confidencialidad: cierre temporal, cierre definitivo, los datos de la empresa están contenidos en otra, erróneamente incluida en la muestra, absorción (Ha sido absorbida y ya no existe como tal), fusión (Se ha fusionado y ya no existe como tal), duplicada (Aparece más de una vez en el directorio), escisión total (Se ha dividido y ya no existe como tal), ilocalizable, negativa final (la empresa se ha negado a proporcionar la información, limpio (Sin incidencia), problema de confidencialidad (por cuestiones de confidencialidad los datos de la empresa han sido censurados), durmiente (la empresa no ha sido entrevistada en el año t).

**INCIEMP:** Indicador de incidencia en el empleo. Este indicador da cuenta de los motivos que justifican una tasa de variación anómala en el empleo, que quizá puede repetirse en otras variables. Clasifica las incidencias en las siguientes categorías: Sin incidencia; Empresa perteneciente a sector de alta temporalidad; Empresa absorbente; Cambio de unidad de referencia: empresa a grupo, grupo a empresa; Cambio de actividad o abandono de parte de actividad; Escisión; Empresa restante de un proceso de absorción; Regulación de empleo o fase de liquidación; Fusión; Empresa con personal cedido por otras empresas; Consecuencia de la crisis; Empresa que cede personal a otras empresas.

**CORREMP:** Indicador de corrección en el empleo. Variable que indica que se ha modificado la cifra de empleo en una revisión realizada con posterioridad al cierre del trabajo de campo. Como consecuencia de esta modificación puede haber inconsistencias con la cifra de empleo de años anteriores que no se han revisado.

**MUESTRA:** Indicador de muestra. Identifica la muestra a la que pertenece la empresa: MEG, MID, MEG y MID, MIDE y por último, MEP.

---

## 4.3 VARIABLES

El Diseño de Registro de los ficheros contiene una relación exhaustiva de las variables con sus posibles valores, año de incorporación, disponibilidad, e indicaciones sobre la anonimización. Así mismo, también contiene la información

sobre la posición de inicio de cada variable y el ancho del campo. La grabación de la información hace uso intensivo de la identidad blanco=no información. La interpretación de la información debe basarse en las siguientes reglas:

a) El año en el que se deja de disponer información de una empresa se graba su identificador de empresa e indicador de incidencia (INCINE) y el resto de variables se dejan en blanco. Los años posteriores a su desaparición, se graba el identificador de empresa y el resto de variables aparecen en blanco.

b) Las variables de las empresas con datos censurados por problemas de confidencialidad aparecen en blanco.

c) Las variables de las que no se dispone de información se graban en blanco. La no disponibilidad de información de una variable se puede deber a: i) No se incluye en el cuestionario; ii) Cambios en la forma o en el contenido de las preguntas del cuestionario. Estos cambios pueden llevar a la necesidad de introducir una nueva variable no comparable con la existente; iii) Alternancia en la disponibilidad de información sobre la variable; iv) La empresa no proporciona información sobre la misma.

# ANEXO 1. Método de anonimización

Este método consiste en, además de eliminar las variables de identificación de la empresa, los valores de las variables sensibles y monetarias de cada empresa se ofrecen como porcentaje de otra variable (siempre que sea posible) o como media de los valores de 3 o 5 empresas. El método de “anonimización” implica lo siguiente:

1. Cada registro contiene un número de identificación único vinculado a cada empresa, cuya relación sólo es conocida por el INE. Esto es necesario para poder identificar cada empresa y poder hacer comparaciones en el tiempo.
2. Cada registro contiene la rama de actividad de la empresa como máximo a dos dígitos de la CNAE-93 Rev.1 hasta 2008 o de la CNAE-2009 desde 2008 (en el año 2008 se proporcionan ambas), salvo en algunos casos en los que alguna actividad contiene muchas empresas y se proporciona una desagregación a tres dígitos; y otros casos en los que hay ramas que se reagrupan con otras porque el número de empresas es reducido. El diseño de registro de los ficheros detalla las ramas de actividad consideradas.
3. Las variables que corresponden a datos de identificación de la empresa no se proporcionan. Únicamente, al ser la ubicación de la empresa una de las variables más demandadas por los investigadores, se proporciona la localización de la sede social de la empresa con el desglose Madrid, Cataluña, Andalucía, y resto de España.
4. Las variables cuantitativas del cuestionario “Cifra de negocios”, “Inversión bruta en Bienes materiales”, “Número de empleados”, “Gastos totales en innovación” y “Personal total en I+D”, son reemplazadas por los siguientes valores simulados.

Se proporcionan medias, calculadas de dos maneras diferentes para cada una de las cinco variables anteriores:

- En primer lugar; se ordena la totalidad de los registros por rama de actividad de la empresa (recodificada según se ha explicado en el punto 2), y dentro de cada rama se ordenan las empresas en orden descendente (según el valor de la variable que se va a simular), seleccionando las 5 mayores. Para estas 5 empresas se calcula la media de los valores que toma dicha variable, que será el valor que toma la variable simulada en esos 5 registros. En el caso de que en alguna rama, recodificada a dos dígitos, existiesen menos de 3 empresas, se agrupan con el resto de empresas en el siguiente paso.
- En segundo lugar; se ordenan los registros de las empresas, sin tener en cuenta la actividad, en orden descendente de los valores de la variable que se va a simular (excluidos los registros que se han modificado para la variable en el paso previo). Se agrupan los registros de tres en tres, y se calculan las medias de los 3 valores consecutivos.

Se reemplaza cada valor por las medias correspondientes.

Si sólo quedara una o dos empresas, al dividir por tres el número total de empresas con valores positivos en la variable en cuestión, el último o dos últimos grupos (los de menor valor) contendrán a cuatro empresas. Así, las variables proporcionadas son, en el primer caso, las medias de los valores de las 5 empresas con los mayores

valores en la variable que se simula por rama de actividad recodificada, y en el segundo, las medias de los 3 ó 4 valores consecutivos tras ordenar las empresas según el orden ascendente de la variable considerada.

El orden en que se anonimizan las cinco variables en cuestión (“Cifra de negocios”, “Inversión bruta en bienes materiales”, “Número de empleados”, “Gastos totales en innovación” y “Personal total en I+D”) es intrascendente, ya que se repite el proceso para cada variable, independientemente de los valores que toman las demás. En el Anexo 2 se muestra un ejemplo de la aplicación del método descrito anteriormente.

5. Todas las variables que se pueden dar como porcentaje de otras no hace falta sustituirlas por medias. En concreto, las variables referidas a exportaciones se proporcionan como porcentaje de la cifra de negocios; el personal remunerado, no remunerado y remunerado con educación superior, como porcentaje del número de empleados; los gastos en I+D interna según la naturaleza del gasto, según el origen de fondos y por Comunidades Autónomas, el gasto en I+D en biotecnología, el importe de las becas de investigación y el gasto en consultores externos trabajando “in situ” en I+D interna, se dan como porcentajes del gasto en I+D interna; los gastos en I+D externa según el proveedor se proporcionan como porcentajes del gasto en I+D externa; los gastos en actividades para la innovación tecnológica y los gastos en innovación tecnológica por Comunidades Autónomas, como porcentajes del gasto total de innovación; el personal en I+D interna según su ocupación, su titulación y por Comunidades Autónomas, los becarios en investigación y los consultores externos trabajando “in situ” en I+D interna se facilitan como porcentaje del personal total en I+D interna. Para todas estas variables el porcentaje se calcula respecto a la cifra original y no respecto a la simulada.

6. Para salvaguardar el secreto estadístico, y teniendo en cuenta la estratificación de la muestra, se oculta el contenido de los registros en los que el número de empresas en la población y en el PITEC sea menor o igual a 3. En el caso de que un registro se censure por secreto estadístico un año, dicho registro se mantendrá censurado en años posteriores para evitar posibles identificaciones.

Finalmente, respecto al acceso a los datos no “anonimizados”, se permite a los investigadores, una vez hayan trabajado con el fichero anonimizado y preparado todos sus programas, el acceso al fichero original (sin variables de identificación) en un ‘secure place’. Habida cuenta que la estructura del fichero original es exacta a la del fichero anonimizado, cabe esperar que no se necesite mucho tiempo para ejecutar sus programas y verificar sus conclusiones.

# ANEXO 2. Caso práctico del método de anonimización

Este documento presenta un caso práctico del método de anonimización. Para ello, se utiliza un conjunto formado por 34 empresas ficticias, y con un número reducido de variables:

- IDENT: código identificador de empresa
- ACTIN: Actividad (código CNAE-2009)
- CLASEN: clase de empresa:
  - 1=Pública;
  - 2=Privada sin participación extranjera;
  - 3=Privada con participación < 10% de capital extranjero;
  - 4=Privada con participación  $\geq$  10% y < 50% de capital extranjero;
  - 5=Privada con participación  $\geq$  50% de capital extranjero;
  - 6=Asociación de investigación y otras instituciones de investigación.
- CIFRA: Cifra de negocios
- EXPORTN: Volumen de exportaciones
- INVER: Inversión bruta en bienes materiales
- TAMANO: Número medio de empleados
- GTINN: Gastos totales en innovación tecnológica
- GINTID: Gastos totales en I+D interna

*NOTA: Es importante hacer notar que en este ejemplo no cabe esperar que los valores anonimizados sean similares a los originales (sin anonimizar) por tratarse de un número reducido de registros.*

## Ejemplo: Valores sin anonimizar

Se parte de los 34 registros siguientes:

ident	actin	clasen	cifra	exportn	inver	tamano	gtinn	gintid
1	1310	2	3.635.311	0	0	70	370.999	279.871
2	0111	5	1.790.431	0	5.944.400	35	0	0
3	0811	3	48.664.533	0	0	350	0	0
4	0812	2	5.100.510	81.990	0	30	0	0
5	0520	2	3.339.933	0	18.586	50	8.879	0
6	1395	5	9.545.555	0	25.000	35	255.555	69.530
7	0113	2	2.780.982	648.105	0	17	46.823	39.500
8	1310	4	13.982.590	433.311	376.659	80	0	0
9	1013	2	310.085	51.510	33.090	13	141.414	0
10	0321	2	5.812.616	532.000	0	155	0	0
11	0111	2	393.170	0	0	15	0	0
12	0321	4	4.005.500	0	0	13	0	0
13	1310	4	33.377.750	13.513.522	421.989	148	855.885	0
14	0520	2	4.906.711	0	69.350	55	9.324	0
15	0520	2	1.950.125	0	5.393.000	46	386.525	194.450
16	0812	2	18.205.519	0	0	375	0	0
17	1102	2	21.191.213	966.969	390.033	80	2.079.700	0
18	1330	2	749.965	0	65.000	11	21.955	0
19	1330	2	895.939	0	92.900	15	93.296	0
20	1013	2	8.995.000	0	149.500	67	1.335.575	0
21	1396	5	31.775.250	0	1.310.611	445	0	0
22	0111	2	30.034.563	0	0	240	0	0
23	1396	5	71.708.000	2.689.136	160.999	560	384.538	250.100
24	0210	2	2.571.599	0	500	195	0	0
25	0729	2	9.021.886	5.242.999	18.900	50	91.100	75.825
26	0812	2	6.845.367	0	350.035	45	208.000	0
27	0510	5	16.325.020	0	850	48	53.531	0
28	0510	2	15.045.268	0	207.099	138	0	0
29	1039	2	958.997	0	0	39	120.099	89.523
30	0111	2	28.422.500	284.000	2.357.750	207	254.799	156.395
31	1013	2	2.152.722	151.151	30.700	25	376.787	0
32	0130	3	1.332.522	0	27.500	5	0	0
33	1082	2	295.641	0	0	28	68.354	0
34	0811	3	985.632	0	0	20	99.890	0

En primer lugar, todas las variables cuantitativas que pueden expresarse como porcentaje de otras, se expresan como tal. En nuestro caso práctico, la variable EXPORTN se expresa como porcentaje de la variable CIFRA (de la original sin simular) y el gasto en I+D se da como porcentaje de la variable original sin simular gasto en innovación. El investigador en la empresa 7, por ejemplo, verá que una empresa exporta el 23,3% de su cifra de negocios original (la cual no conoce) y si quiere un dato tendrá que aplicar ese 23,3% a la variable CIFRA simulada. Si quiere el gasto en I+D, sabrá que es el 84,4% del valor original del gasto en innovación original sin simular (el cual no conoce), pero si quiere puede aplicar el porcentaje sobre el valor simulado y así tener un valor también simulado del gasto en I+D. Por otro lado, en cuanto a la actividad, se recodifica, agrupando actividades similares, con el fin de salvaguardar la identificación de las empresas (nunca se da la actividad a más de 2 dígitos).

ident	actin	clasen	cifra	exportn	inver	tamano	gtinn	gintid
1	0004	2	3.635.311	0,0	0	70	370.999	75,4
2	0000	5	1.790.431	0,0	5.944.400	35	0	0,0
3	0001	3	48.664.533	0,0	0	350	0	0,0
4	0001	2	5.100.510	1,6	0	30	0	0,0
5	0001	2	3.339.933	0,0	18.586	50	8.879	0,0
6	0004	5	9.545.555	0,0	25.000	35	255.555	27,2
7	0000	2	2.780.982	23,3	0	17	46.823	84,4
8	0004	4	13.982.590	3,1	376.659	80	0	0,0
9	0003	2	310.085	16,6	33.090	13	141.414	0,0
10	0000	2	5.812.616	9,2	0	155	0	0,0
11	0000	2	393.170	0,0	0	15	0	0,0
12	0000	4	4.005.500	0,0	0	13	0	0,0
13	0004	4	33.377.750	40,5	421.989	148	855.885	0,0
14	0001	2	4.906.711	0,0	69.350	55	9.324	0,0
15	0001	2	1.950.125	0,0	5.393.000	46	386.525	50,3
16	0001	2	18.205.519	0,0	0	375	0	0,0
17	0003	2	21.191.213	4,6	390.033	80	2.079.700	0,0
18	0004	2	749.965	0,0	65.000	11	21.955	0,0
19	0004	2	895.939	0,0	92.900	15	93.296	0,0
20	0003	2	8.995.000	0,0	149.500	67	1.335.575	0,0
21	0004	5	31.775.250	0,0	1.310.611	445	0	0,0
22	0000	2	30.034.563	0,0	0	240	0	0,0
23	0004	5	71.708.000	3,8	160.999	560	384.538	65,0
24	0000	2	2.571.599	0,0	500	195	0	0,0
25	0001	2	9.021.886	58,1	18.900	50	91.100	83,2
26	0001	2	6.845.367	0,0	350.035	45	208.000	0,0
27	0001	5	16.325.020	0,0	850	48	53.531	0,0
28	0001	2	15.045.268	0,0	207.099	138	0	0,0
29	0003	2	958.997	0,0	0	39	120.099	74,5
30	0000	2	28.422.500	1,0	2.357.750	207	254.799	61,4
31	0003	2	2.152.722	7,0	30.700	25	376.787	0,0
32	0000	3	1.332.522	0,0	27.500	5	0	0,0
33	0003	2	295.641	0,0	0	28	68.354	0,0
34	0001	3	985.632	0,0	0	20	99.890	0,0

A continuación, se procede a obtener el valor anonimizado para las demás variables cuantitativas. El orden en que se procesan las variables a anonimizar es intrascendente, ya que se repite el proceso una vez para cada variable, independientemente de los valores que toman las demás variables (excepto el de la variable recodificada ACTIN, evidentemente).

## 1 Variable CIFRA

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable CIFRA en segundo lugar. A continuación, se "agrupan" las cinco empresas con mayor valor en la variable CIFRA de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de cinco en todos los casos, y para estos grupos se calcula la media aritmética:

ident	actin	cifra
1	0004	3.635.311
2	0000	1.790.431
3	0001	48.664.533
4	0001	5.100.510
5	0001	3.339.933
6	0004	9.545.555
7	0000	2.780.982
8	0004	13.982.590
9	0003	310.085
10	0000	5.812.616
11	0000	393.170
12	0000	4.005.500
13	0004	33.377.750
14	0001	4.906.711
15	0001	1.950.125
16	0001	18.205.519
17	0003	21.191.213
18	0004	749.965
19	0004	895.939
20	0003	8.995.000
21	0004	31.775.250
22	0000	30.034.563
23	0004	71.708.000
24	0000	2.571.599
25	0001	9.021.886
26	0001	6.845.367
27	0001	16.325.020
28	0001	15.045.268
29	0003	958.997
30	0000	28.422.500
31	0003	2.152.722
32	0000	1.332.522
33	0003	295.641
34	0001	985.632

ident	actin	cifra
22	0000	30.034.563
30	0000	28.422.500
10	0000	5.812.616
12	0000	4.005.500
7	0000	2.780.982
24	0000	2.571.599
2	0000	1.790.431
32	0000	1.332.522
11	0000	393.170
3	0001	48.664.533
16	0001	18.205.519
27	0001	16.325.020
28	0001	15.045.268
25	0001	9.021.886
26	0001	6.845.367
4	0001	5.100.510
14	0001	4.906.711
5	0001	3.339.933
15	0001	1.950.125
34	0001	985.632
17	0003	21.191.213
20	0003	8.995.000
31	0003	2.152.722
29	0003	958.997
9	0003	310.085
33	0003	295.641
23	0004	71.708.000
13	0004	33.377.750
21	0004	31.775.250
8	0004	13.982.590
6	0004	9.545.555
1	0004	3.635.311
19	0004	895.939
18	0004	749.965

El segundo paso de anonimización para esta variable, como ya se explicaba en el documento, no tiene en cuenta la actividad, y las medias se hacen de tres en tres. Por tanto, una vez cogidas las empresas más grandes de cada actividad, se ordenan decrecientemente los registros restantes según el valor de la variable CIFRA (esta vez sin considerar la actividad recodificada) y se "agrupan" de tres en tres, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay dos grupos de cuatro. Para cada grupo también se calcula la media aritmética, que es el valor de la variable CIFRA anonimizada:

ident	actin	cifra
22	0000	30.034.563
30	0000	28.422.500
10	0000	5.812.616
12	0000	4.005.500
7	0000	2.780.982
3	0001	48.664.533
16	0001	18.205.519
27	0001	16.325.020
28	0001	15.045.268
25	0001	9.021.886
17	0003	21.191.213
20	0003	8.995.000
31	0003	2.152.722
29	0003	958.997
9	0003	310.085
23	0004	71.708.000
13	0004	33.377.750
21	0004	31.775.250
8	0004	13.982.590
6	0004	9.545.555
24	0000	2.571.599
2	0000	1.790.431
32	0000	1.332.522
11	0000	393.170
26	0001	6.845.367
4	0001	5.100.510
14	0001	4.906.711
5	0001	3.339.933
15	0001	1.950.125
34	0001	985.632
33	0003	295.641
1	0004	3.635.311
19	0004	895.939
18	0004	749.965

ident	actin	cifra	
22	0000	30.034.563	
30	0000	28.422.500	
10	0000	5.812.616	14.211.232,2
12	0000	4.005.500	
7	0000	2.780.982	
3	0001	48.664.533	
16	0001	18.205.519	
27	0001	16.325.020	21.452.445,2
28	0001	15.045.268	
25	0001	9.021.886	
17	0003	21.191.213	
20	0003	8.995.000	
31	0003	2.152.722	6.721.603,4
29	0003	958.997	
9	0003	310.085	
23	0004	71.708.000	
13	0004	33.377.750	
21	0004	31.775.250	32.077.829,0
8	0004	13.982.590	
6	0004	9.545.555	
26	0001	6.845.367	
4	0001	5.100.510	5.617.529,3
14	0001	4.906.711	
1	0004	3.635.311	
5	0001	3.339.933	3.182.281,0
24	0000	2.571.599	
15	0001	1.950.125	
2	0000	1.790.431	1.514.677,5
32	0000	1.332.522	
34	0001	985.632	
19	0004	895.939	
18	0004	749.965	583.678,8
11	0000	393.170	
33	0003	295.641	

Resultando así:

ident	actin	cifra
1	0004	3.182.281,0
2	0000	1.514.677,5
3	0001	21.452.445,2
4	0001	5.617.529,3
5	0001	3.182.281,0
6	0004	32.077.829,0
7	0000	14.211.232,2
8	0004	32.077.829,0
9	0003	6.721.603,4
10	0000	14.211.232,2
11	0000	583.678,8
12	0000	14.211.232,2
13	0004	32.077.829,0
14	0001	5.617.529,3
15	0001	1.514.677,5
16	0001	21.452.445,2
17	0003	6.721.603,4
18	0004	583.678,8
19	0004	583.678,8
20	0003	6.721.603,4
21	0004	32.077.829,0
22	0000	14.211.232,2
23	0004	32.077.829,0
24	0000	3.182.281,0
25	0001	21.452.445,2
26	0001	5.617.529,3
27	0001	21.452.445,2
28	0001	21.452.445,2
29	0003	6.721.603,4
30	0000	14.211.232,2
31	0003	6.721.603,4
32	0000	1.514.677,5
33	0003	583.678,8
34	0001	1.514.677,5

---

## 2 Variable INVER

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable INVER en segundo lugar. A continuación, se "agrupan" las cinco empresas con mayor valor en la variable INVER de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de tres en todos los casos: dos que sólo tienen cuatro y las demás, tienen más de cinco, y para estos grupos se calcula la media aritmética, que es el valor anonimizado de la variable INVER:

ident	actin	inver
1	0004	0
2	0000	5.944.400
3	0001	0
4	0001	0
5	0001	18.586
6	0004	25.000
7	0000	0
8	0004	376.659
9	0003	33.090
10	0000	0
11	0000	0
12	0000	0
13	0004	421.989
14	0001	69.350
15	0001	5.393.000
16	0001	0
17	0003	390.033
18	0004	65.000
19	0004	92.900
20	0003	149.500
21	0004	1.310.611
22	0000	0
23	0004	160.999
24	0000	500
25	0001	18.900
26	0001	350.035
27	0001	850
28	0001	207.099
29	0003	0
30	0000	2.357.750
31	0003	30.700
32	0000	27.500
33	0003	0
34	0001	0

ident	actin	inver
2	0000	5.944.400
30	0000	2.357.750
32	0000	27.500
24	0000	500
15	0001	5.393.000
26	0001	350.035
28	0001	207.099
14	0001	69.350
25	0001	18.900
5	0001	18.586
27	0001	850
17	0003	390.033
20	0003	149.500
9	0003	33.090
31	0003	30.700
21	0004	1.310.611
13	0004	421.989
8	0004	376.659
23	0004	160.999
19	0004	92.900
18	0004	65.000
6	0004	25.000
1	0004	0
3	0001	0
4	0001	0
7	0000	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
22	0000	0
29	0003	0
33	0003	0
34	0001	0

*NOTA: Sólo se considera para anonimizar los valores no nulos de la variable.*

A continuación, se pasa al segundo paso de anonimización de la variable INVER de la misma manera que la variable CIFRA. Se ordenan decrecientemente los registros restantes según el valor de la variable INVER (esta vez sin considerar la actividad recodificada) y se "agrupan" de tres en tres, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay un grupo de cuatro.

Para cada grupo se calcula la media aritmética, que es el valor de la variable INVER anonimizada:

ident	actin	inver
2	0000	5.944.400
30	0000	2.357.750
32	0000	27.500
24	0000	500
15	0001	5.393.000
26	0001	350.035
28	0001	207.099
14	0001	69.350
25	0001	18.900
17	0003	390.033
20	0003	149.500
9	0003	33.090
31	0003	30.700
21	0004	1.310.611
13	0004	421.989
8	0004	376.659
23	0004	160.999
19	0004	92.900
5	0001	18.586
27	0001	850
18	0004	65.000
6	0004	25.000
1	0004	0
3	0001	0
4	0001	0
7	0000	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
22	0000	0
29	0003	0
33	0003	0
34	0001	0

ident	actin	inver	
2	0000	5.944.400	
30	0000	2.357.750	2.082.537,5
32	0000	27.500	
24	0000	500	
15	0001	5.393.000	
26	0001	350.035	1.207.676,8
28	0001	207.099	
14	0001	69.350	
25	0001	18.900	
17	0003	390.033	
20	0003	149.500	150.830,8
9	0003	33.090	
31	0003	30.700	
21	0004	1.310.611	
13	0004	421.989	472.631,6
8	0004	376.659	
23	0004	160.999	
19	0004	92.900	
18	0004	65.000	
6	0004	25.000	27.359,0
5	0001	18.586	
27	0001	850	
1	0004	0	0
3	0001	0	0
4	0001	0	0
7	0000	0	0
10	0000	0	0
11	0000	0	0
12	0000	0	0
16	0001	0	0
22	0000	0	0
29	0003	0	0
33	0003	0	0
34	0001	0	0

Resultando así:

ident	actin	inver
1	0004	0
2	0000	2.082.537,5
3	0001	0
4	0001	0
5	0001	27.359,0
6	0004	27.359,0
7	0000	0
8	0004	472.631,6
9	0003	150.830,8
10	0000	0
11	0000	0
12	0000	0
13	0004	472.631,6
14	0001	1.207.676,8
15	0001	1.207.676,8
16	0001	0
17	0003	150.830,8
18	0004	27.359,0
19	0004	472.631,6
20	0003	150.830,8
21	0004	472.631,6
22	0000	0
23	0004	472.631,6
24	0000	2.082.537,5
25	0001	1.207.676,8
26	0001	1.207.676,8
27	0001	27.359,0
28	0001	1.207.676,8
29	0003	0
30	0000	2.082.537,5
31	0003	150.830,8
32	0000	2.082.537,5
33	0003	0
34	0001	0

### 3 Variable TAMANO

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable TAMANO en segundo lugar. A continuación, se “agrupan” las cinco empresas con mayor valor en la variable TAMANO de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de cinco (y por tanto, de tres) en todos los casos, y para estos grupos se calcula la media aritmética, que es el valor simulado de la variable TAMANO:

ident	actin	tamano
1	0004	70
2	0000	35
3	0001	350
4	0001	30
5	0001	50
6	0004	35
7	0000	17
8	0004	80
9	0003	13
10	0000	155
11	0000	15
12	0000	13
13	0004	148
14	0001	55
15	0001	46
16	0001	375
17	0003	80
18	0004	11
19	0004	15
20	0003	67
21	0004	445
22	0000	240
23	0004	560
24	0000	195
25	0001	50
26	0001	45
27	0001	48
28	0001	138
29	0003	39
30	0000	207
31	0003	25
32	0000	5
33	0003	28
34	0001	20

ident	actin	tamano
22	0000	240
30	0000	207
24	0000	195
10	0000	155
2	0000	35
7	0000	17
11	0000	15
12	0000	13
32	0000	5
16	0001	375
3	0001	350
28	0001	138
14	0001	55
5	0001	50
25	0001	50
27	0001	48
15	0001	46
26	0001	45
4	0001	30
34	0001	20
17	0003	80
20	0003	67
29	0003	39
33	0003	28
31	0003	25
9	0003	13
23	0004	560
21	0004	445
13	0004	148
8	0004	80
1	0004	70
6	0004	35
19	0004	15
18	0004	11

A continuación, se pasa al segundo paso de anonimización de la variable TAMANO de la misma manera que hicimos con las variables CIFRA e INVER. Se ordenan decrecientemente los registros restantes según el valor de la variable TAMANO (esta vez sin considerar la actividad recodificada) y se "agrupan" de tres en tres, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, hay dos grupos de cuatro. Para cada grupo se calcula la media aritmética, que es el valor de la variable TAMANO anonimizada:

ident	actin	tamano
22	0000	240
30	0000	207
24	0000	195
10	0000	155
2	0000	35
16	0001	375
3	0001	350
28	0001	138
14	0001	55
5	0001	50
17	0003	80
20	0003	67
29	0003	39
33	0003	28
31	0003	25
23	0004	560
21	0004	445
13	0004	148
8	0004	80
1	0004	70
7	0000	17
11	0000	15
12	0000	13
32	0000	5
25	0001	50
27	0001	48
15	0001	46
26	0001	45
4	0001	30
34	0001	20
9	0003	13
6	0004	35
19	0004	15
18	0004	11

ident	actin	tamano	
22	0000	240	
30	0000	207	
24	0000	195	166,4
10	0000	155	
2	0000	35	
16	0001	375	
3	0001	350	
28	0001	138	193,6
14	0001	55	
5	0001	50	
17	0003	80	
20	0003	67	
29	0003	39	47,8
33	0003	28	
31	0003	25	
23	0004	560	
21	0004	445	
13	0004	148	260,6
8	0004	80	
1	0004	70	
25	0001	50	
27	0001	48	48,0
15	0001	46	
26	0001	45	
6	0004	35	36,7
4	0001	30	
34	0001	20	
7	0000	17	16,8
11	0000	15	
19	0004	15	
12	0000	13	
9	0003	13	10,5
18	0004	11	
32	0000	5	

Resultando así:

ident	actin	tamano
1	0004	260,6
2	0000	166,4
3	0001	193,6
4	0001	36,7
5	0001	193,6
6	0004	36,7
7	0000	16,8
8	0004	260,6
9	0003	10,5
10	0000	166,4
11	0000	16,8
12	0000	10,5
13	0004	260,6
14	0001	193,6
15	0001	48,0
16	0001	193,6
17	0003	47,8
18	0004	10,5
19	0004	16,8
20	0003	47,8
21	0004	260,6
22	0000	166,4
23	0004	260,6
24	0000	166,4
25	0001	48,0
26	0001	36,7
27	0001	48,0
28	0001	193,6
29	0003	47,8
30	0000	166,4
31	0003	47,8
32	0000	10,5
33	0003	47,8
34	0001	16,8

#### 4 Variable GTINN

Se ordenan los registros según la actividad recodificada en primer lugar y según el valor de la variable GTINN en segundo lugar. A continuación, se "agrupan" las cinco empresas con mayor valor en la variable GTINN de cada actividad recodificada (si hay al menos tres). En el ejemplo, hay más de tres en todos los casos salvo en uno, que por tanto no se considera y se agrupa con los demás en el siguiente paso. Para todos los grupos de más de tres, se calcula la media aritmética, que es el valor de la variable GTINN anonimizada:

ident	actin	gtinn
1	0004	370.999
2	0000	0
3	0001	0
4	0001	0
5	0001	8.879
6	0004	255.555
7	0000	46.823
8	0004	0
9	0003	141.414
10	0000	0
11	0000	0
12	0000	0
13	0004	855.885
14	0001	9.324
15	0001	386.525
16	0001	0
17	0003	2.079.700
18	0004	21.955
19	0004	93.296
20	0003	1.335.575
21	0004	0
22	0000	0
23	0004	384.538
24	0000	0
25	0001	91.100
26	0001	208.000
27	0001	53.531
28	0001	0
29	0003	120.099
30	0000	254.799
31	0003	376.787
32	0000	0
33	0003	68.354
34	0001	99.890

ident	actin	gtinn
30	0000	254.799
7	0000	46.823
15	0001	386.525
26	0001	208.000
34	0001	99.890
25	0001	91.100
27	0001	53.531
14	0001	9.324
5	0001	8.879
17	0003	2.079.700
20	0003	1.335.575
31	0003	376.787
9	0003	141.414
29	0003	120.099
33	0003	68.354
13	0004	855.885
23	0004	384.538
1	0004	370.999
6	0004	255.555
19	0004	93.296
18	0004	21.955
2	0000	0
3	0001	0
4	0001	0
8	0004	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
21	0004	0
22	0000	0
24	0000	0
28	0001	0
32	0000	0

*NOTA: Sólo se considera para anonimizar los valores no nulos de la variable.*

A continuación, se pasa al segundo paso como en el resto de variables. Se ordenan decrecientemente los registros restantes según el valor de la variable GTINN (esta vez sin considerar la actividad recodificada) y se "agrupan" de tres en tres, salvo el último o dos últimos grupos (los de menor valor) que pueden ser de cuatro empresas. En el ejemplo, no hay ningún grupo de cuatro. Para cada grupo se calcula la media aritmética, que es el valor de la variable GTINN anonimizada:

ident	actin	gtinn
15	0001	386.525
26	0001	208.000
34	0001	99.890
25	0001	91.100
27	0001	53.531
17	0003	2.079.700
20	0003	1.335.575
31	0003	376.787
9	0003	141.414
29	0003	120.099
13	0004	855.885
23	0004	384.538
1	0004	370.999
6	0004	255.555
19	0004	93.296
30	0000	254.799
7	0000	46.823
14	0001	9.324
5	0001	8.879
33	0003	68.354
18	0004	21.955
2	0000	0
3	0001	0
4	0001	0
8	0004	0
10	0000	0
11	0000	0
12	0000	0
16	0001	0
21	0004	0
22	0000	0
24	0000	0
28	0001	0
32	0000	0

ident	actin	gtinn	
15	0001	386.525	
26	0001	208.000	
34	0001	99.890	167.809,2
25	0001	91.100	
27	0001	53.531	
17	0003	2.079.700	
20	0003	1.335.575	
31	0003	376.787	810.715,0
9	0003	141.414	
29	0003	120.099	
13	0004	855.885	
23	0004	384.538	
1	0004	370.999	392.054,6
6	0004	255.555	
19	0004	93.296	
30	0000	254.799	
33	0003	68.354	123.325,3
7	0000	46.823	
18	0004	21.955	
14	0001	9.324	13.386,0
5	0001	8.879	
2	0000	0	0
3	0001	0	0
4	0001	0	0
8	0004	0	0
10	0000	0	0
11	0000	0	0
12	0000	0	0
16	0001	0	0
21	0004	0	0
22	0000	0	0
24	0000	0	0
28	0001	0	0
32	0000	0	0

Resultando así:

ident	actin	qtinn
1	0004	392.054,6
2	0000	0,0
3	0001	0,0
4	0001	0,0
5	0001	13.386,0
6	0004	392.054,6
7	0000	123.325,3
8	0004	0,0
9	0003	810.715,0
10	0000	0,0
11	0000	0,0
12	0000	0,0
13	0004	392.054,6
14	0001	13.386,0
15	0001	167.809,2
16	0001	0,0
17	0003	810.715,0
18	0004	13.386,0
19	0004	392.054,6
20	0003	810.715,0
21	0004	0,0
22	0000	0,0
23	0004	392.054,6
24	0000	0,0
25	0001	167.809,2
26	0001	167.809,2
27	0001	167.809,2
28	0001	0,0
29	0003	810.715,0
30	0000	123.325,3
31	0003	810.715,0
32	0000	0,0
33	0003	123.325,3
34	0001	167.809,2

Finalmente, se obtienen los 34 registros siguientes:

ident	actin	clasen	cifra	exportn	inver	tamano	qtinn	gintid
1	0004	2	3.182.281,0	0,0	0,0	260,6	392.054,6	75,4
2	0000	5	1.514.677,5	0,0	2.082.537,5	166,4	0,0	0,0
3	0001	3	21.452.445,2	0,0	0,0	193,6	0,0	0,0
4	0001	2	5.617.529,3	1,6	0,0	36,7	0,0	0,0
5	0001	2	3.182.281,0	0,0	27.359,0	193,6	13.386,0	0,0
6	0004	5	32.077.829,0	0,0	27.359,0	36,7	392.054,6	27,2
7	0000	2	14.211.232,2	23,3	0,0	16,8	123.325,3	84,4
8	0004	4	32.077.829,0	3,1	472.631,6	260,6	0,0	0,0
9	0003	2	6.721.603,4	16,6	150.830,8	10,5	810.715,0	0,0
10	0000	2	14.211.232,2	9,2	0,0	166,4	0,0	0,0
11	0000	2	583.678,8	0,0	0,0	16,8	0,0	0,0
12	0000	4	14.211.232,2	0,0	0,0	10,5	0,0	0,0
13	0004	4	32.077.829,0	40,5	472.631,6	260,6	392.054,6	0,0
14	0001	2	5.617.529,3	0,0	1.207.676,8	193,6	13.386,0	0,0
15	0001	2	1.514.677,5	0,0	1.207.676,8	48,0	167.809,2	50,3
16	0001	2	21.452.445,2	0,0	0,0	193,6	0,0	0,0
17	0003	2	6.721.603,4	4,6	150.830,8	47,8	810.715,0	0,0
18	0004	2	583.678,8	0,0	27.359,0	10,5	13.386,0	0,0
19	0004	2	583.678,8	0,0	472.631,6	16,8	392.054,6	0,0
20	0003	2	6.721.603,4	0,0	150.830,8	47,8	810.715,0	0,0
21	0004	5	32.077.829,0	0,0	472.631,6	260,6	0,0	0,0
22	0000	2	14.211.232,2	0,0	0,0	166,4	0,0	0,0
23	0004	5	32.077.829,0	3,8	472.631,6	260,6	392.054,6	65,0
24	0000	2	3.182.281,0	0,0	2.082.537,5	166,4	0,0	0,0
25	0001	2	21.452.445,2	58,1	1.207.676,8	48,0	167.809,2	83,2
26	0001	2	5.617.529,3	0,0	1.207.676,8	36,7	167.809,2	0,0
27	0001	5	21.452.445,2	0,0	27.359,0	48,0	167.809,2	0,0
28	0001	2	21.452.445,2	0,0	1.207.676,8	193,6	0,0	0,0
29	0003	2	6.721.603,4	0,0	0,0	47,8	810.715,0	74,5
30	0000	2	14.211.232,2	1,0	2.082.537,5	166,4	123.325,3	61,4
31	0003	2	6.721.603,4	7,0	150.830,8	47,8	810.715,0	0,0
32	0000	3	1.514.677,5	0,0	2.082.537,5	10,5	0,0	0,0
33	0003	2	583.678,8	0,0	0,0	47,8	123.325,3	0,0
34	0001	3	1.514.677,5	0,0	0,0	16,8	167.809,2	0,0

# ANEXO 3: Muestra de Empresas

**Tabla Abis. Muestra de empresas. Año 2003 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	3794	1044	4838
Empresas sin gasto en I+D interna	--	2426	2426
<b>TOTAL</b>	<b>3794</b>	<b>3470</b>	<b>7264</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	--	--	--
Empresas sin gastos en innovación <sup>3</sup>	19	--	19
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>7283</b>

**Tabla Abis. Muestra de empresas. Año 2004 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	5197	1139	6336
Empresas sin gasto en I+D interna	--	2366	2366
<b>TOTAL</b>	<b>5197</b>	<b>3505</b>	<b>8702</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	437	--	437
Empresas sin gastos en innovación <sup>3</sup>	1017	--	1017
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>10156</b>

**Tabla Abis. Muestra de empresas. Año 2005 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	7400	1194	8594
Empresas sin gasto en I+D interna	--	2219	2219
<b>TOTAL</b>	<b>7400</b>	<b>3413</b>	<b>10813</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	412	--	412
Empresas sin gastos en innovación <sup>3</sup>	954	--	954
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>12179</b>

<sup>1</sup> Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel (200 o más trabajadores, hacer I+D interna, menos de 200 trabajadores con gastos en I+D externa y sin gastos en I+D interna, menos de 200 trabajadores sin gastos en innovación, respectivamente). En el caso de tratarse de una empresa incorporada por recuperación se considerará el rasgo que caracterizaba a la empresa en el año inicial de incorporación.

<sup>2</sup>Muestra de empresas incorporada en 2004.

<sup>3</sup>Dos empresas incorporadas a esta muestra en 2003 no cumplen los requisitos de la muestra.

**Tabla Abis. Muestra de empresas. Año 2006 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	7355	1167	8522
Empresas sin gasto en I+D interna	--	2224	2224
<b>TOTAL</b>	<b>7355</b>	<b>3391</b>	<b>10746</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	417	--	417
Empresas sin gastos en innovación <sup>3</sup>	961	--	961
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>12124</b>

**Tabla Abis. Muestra de empresas. Año 2007 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	7098	1120	8218
Empresas sin gasto en I+D interna	--	2156	2156
<b>TOTAL</b>	<b>7098</b>	<b>3276</b>	<b>10374</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	405	--	405
Empresas sin gastos en innovación	907	--	907
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>11686</b>

**Tabla Abis. Muestra de empresas. Año 2008 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	6829	1092	7921
Empresas sin gasto en I+D interna	--	2086	2086
<b>TOTAL</b>	<b>6829</b>	<b>3178</b>	<b>10007</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	396	--	396
Empresas sin gastos en innovación <sup>3</sup>	872	--	872
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>11275</b>

<sup>1</sup> Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel (200 o más trabajadores, hacer I+D interna, menos de 200 trabajadores con gastos en I+D externa y sin gastos en I+D interna, menos de 200 trabajadores sin gastos en innovación, respectivamente). En el caso de tratarse de una empresa incorporada por recuperación se considerará el rasgo que caracterizaba a la empresa en el año inicial de incorporación.

<sup>2</sup>Muestra de empresas incorporada en 2004.

<sup>3</sup>Dos empresas incorporadas a esta muestra en 2003 no cumplen los requisitos de la muestra.

**Tabla Abis. Muestra de empresas. Año 2009 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	6601	1064	7665
Empresas sin gasto en I+D interna	--	2024	2024
<b>TOTAL</b>	<b>6601</b>	<b>3088</b>	<b>9689</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	385	--	385
Empresas sin gastos en innovación	817	--	817
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>10891</b>

**Tabla Abis. Muestra de empresas. Año 2010 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	6330	1030	7360
Empresas sin gasto en I+D interna	--	1978	1978
<b>TOTAL</b>	<b>6330</b>	<b>3008</b>	<b>9338</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	364	--	364
Empresas sin gastos en innovación	768	--	768
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>10470</b>

**Tabla Abis. Muestra de empresas. Año 2011 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	6076	999	7075
Empresas sin gasto en I+D interna	--	1912	1912
<b>TOTAL</b>	<b>6076</b>	<b>2911</b>	<b>8987</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	345	--	345
Empresas sin gastos en innovación	742	--	742
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>10074</b>

<sup>1</sup> Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel (200 o más trabajadores, hacer I+D interna, menos de 200 trabajadores con gastos en I+D externa y sin gastos en I+D interna, menos de 200 trabajadores sin gastos en innovación, respectivamente). En el caso de tratarse de una empresa incorporada por recuperación se considerará el rasgo que caracterizaba a la empresa en el año inicial de incorporación.

<sup>2</sup>Muestra de empresas incorporada en 2004.

**Tabla Abis. Muestra de empresas. Año 2012 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	5851	962	6813
Empresas sin gasto en I+D interna	--	1856	1856
<b>TOTAL</b>	<b>5851</b>	<b>2818</b>	<b>8669</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	331	--	331
Empresas sin gastos en innovación	709	--	709
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>9709</b>

**Tabla Abis. Muestra de empresas. Año 2013 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	5577	927	6504
Empresas sin gasto en I+D interna	--	1798	1798
<b>TOTAL</b>	<b>5577</b>	<b>2725</b>	<b>8302</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	309	--	309
Empresas sin gastos en innovación	657	--	657
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>9268</b>

**Tabla Abis. Muestra de empresas. Año 2014 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	4142	892	5034
Empresas sin gasto en I+D interna	--	1656	1656
<b>TOTAL</b>	<b>4142</b>	<b>2548</b>	<b>6690</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	201	--	201
Empresas sin gastos en innovación	255	--	255
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>7146</b>

<sup>1</sup> Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel (200 o más trabajadores, hacer I+D interna, menos de 200 trabajadores con gastos en I+D externa y sin gastos en I+D interna, menos de 200 trabajadores sin gastos en innovación, respectivamente). En el caso de tratarse de una empresa incorporada por recuperación se considerará el rasgo que caracterizaba a la empresa en el año inicial de incorporación.

<sup>2</sup>Muestra de empresas incorporada en 2004.

**Tabla Abis. Muestra de empresas. Año 2015 <sup>1</sup>**

	Empresas con menos de 200 trabajadores	Empresas con 200 o más trabajadores	TOTAL
Empresas con gasto en I+D interna	4038	878	4916
Empresas sin gasto en I+D interna	--	1604	1604
<b>TOTAL</b>	<b>4038</b>	<b>2482</b>	<b>6520</b>
Empresas con gastos en I+D externa y sin gastos en I+D interna <sup>2</sup>	189	--	189
Empresas sin gastos en innovación	249	--	249
<b>TOTAL MUESTRA</b>	<b>--</b>	<b>--</b>	<b>6958</b>

<sup>1</sup> Las empresas se incorporan a cada una de las cuatro muestras consideradas de acuerdo con el rasgo que caracterizaba a las empresas en el año de incorporación al panel (200 o más trabajadores, hacer I+D interna, menos de 200 trabajadores con gastos en I+D externa y sin gastos en I+D interna, menos de 200 trabajadores sin gastos en innovación, respectivamente). En el caso de tratarse de una empresa incorporada por recuperación se considerará el rasgo que caracterizaba a la empresa en el año inicial de incorporación.

<sup>2</sup>Muestra de empresas incorporada en 2004.