



## Medida sintética de asociación para determinar el tamaño empresarial de las PyMEs

### Synthetic measure of association to determinate the business size of SMEs

Revista FIR, FAEDPYME International Review // Vol. 5 N° 8 // enero - junio de 2016 // pp. --- // e-ISSN: 2255-078X

Gerardo Reyes Ruiz

Profesor, Tiempo Completo, Centro Universitario UAEM Valle de Chalco Solidaridad, México. Doctor en Estudios Empresariales (*perfil actuariales*), Facultad de Economía y Empresa, Universidad de Barcelona. Miembro del Sistema Nacional de Investigadores (Candidato a Investigador) del CONACYT.

E-mail: greyesru@uaemex.mx

Ma. de Lourdes Elena García Vargas

Profesora investigadora titular "B", Universidad Autónoma del Estado de Hidalgo, Pachuca de Soto, México. Escuela superior de administración. Doctora en Administración, Universidad del Celaya. Estancia posdoctoral, Facultad Latinoamericana de Ciencias Sociales (FLACSO, México). Miembro del Sistema Nacional de Investigación (candidato).

E-mail: maria\_garcia10476@uaeh.edu.mx

Orcid id 0000-0002-1724-4744

**Fecha de recepción:** 2016-02-04

**Fecha de aprobación:** 2016-07-07

#### Resumen

Las PyMEs desempeñan un papel preponderante en toda economía actual contribuyendo tanto a la generación de riqueza como a la creación de empleos, por lo que toma sentido el presente estudio al mostrar el nivel de asociación entre *ratios* financieros para construir una medida sintética que explique el tamaño empresarial para una muestra de PyMEs catalanas. Para ello se utilizaron tres técnicas de análisis de datos; primero se utilizó la técnica de Conglomerados en Dos Fases (CDF) para definir una correspondencia con el tamaño empresarial, una vez obtenido el número óptimo de clusters se estimó, a través de un análisis *fuzzy clustering*, el tamaño empresarial más probable para cada empresa. Los resultados obtenidos arrojan evidencias de que los grupos iniciales no deberían ser tres sino dos para las 12,658 PyMEs de Cataluña donde una gran parte de las denominadas pequeñas empresas presentan información financiera que bien pudiesen clasificarlas como empresas medianas. El estudio aporta una forma diferente para considerar el tamaño empresarial.

Palabras clave: tamaño empresarial; PyMEs; *fuzzy clustering*; Modelo Probit.

JEL: C4, C8, M1.

#### Abstract

SMEs play a major role in today's economy all contributing both to the generation of wealth and the creation of jobs, so it makes sense this study to show the level of association between financial ratios to construct a synthetic measure that explains the size business for a sample of Catalan SMEs. To do three techniques of data analysis were used; Technical Cluster was first used in two phases (CDF) to define a correspondence with firm size, after obtaining the optimal number of clusters was estimated through a fuzzy analysis clustering, business size most likely for each company. The results yield evidence that the initial groups should not be three but two for 12,658 SMEs in Catalonia where a large proportion of small businesses called present financial information that might well classify as medium enterprises. The study provides a different way to consider business size.

Keywords: Business size, SMEs, Fuzzy clustering, Probit model.



## 1. Introducción

La trascendencia y relevancia que adquiere el estudio en torno de las PyMEs radica en la importancia que éstas tienen en el desarrollo económico del país. Si bien es cierto que su crecimiento y desarrollo se ha enmarcado en un contexto económico y político particular, se vuelve cada vez más necesaria la implementación de modelos que permitan a este sector desenvolverse de forma pertinente, conforme las condiciones que las competencias globales lo requieran. Es importante considerar que las PyMEs generan empleos apoyando a la economía nacional e internacional. Las pequeñas y medianas empresas PyMEs europeas crearon el 85% de los nuevos puestos de trabajo entre 2001 y 2010, es decir 1,1 millones de nuevos empleos cada año (European Commission, 2012).

Según la Comisión Europea (CE) la tasa de creación de empleo anual de las PyMEs a nivel comunitario en este período ha sido el doble que en el caso de las grandes empresas europeas con un aumento anual medio del 1% frente al 0.5% de las grandes (Valda 2012). Hasta el 2004, las PyMEs representaba el 65.4% del empleo en los países miembros de la Unión Europea (UE), el 69% en los países no miembros de la UE, el 61.9% en Asia, el 55.43% en Norteamérica y el 66.9% en Latinoamérica (Castelleti, 2005). Por tanto, es innegable que su subsistencia soporta a más de la mitad de la población económicamente activa del planeta, debido a lo cual es imprescindible conocer más de ellas y buscar la forma en que dichas empresas puedan alcanzar niveles de excelencia para sí mismas y los países que las poseen.

De la información que se cuenta sobre las características de las PyMEs comunes entre los empresarios europeos y los latinoamericanos, se puede decir lo siguiente: En Europa, el 93% del total de empresas, cuentan con menos de 10 empleados, el 6% incluye entre 10 y hasta 49 empleados; el 0.8% comprende entre 50 y 249 colaboradores y finalmente, tan solo el 0.2% del total cuenta con 250 o más empleados (Zevallos, 2007). En América Latina, la sola existencia de las PyMEs representa un promedio de entre el 90% (en el caso de Uruguay) y 98.5% del total de las empresas en países como Guatemala, Argentina, Brasil, Chile y México (Zevallos, 2003). Si se toma en cuenta que en algunos países latinoamericanos refieren problemas crónicos sobre las bases y formas en que estas empresas son creadas y sostenidas como la excesiva burocracia, la lentitud de los trámites, la falta de confianza en los reglamentos y la falta de promoción de los gobiernos federales sobre programas para apoyo exclusivo de las PyMEs no se puede dejar de considerar la tenacidad y verdadera confianza que poseen los empresarios que han decidido iniciar negocios donde existen elementos en contra (Guaipatín, 2003).

Existen dos formas de surgimiento y clasificación de las PyMEs. Por un lado aquellas que se originan como empresas propiamente dichas, es decir, en las que se distingue correc-

tamente una organización y una estructura, una gestión empresarial y el trabajo remunerado, se desarrollan dentro del sector formal de la economía; por otra parte, están aquellas que tuvieron un origen familiar caracterizadas por una gestión, sólo les preocupó su supervivencia sin considerar el costo de oportunidad del capital, o inversión que permite el crecimiento. De acuerdo a Longenecker (2001) entre las ventajas de las PyMEs se consideran las siguientes: son un importante motor de desarrollo del país; tienen una gran movilidad, permitiéndoles ampliar o disminuir el tamaño de la planta así como cambiar los procesos técnicos necesarios; por su dinamismo conservan la posibilidad de crecimiento para llegar a convertirse en una empresa grande y absorber una porción importante de la población. Entre las desventajas se cuenta con la dificultad de reinvertir las utilidades para mejorar el equipo y las técnicas de producción; generalmente no cuenta con personal especializado y capacitado además no paga salarios competitivos, presentan algunas deficiencias en la calidad de la producción ya que en muchas de ellas los controles de calidad son mínimos o no existen. También tienen otros problemas derivados de la falta de organización como: ventas insuficientes, debilidad competitiva, mal servicio, mala atención al público, precios altos o mala calidad, activos fijos excesivos, mala ubicación, descontrol de inventarios, problemas de impuestos, falta de financiamiento adecuado y oportuno, dificultad de acceso a la tecnología, insumos, mercado, información, crédito y servicios de apoyo (Schmitz, 1982). De acuerdo a Schmitz (1995) gran parte de la evidencia empírica demuestra que las pequeñas empresas ubicadas en concentraciones empresariales son capaces de superar algunos de los obstáculos mencionados con anterioridad.

La presente investigación pretende mostrar un nivel de asociación entre algunos ratios financieros, previamente analizados, para con ello construir una medida sintética que explique el tamaño empresarial. De esta manera, también se adopta la idea propuesta por Galindo (2005) en cuanto a que el tamaño empresarial bien puede verse como una multiplicidad de variables que actuarían como subrogados de dicho concepto. Por lo tanto, el objetivo del estudio es determinar el tamaño empresarial a través de una medida sintética de asociación así como validar la probabilidad de pertenencia al conglomerado pronosticado.

## 2. Marco de referencia

La perspectiva del conocimiento gerencial sostiene que los gerentes operan en entornos complejos y dinámicos, por lo que necesitan focalizar su atención, interpretación y acciones para desarrollar conocimiento y estructuras que permitan desarrollar sus competencias para ser fuertes en el mercado. Miles y Snow (1978) exploran el contenido y la estructura del conocimiento estratégico de los altos directivos mediante el cálculo de las diferencias en el nivel de atención que pre-

sentan los informes anuales para asuntos vitales; su modelo se estructura con veintidós temas que, integrados en siete factores principales describen la cognición estratégica identificada por los directivos. Miles y Snow (1978) muestran que los expertos calificadores pueden reconocer estos factores al leer los informes financieros anuales que los contienen, los investigadores emplean el análisis de conglomerados para identificar grupos de empresas que comparten perfiles similares en estas dimensiones estratégicas que son interpretadas como ejemplos de grupos estratégicos cognitivos. Estos grupos muestran alineación con tipos vitales relativamente estables en el tiempo y difieren en rendimiento financiero.

En el período 1992 a 2003 Kabanoff y Brown (2008) demostraron a partir de una muestra de 1038 empresas australianas que, los tipos de estrategias según el modelo de Miles y Snow (1978) se mantienen relativamente estables en este período y que difieren respecto a su rendimiento financiero. Mosqueda (2004) presentó un modelo dinámico de valoración de riesgo para aumentar la asertividad de predicción de quiebra empresarial conocido como Ratio de Valoración Ponderado (RPV) para empresas mexicanas, orientado a las PyMEs; integra variables cuantitativas como la Rentabilidad del ejercicio =  $(\text{Resultado neto del ejercicio} / \text{Activos totales}) * 100$ ,  $CC / AT$ ,  $\text{Ventas totales} / AT$ , Tasa de crecimiento operativo =  $(\text{NOPLAT actual} - \text{NOPLAT año anterior}) / \text{NOPLAT año anterior}$ ,  $\text{Costo de ventas} / \text{Inventarios}$ ,  $AC - \text{Inventarios} / PC$ ,  $PT / CC$ ,  $CC / (\text{ventas totales} / 360)$ , Utilidad del ejercicio /  $CC$ , Variación porcentual UPA / Variación porcentual, utilidad antes de intereses e impuestos; para aumentar la consistencia del cálculo de un Ratio de Valoración Ponderado (RPV). También involucra variables cualitativas como el conocimiento del mercado, administración estratégica, operaciones de negocio, compras, entre otras.

Sabatino, y Smith (2004) presentan un modelo que se elabora a partir de información contable de la empresa de tres ejercicios fiscales anteriores donde los usuarios transcribirán los saldos de las cuentas al balance de comprobación respetando los signos de cada cuenta. El balance de comprobación contiene los siguientes datos a llenar: Cálculo de razones financieras de los estados financieros históricos, los estados financieros proyectados, cálculo de las razones financieras proyectadas y análisis de inversión. Estas estrategias se manifiestan como un aumento o disminución de las ventas, estimaciones de impuesto, gastos por intereses, dividendos, gastos de equipos, adquisición de deudas, amortizaciones de capital. Las estrategias son consideradas para proyectar cinco años del plan a implementar. En un estudio realizado sobre 108 PYMES en el sector de reparación de automóviles en el sur de España, Aragón, Hurtado, Sharma, y García (2008) encontraron que las empresas con prácticas más proactivas conseguían un mayor rendimiento financiero significativamente positivo.

Labatut, Pozuelo, y Veres (2009) han considerado en su modelo los siguientes ratios financieros: Resultado antes de intereses e impuestos / Gastos Financieros, Resultado del ejercicio / Ventas, Cash Flow recursos generados / Ventas, Resultado antes de impuestos / Ventas, Resultado antes de intereses e impuestos / Ventas, Resultado de actividades ordinarias / Ventas, Valor añadido / ventas, Pasivo fijo / Ventas; por considerarlos más eficaces y rápidos para detectar alarmas de fracaso empresarial. Avci, Madanoglu y Okumus (2011) analizan si las empresas de turismo en un país en desarrollo como Turquía adoptan ubicaciones estratégicas y si existen diferencias en su rendimiento siguiendo la misma clasificación de Miles y Snow (1978), para ello analizan 756 PyMEs del sector turismo. Los resultados del estudio muestran que existe una diferencia significativa entre las empresas que siguen una estrategia exploradora porque alcanzan un mayor rendimiento que las empresas de tipo defensivo y las empresas que siguen una estrategia analizadora mostraron rendimientos comparables a las empresas exploradoras.

Mongrut, Alberti, Fuenzalida, y Akamine (2011) estimaron en su modelo los siguientes ratios financieros: Retorno sobre activos =  $BAIT (1-t) / \text{Activo neto promedio}$ , Predicción de insolvencia empresarial, Coeficiente de solvencia =  $\text{Total Pasivo} / \text{Patrimonio}$ , para predecir Insolvencia Empresarial. Sánchez, Mosqueda y García (2014) realizaron un contraste del modelo de predicción de quiebra empresarial en las PyMEs manufactureras, de comercio y servicios basado en el Modelo de Predicción de Quiebra Empresarial de Mosqueda (2010) considerando ratios financieros como rentabilidad de la inversión  $ROI = \text{Capital contable} / \text{activos totales}$  y Presión financiera =  $\text{pago a proveedores} / \text{Utilidad neta de operación}$ . El modelo presenta robustez al predecir con un año de antelación el fracaso empresarial.

### 3. Metodología

El presente estudio empírico se divide en dos etapas principalmente, en la primera se utiliza un análisis de *clusters* para obtener una medida de asociación entre una batería de *ratios* financieros y en la segunda, se lleva a cabo una regresión *probit* para validar la probabilidad de pertenencia al conglomerado pronosticado. En la primera parte se analizan los *ratios* financieros mediante la técnica denominada Conglomerados en Dos Fases (CDF) para con ello obtener el número óptimo de *clusters* a formar y, posteriormente, mediante un algoritmo de *Fuzzy Clustering* llevar a cabo la discriminación de los elementos y con ello integrar los *clusters* que serán sometidos posteriormente a un modelo *Probit* para comprobar su probabilidad de pertenencia al *cluster* asignado.

**Minería de datos.** La minería de datos<sup>1</sup> surge como una tecnología que intenta ayudar a comprender el contenido de

una base de datos. En el contexto de los sistemas de bases de datos es considerada como uno de los desarrollos más prometedores en la industria de la información interdisciplinaria (Han y Kamber, 2006). Además, representa la posibilidad de buscar información dentro de un conjunto de datos con la finalidad de extraer, a su vez, información nueva y útil que se encuentra oculta en grandes volúmenes de datos (Tugrul, Rueda, Martin, y Gerdstri, 2006). El proceso *Knowledge Discovery in Data bases* (KDD) que incluye a la minería de datos, enfatiza que el *clustering* de datos juega un papel muy importante para detectar patrones ocultos en la información. Es decir, el *clustering* es una tarea primordial en el proceso de la minería de datos para expresar grupos, identificar distribuciones y extraer información interesante de los datos. Así, el *clustering* consiste en agrupar una colección dada de datos no etiquetados en un conjunto de grupos de tal manera que los objetos que pertenecen a un grupo sean homogéneos entre sí, buscando además que la heterogeneidad entre los distintos grupos sea lo más elevada posible (Tan, Steinbach y Kumar, 2006). En términos de variabilidad el análisis de *clusters* se focaliza en minimizar dicha variabilidad dentro de los grupos, pero al mismo tiempo, busca maximizar la variabilidad entre los distintos grupos considerados. Así, los grupos se crean de acuerdo a las características de los datos, y no siguen una asignación de clases previamente definidas, por lo que el proceso *clustering* es también conocido como clasificación no supervisada (Halkidi, Batistakis y Vazirgianis, 2001).

**Técnicas de minería de datos.** Pueden dividirse en dos categorías principales: a) Predictivas, cuando el aprendizaje es supervisado y, b) Descriptivas, cuando el aprendizaje es no supervisado (Mittra y Acharya, 2003). En los procesos predictivos cada observación incluye un valor de la clase a la que corresponde. El principal objetivo de estos procesos es predecir el valor de un atributo en particular basado en valores de otros atributos. El atributo que ha de ser predicho se conoce como variable dependiente u objetivo (*target*), mientras que a los atributos utilizados para realizar una predicción se les conoce como variables de exploración. Entre los procesos descriptivos sobresalen: 1) *Clustering*; 2) Asociaciones y 3) Dependencias.

En los procesos descriptivos, el conjunto de observaciones no tienen clases asociadas. Por tanto, su objetivo será derivar características (correlaciones, *clusters*, trayectorias, anomalías) que describan posibles relaciones entre los datos. Como bien menciona Pérez, Puldón y Espín (2012), estos procesos son comúnmente de exploración natural y frecuentemente requieren de técnicas de *post* procesamiento para explicar los resultados, por ello, y de manera generalizada, a estos procesos predictivos se les asocia: 1) Clasificación; 2) Regresión y; 3) Predicción.

**Objetivo de la técnica *clustering*.** A pesar de que el algoritmo de *clustering* es un proceso de aprendizaje no supervisado,

en el cual no se requiere predefinir una clasificación, su principal objetivo no se limita a establecer tan solo una partición de los datos para extraer el conocimiento de acuerdo a las características de los mismos. De hecho, las "particiones" de datos no se vislumbran en el mismo conjunto de datos, por lo tanto, los objetos son agrupados basándose en el principio de maximización de similitud dentro de cada *cluster* y minimización de similitud entre diferentes *clusters*. Ello ha hecho posible que el algoritmo de *clustering* sea considerado como una de las mejores técnicas para obtener conocimiento y realizar exploraciones a partir de los datos (Han y Kamber, 2006). En este contexto, un *cluster*, en términos generales, es un conjunto de elementos que presenta una fuerte similitud matemática entre ellos y, simultáneamente, una débil similitud con otros elementos. Se le denomina *clustering* a la detección de subespacios (*clusters*) en un determinado espacio de datos.

**Calidad de los datos: la selección de los ratios relevantes.**

Sin duda, la calidad de los datos es una de las características más deseables en cualquier proceso que considere su análisis. No obstante, es importante tomar en cuenta la presencia de valores faltantes, *outliers*, inconsistencia, datos duplicados y otras imperfecciones que bien pudiesen deformar al conjunto de datos. En general, la mayoría de los procesos de *clustering* no distinguen las anomalías dentro de los grupos. La manera más idónea de tratar dichas anomalías es definir una partición del conjunto de datos y conservar al subconjunto de anomalías por separado para tratarlas en forma diferente para con ello no interferir con el proceso principal de *clustering*. Existen múltiples maneras para tratar las anomalías pero es importante mencionar que deben de tomarse en cuenta desde una etapa de *pre* procesamiento. Algunos algoritmos han implementado ciertas metodologías para trabajar con datos denominados *outliers*, por ejemplo CURE (Guha, Rastogi y Shim, 2001) y DBSCAN (Ester, Kgiegel, Sander y Xu, 1996). Los *ratios* relevantes constituyen las variables explicativas con las cuales, y a partir de un determinado procedimiento, se pueden clasificar a las empresas con relación a su tamaño. Sin embargo, la selección de estas variables explicativas ha sido, hasta el día de hoy, de naturaleza empírica, es decir, los modelos desarrollados hasta ahora han identificado los *ratios* que utilizarían a partir de un *test* para cada uno de ellos al momento de separar a las empresas en pequeñas, medianas o microempresas.

Este último análisis implica que no existe ningún modelo teórico que fundamente la elección de las variables independientes de los estudios realizados. Un aspecto importante a considerar en la selección de los *ratios* es el problema de la multicolinealidad: Es decir, la correlación que pudiese existir entre los *ratios* que se seleccionen. No obstante, la eliminación de determinados *ratios*, salvo que alguno sea una combinación lineal exacta de otro, significa perder información y, por lo tanto, reducir el poder del modelo (Lev, 1978). Otra situación de suma importancia, se relaciona con el

problema de la distribución normal en los *ratios*. No obstante, en la mayoría de las técnicas estadísticas se asume el supuesto de normalidad de las variables. En el caso de los *ratios*, algunos estudios que evaluaron este comportamiento aseguran que la mayoría de ellos siguen una distribución no normal (Deakin, 1976). Por su parte, la inestabilidad de los *ratios* a través del tiempo es otro problema que afecta a los modelos, ya que utilizan un horizonte temporal en su estimación. Como es lógico, los *ratios* pueden modificarse a través del tiempo por varias razones, entre las cuales se puede mencionar; el ciclo económico, cambios en los mercados, en la tecnología, etc. (Boedo, 2001). Por ello, cabe mencionar que la normalidad para las variables de inicio en este análisis no se ha conseguido totalmente y se asumió una determinada asimetría en algunas de ellas.

El estudio se divide en dos etapas principalmente, en la primera se utiliza un análisis de *clusters* para obtener una medida de asociación entre una batería de *ratios* financieros y en la segunda, se lleva a cabo una regresión *probit* para validar la probabilidad de pertenencia al conglomerado pronosticado. En la primera parte se analizan los *ratios* financieros mediante la técnica denominada Conglomerados en Dos Fases (CDF) para con ello obtener el número óptimo de *clusters* a formar y posteriormente, mediante un algoritmo de *FuzzyClustering* llevar a cabo la discriminación de los elementos y con ello integrar los *clusters* que serán sometidos posteriormente a un modelo *Probit* para comprobar su probabilidad de pertenencia al *cluster* asignado.

**Conglomerados en dos fases (CDF).** El procedimiento de CDF es una herramienta de exploración que descubre las agrupaciones naturales (o conglomerados) de un conjunto de datos que de otra manera no sería posible detectar. Esta técnica tiene la gran ventaja de que al comparar los valores para un criterio de selección del modelo puede determinar automáticamente el número óptimo de conglomerados. Para esta técnica de análisis de datos, la medida de la distancia de la verosimilitud supone que las variables del modelo de conglomerados son independientes. Además, se supone que cada variable tiene una distribución normal, pero las comprobaciones empíricas indican que este procedimiento es bastante robusto frente a las violaciones tanto del supuesto de independencia como de las distribuciones, pero aun así es de suma importancia tener en cuenta hasta qué punto se cumplen estos supuestos (Renchery Christensen, 2012). También es importante resaltar que, para los fines del presente trabajo, esta técnica será utilizada únicamente para establecer el número óptimo de *clusters* a conformar por lo que sus resultados estadísticos no serán analizados.

**Fuzzy C-Means.** En los algoritmos de *clustering* tradicionales, cada elemento pertenece única y exclusivamente a un solo *cluster*. El *fuzzyclustering* asocia cada elemento con cada *cluster* utilizando funciones de pertenencia (Soto, Flores y Virgo, 2004). La salida de cada algoritmo es un agrupamiento, no una partición excluyente. Las técnicas de agrupación difusa hacen énfasis, en

gran parte de la literatura existente, en la minimización de las distancias de los elementos perteneciente a una determinada muestra. La función objetivo para una familia de algoritmos de agrupamiento borroso es la siguiente (Díez, Navarro y Sala, 2004):

$$J(Z,U,C) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{iz})^m \|z_k - c_i\|_B^2 \quad (1)$$

Donde  $Z = \{Z_1, Z_2, \dots, Z_N\}$  son los datos que deben ser clasificados y  $U = [\mu_{ik}] \in M_{fc}$  es una matriz con partición borrosa de  $Z$ . Además:

$$0 < \sum_{k=1}^N \mu_{iz} < N \quad \text{con } i = 1, 2, \dots, c \quad (2)$$

Donde  $\mu_{ik}$  representa el grado de pertenencia para cada elemento  $Z_k$  al centro del prototipo  $C_i$ ; con  $C = [c_1, c_2, \dots, c_c] \in R^n$  como el vector de centroides a determinar y:

$$D_{ikB}^2 = \|z_k - c_i\|_B^2 \quad (3)$$

Es una norma que viene determinada por la elección de la matriz  $B$  (por ejemplo, la identidad hace que la norma sea la distancia euclídea), y  $m \in (1, \infty)$  es un exponente que determina la borrosidad de las clases resultantes (cuanto mayor es  $m$ , más borrosos son los conjuntos identificados). El valor de la función objetivo (1) es una medida ponderada del error cuadrático que se comete al representar las clases por los prototipos. La minimización de la función objetivo (1) es un problema de optimización no lineal que puede ser resuelto de muchas formas, pero la más habitual es la conocida como algoritmo *Fuzzy C-Means* (Díez et al, 2004). Los puntos estacionarios de la función objetivo se encuentran añadiendo, al índice  $J$ , la condición de que la suma de las pertenencias de un punto a todas las clases debe ser igual a 1 mediante los multiplicadores de Lagrange:

$$J(Z,U,C,\lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik}^2 + \sum_{k=1}^N \lambda_k (\sum_{i=1}^c \mu_{ik} - 1) \quad (4)$$

$$\text{Sujeta a que } \sum_{i=1}^c \mu_{ik} = 1 \quad (5)$$

Entonces, igualando a cero las derivadas parciales de  $J$  con respecto a  $U, C$  y  $\lambda$ , las condiciones necesarias para que la función objetivo alcance su mínimo son (Dunn, 1974):

$$\mu_{ik} = \frac{1}{\sum_{i=1}^c \left( \frac{D_{ikB}}{D_{jkB}} \right)^{2/(m-1)}} \quad \text{con } i \leq c \text{ y } 1 \leq k \leq N. \quad (6)$$

$$c_i = \frac{\sum_{k=1}^N (\mu_{ik})^m z_k}{\sum_{k=1}^N (\mu_{ik})^m} \quad \text{con } 1 \leq i \leq c. \quad (7)$$

<sup>2</sup> El símbolo  $J$  denota los multiplicadores de Lagrange, es decir, define la función  $J(Z,U,C,\lambda)$  que, a su vez, depende de las variables  $Z, U, C$  y  $\lambda$ .

Así, la ecuación (7) proporciona un valor para  $c_i$  como la media ponderada de los datos que pertenecen a una clase y donde los pesos son las funciones de pertenencia.

**Modelo Probit.** El modelo *Probit* mide la relación entre la intensidad de un estímulo y la proporción de casos que presentan una cierta respuesta a dicho estímulo. Es útil para las situaciones en las que se dispone de una respuesta dicotómica que se piensa puede estar influenciada o causada por los niveles de algunas variables independientes y es, particularmente, adecuado para datos experimentales. Este tipo de análisis permite estimar la intensidad necesaria para que un estímulo llegue a inducir una determinada proporción de respuestas (Wooldridge, 2012). Para cada valor de la variable independiente (o para cada combinación de valores para múltiples variables independientes), la variable de respuesta debe contener el recuento del número de casos que presenta la respuesta de interés y que toma dichos valores de la variable independiente. Además la variable del total observado debe ser el recuento del número total de casos con dichos valores para la variable independiente. El análisis *Probit* está estrechamente relacionado con la regresión logística. En general, el análisis *Probit* es apropiado para los diseños experimentales, mientras que la regresión logística es más adecuada para los estudios observacionales. El procedimiento *Probit* informa las estimaciones de los valores efectivos para las diferentes tasas de respuesta, mientras que

la regresión logística informa las estimaciones de las razones de las ventajas para las variables independientes (Bierens, 2007).

**Muestra.** Los criterios más usados para definir el tamaño empresarial, son de tipo cuantitativo y, en concreto, el número de empleados, el activo total y la cifra de ventas (Galindo, 2005). En términos generales, el establecimiento de límites cuantitativos para estas variables no corresponde al empleo de ningún método que discrimine por qué una empresa debe de pertenecer ya sea al conglomerado de pequeñas, medianas o microempresas. La base de datos está integrada por 12,658 PyMEs de Cataluña y 22 ratios financieros. Éstos corresponden únicamente para el año 2002<sup>3</sup> y, los cuales, se muestran en la tabla 1.

### 3. Resultados

#### 3.1. Análisis de los conglomerados

Trabajar con ratios financieros siempre será complicado, debido principalmente a la construcción de los mismos, ya que es muy común que existan problemas de dependencia lineal entre varios de ellos. En este sentido, y debido al problema de multicolinealidad perfecta entre los 22 ratios financieros involucrados en este análisis, se seleccionaron y agruparon<sup>4</sup> 19 de ellos para ser considerados en el presente estudio empírico (Tabla 2).

Tabla 1. Ratios financieros

No.	Concepto	Etiqueta	Ratio
1	Activo circulante	act_circ	Act circexplot / Actiu net
2	Activo neto	act_net	Act net explot / Actiu net
3	Rentabilidad de las ventas	altres_d	Altresdespexplot / Ingexplot
4	Liquidez	cash_flo	Cash Flow / creditors a c/t
5	Operación	consums	Consumsexplot / Ingexplot
6	Créditos	creditor	Creditors c/t / Passiuremunerat
7	Gastos de personal por empleado	desp_per	Desp personal / n treballadors
8	Fondos propios	fons_pro	Fons propis / passiuremunerat
9	Financiación ajena del inmovilizado	immob_ne	Immob net explot / Actiu net
10	Margen	marge	Marge
11	Endeudamiento	rec_alie	Rec aliens II/t / passiuremunerat
12	Recursos generales	recursos	Recursos Generats
13	Rentabilidad económica	rendib1	Rendibilitateconòmica
14	Rentabilidad financiera	rendibil	Rendibilitatfinancera
15	Cobertura de Gastod financieros	rene_d	Rene / Despfinanceres
16	Coberetura de efectivo	rene_i	Rene / Ingexplot
17	Rotación	rotació	Rotació
18	Solvencia	tresorer	Tresoreria/Ingressosexplot
19	Productividad del activo	vab_ac	VAB / Act net explot
20	Productividad del inmovilizado	vab_im	VAB / Immob net explot
21	Productividad del ingreso	vab_in	VAB / Ingexplot
22	VAB (valor añadido bruto) por ocupante	vab_nom	VAB/ Nombre treb

Fuente: Elaboración de los autores

<sup>3</sup> Se tiene el entendido de que al considerar tan solo un año de estudio de ninguna manera delimita el objetivo propuesto en este trabajo de investigación.

<sup>4</sup> La discriminación de los ratios se llevó a cabo mediante la distancia de Chebychev, la cual se calculó para 3 grupos de ratios, esto debido a la construcción de los mismos y a su alta dependencia lineal.

**Tabla 2. Agrupamiento de los ratios financieros seleccionados**

	No.	Concepto	Etiqueta	Ratio
A	1	Fondos propios	fons_pro	Fons propis / passiu remunerat
	2	Créditos	creditor	Creditors c/t / Passiu remunerat
B	3	Financiación ajena del inmovilizado	immob_ne	Immob net explot / Actiu net
	4	act_circ	act_circ	Act circ explot / Actiu net
C	5	Operación	consums_	Consumsexplot / Ingexplot
	6	Productividad del ingreso	vab_in	VAB / Ingexplot
	7	Coberetura de efectivo	rene_i	Rene / Ingexplot
	8	Solvencia	tresorer	Tresoreria/Ingressosexplot
D	9	Rentabilidad financiera	rendibil	Rendibilitatfinancera
	10	Rentabilidad económica	rendib1	Rendibilitateconòmica
	11	Margen	marge	Marge
	12	Rotación	rotació	Rotació
	13	Productividad del inmovilizado	vab_im	VAB / Immob net explot
	14	Recursos generales	recursos	Recursos Generats
	15	Liquidez	cash_flo	Cash Flow / creditors a c/t
	16	Cobertura de Gastod financieros	rene_d	Rene / Despfincanceres
	17	Productividad del activo	vab_ac	VAB / Act net explot
	18	VAB (valor añadido bruto) por ocupante	vab_nom	VAB/ Nombre treb
	19	Gastos de personal por empleado	desp_per	Desp personal / n treballadors

Fuente: Elaboración de los autores

Para estos 19 *ratios* se obtuvo que, a través de un análisis de Conglomerados en Dos Fases y que fue realizado con el paquete estadístico SPSS 20.0, el número óptimo de conglomerados resultante fue dos (Tabla 3). A su vez, la importancia de los atributos se presenta en el gráfico 1.

Para estos dos conglomerados se realizó una prueba de homogeneidad de varianzas, donde el factor de análisis fue el grupo pronosticado (Tabla 4). Como puede apreciarse en la mencionada tabla 4, el estadístico de Levene (el cual no depende del supuesto de normalidad) para contrastar la igualdad de las varianzas en cada grupo no es significativo al 95% para casi todos los *ratios*, excepto para el denominador Creditor. Es decir, esta variable no presenta diferencias estadísticas entre los dos grupos pronosticados en cuanto a sus varianzas se refiere. Ello significa que dicho *ratio* puede ser omitido en el análisis.

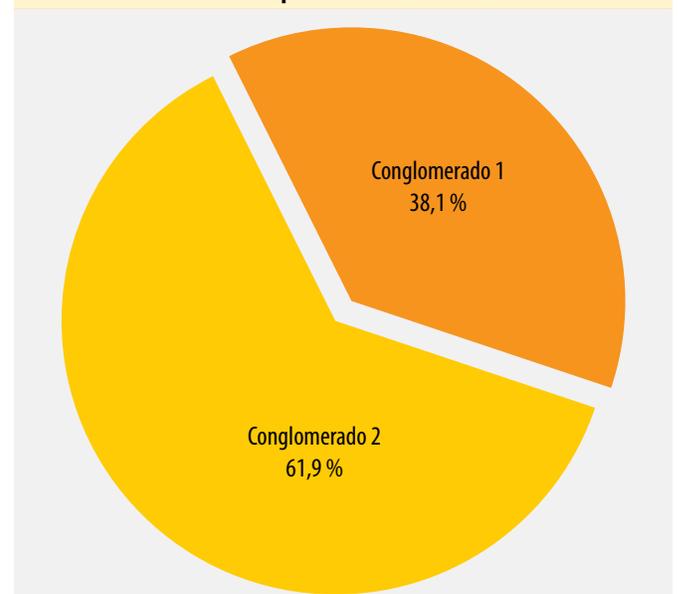
No obstante, también se realizó un Análisis de Varianza (ANOVA) para estos *ratios*, suponiendo normalidad en cada uno de ellos y tomando en cuenta como factor el conglomerado pronosticado (Tabla 5). Los resultados de la tabla 5 implican que existen diferencias significativas en cuanto a las medias de cada *ratio* en cada grupo pronosticado. Sin embargo, si se asume el supuesto *a priori* de la no igualdad de varianzas entre dichos grupos, entonces puede llevarse a cabo un contraste robusto de igualdad de medias (Tabla 6). En este sentido, el estadístico de Brown-Forsythe y el de Welch, los cuales son preferibles al estadístico F si no se supone la igualdad de las varianzas, implican que las medias difieren estadísticamente en cada grupo pronosticado, ya que el nivel de significancia es menor al 5%. Este resultado implica, a su vez, que a través de un proceso más robusto

todos los *ratios* son importantes para el presente estudio empírico.

**Tabla 3. Distribución de conglomerados**

Conglomerado	N	% de combinados	% del total
1	7726	61.0%	61.0%
2	4932	39.0%	39.0%
Combinados	12658	100.0%	100.0%
Total	12658		100.0%

Fuente: Elaboración de los autores

**Gráfico 1. Importancia de los atributos**

Fuente: Elaboración de los autores

**Tabla 4. Prueba de homogeneidad de varianzas**

<i>Ratio</i>	<i>Estadístico de Levene</i>	<i>gl1</i>	<i>gl2</i>	<i>Sig.</i>
Rendibilitatfinancera	5.860	1	12,656	0.016
Rendibilitateconómica	468.895	1	12,656	0.000
Marge	825.574	1	12,656	0.000
Rotació	157.722	1	12,656	0.000
Fons propis / passiuremunerat	177.918	1	12,656	0.000
Creditors c/t / Passiuremunerat	0.681	1	12,656	0.409
Immob net explot / Actiu net	1,061.206	1	12,656	0.000
Act circexplot / Actiu net	976.538	1	12,656	0.000
Consumsexplot / Ingexplot	212.744	1	12,656	0.000
VAB / Ingexplot	227.004	1	12,656	0.000
VAB / Immob net explot	769.832	1	12,656	0.000
VAB / Act net explot	138.218	1	12,656	0.000
Desp personal / n treballadors	10.524	1	12,656	0.001
VAB/ Nombre treb	6.766	1	12,656	0.009
Tresoreria/Ingressosexplot	438.386	1	12,656	0.000
Recursos Generats	430.539	1	12,656	0.000
Cash Flow / creditors a c/t	3,316.556	1	12,656	0.000
Rene / Ingexplot	793.174	1	12,656	0.000
Rene / Despfinanceres	707.765	1	12,656	0.000

Fuente: Elaboración de los autores mediante SPSS

**Tabla 5. ANOVA para los ratios involucrados**

<i>Concepto</i>		<i>Suma de cuadrados</i>	<i>gl</i>	<i>Media cuadrática</i>	<i>F</i>	<i>Sig.</i>
Rendibilitatfinancera	Inter-grupos	0.423	1	0.423	17.189	0.000
	Intra-grupos	311.499	12,656	0.025		
	Total	311.922	12,657			
Rendibilitateconómica	Inter-grupos	1.102	1	1.102	763.251	0.000
	Intra-grupos	18.266	12,656	0.001		
	Total	19.368	12,657			
Marge	Inter-grupos	0.708	1	0.708	1,729.923	0.000
	Intra-grupos	5.183	12,656	0.000		
	Total	5.891	12,657			
Rotació	Inter-grupos	291.652	1	291.652	565.978	0.000
	Intra-grupos	6,521.712	12,656	0.515		
	Total	6,813.364	12,657			
Fons propis / passiuremunerat	Inter-grupos	71.804	1	71.804	2,560.964	0.000
	Intra-grupos	354.847	12,656	0.028		
	Total	426.651	12,657			
Creditors c/t / Passiuremunerat	Inter-grupos	142.340	1	142.340	5,352.108	0.000
	Intra-grupos	336.588	12,656	0.027		
	Total	478.929	12,657			
Immob net explot / Actiu net	Inter-grupos	134.401	1	134.401	5,842.258	0.000
	Intra-grupos	291.151	12,656	0.023		
	Total	425.552	12,657			
Act circexplot / Actiu net	Inter-grupos	142.009	1	142.009	5,999.207	0.000
	Intra-grupos	299.584	12,656	0.024		
	Total	441.594	12,657			
Consumsexplot / Ingexplot	Inter-grupos	141.085	1	141.085	4,901.806	0.000
	Intra-grupos	364.269	12,656	0.029		
	Total	505.355	12,657			
VAB / Ingexplot	Inter-grupos	88.701	1	88.701	4,906.609	0.000
	Intra-grupos	228.794	12,656	0.018		
	Total	317.495	12,657			

**Tabla 5. ANOVA para los ratios involucrados**

Concepto		Suma de cuadrados	gl	Media cuadrática	F	Sig.
VAB / Immob net explot	Inter-grupos	5,010.741	1	5,010.741	821.059	0.000
	Intra-grupos	77,236.710	12,656	6.103		
	Total	82,247.450	12,657			
VAB / Act net explot	Inter-grupos	156.849	1	156.849	1,717.228	0.000
	Intra-grupos	1,155.981	12,656	0.091		
	Total	1,312.830	12,657			
Desp personal / n treballadors	Inter-grupos	6,199.915	1	6,199.915	114.842	0.000
	Intra-grupos	683,249.903	12,656	53.986		
	Total	689,449.818	12,657			
VAB/ Nombre treb	Inter-grupos	15,928.800	1	15,928.800	157.248	0.000
	Intra-grupos	1,282,020.200	12,656	101.297		
	Total	1,297,949.000	12,657			
Tresoreria/Ingressosexplot	Inter-grupos	1.388	1	1.388	530.009	0.000
	Intra-grupos	33.147	12,656	0.003		
	Total	34.535	12,657			
Recursos Generats	Inter-grupos	808,799.852	1	808,799.852	689.738	0.000
	Intra-grupos	14,840,665.589	12,656	1,172.619		
	Total	15,649,465.441	12,657			
Cash Flow / creditors a c/t	Inter-grupos	73.558	1	73.558	6,942.909	0.000
	Intra-grupos	134.086	12,656	0.011		
	Total	207.643	12,657			
Rene / Ingexplot	Inter-grupos	0.708	1	0.708	1,701.089	0.000
	Intra-grupos	5.266	12,656	0.000		
	Total	5.974	12,657			
Rene / Despfinanceres	Inter-grupos	4,072.933	1	4,072.933	680.743	0.000
	Intra-grupos	75,721.723	12,656	5.983		
	Total	79,794.656	12,657			

Fuente: Elaboración de los autores mediante SPSS

**Tabla 6. Pruebas robustas de igualdad de las medias**

Concepto		Estadístico(a)	Egl1	gl2	Sig.
Rendibilitatfinancera	Welch	17.373	1	10,689.009	0.000
	Brown-Forsythe	17.373	1	10,689.009	0.000
Rendibilitateconómica	Welch	666.224	1	8,228.731	0.000
	Brown-Forsythe	666.224	1	8,228.731	0.000
Marge	Welch	1,406.806	1	7,237.215	0.000
	Brown-Forsythe	1,406.806	1	7,237.215	0.000
Rotació	Welch	604.440	1	11,594.782	0.000
	Brown-Forsythe	604.440	1	11,594.782	0.000
Fons propis / passiumunerat	Welch	2,405.869	1	9,417.170	0.000
	Brown-Forsythe	2,405.869	1	9,417.170	0.000
Creditors c/t / Passiumunerat	Welch	5,374.341	1	10,575.925	0.000
	Brown-Forsythe	5,374.341	1	10,575.925	0.000
Immob net explot / Actiu net	Welch	4,974.558	1	7,862.782	0.000
	Brown-Forsythe	4,974.558	1	7,862.782	0.000
Act circexplot / Actiu net	Welch	5,140.607	1	7,954.176	0.000
	Brown-Forsythe	5,140.607	1	7,954.176	0.000

Tabla 6. Pruebas robustas de igualdad de las medias

Concepto		Estadístico(a)	Eg1	gl2	Sig.
Consumsexplot / Ingexplot	Welch	4,514.059	1	9,081.182	0.000
	Brown-Forsythe	4,514.059	1	9,081.182	0.000
VAB / Ingexplot	Welch	4,530.352	1	9,124.908	0.000
	Brown-Forsythe	4,530.352	1	9,124.908	0.000
VAB / Immob net explot	Welch	964.707	1	12,600.848	0.000
	Brown-Forsythe	964.707	1	12,600.848	0.000
VAB / Act net explot	Welch	1,621.014	1	9,499.475	0.000
	Brown-Forsythe	1,621.014	1	9,499.475	0.000
Desp personal / n treballadors	Welch	113.251	1	10,259.782	0.000
	Brown-Forsythe	113.251	1	10,259.782	0.000
VAB/ Nombre treb	Welch	155.846	1	10,347.259	0.000
	Brown-Forsythe	155.846	1	10,347.259	0.000
Tresoreria/Ingressosexplot	Welch	469.300	1	8,448.452	0.000
	Brown-Forsythe	469.300	1	8,448.452	0.000
Recursos Generats	Welch	606.713	1	8,346.271	0.000
	Brown-Forsythe	606.713	1	8,346.271	0.000
Cash Flow / creditors a c/t	Welch	5,081.171	1	6,049.310	0.000
	Brown-Forsythe	5,081.171	1	6,049.310	0.000
Rene / Ingexplot	Welch	1,386.195	1	7,263.651	0.000
	Brown-Forsythe	1,386.195	1	7,263.651	0.000
Rene / Despfinanceres	Welch	573.608	1	7,714.477	0.000
	Brown-Forsythe	573.608	1	7,714.477	0.000

a. Distribuidos en F asintóticamente.

Fuente: Elaboración de los autores mediante SPSS

Es importante mencionar que si se toma en cuenta únicamente el *ratio* que hace referencia al número de trabajadores para los tres grupos de empresas iniciales<sup>5</sup>, entonces se observa que para dichos grupos sí existen diferencias estadísticas en cuanto a sus varianzas y medias. Por lo que definitivamente la variable que hace énfasis al número de trabajadores en cada empresa resulta preponderante para este estudio. Por su parte, si se omite el *ratio* que hace referencia al número de trabajadores y se consideran únicamente los 19 *ratios* involucrados en el análisis, entonces se encontró que para los grupos iniciales 1 y 2 se obtienen diferencias estadísticas tanto para la varianza como para la media en la mayoría de los *ratios*. Sin embargo, para los grupos iniciales 2 y 3 se aprecian evidencias estadísticas en cuanto a la igualdad de varianzas, pero ligeras evidencias en cuanto a la igualdad de medias. Por lo tanto, y a través de estos resultados, se ha verificado que existen grandes evidencias para asumir el hecho de que en realidad deberían ser dos grupos de empresas y no tres. Por

lo que con base a estos hechos, se optó por considerar dos conglomerados, como se había obtenido desde un principio mediante el análisis de Conglomerados en Dos Fases.

### 3.2. Fuzzyclustering

Después de verificar el número óptimo de *clusters*, se procedió a la integración de los dos conglomerados mediante el algoritmo de *FuzzyC-Means*. Para ello, los 19 *ratios* financieros de la tabla 2 se suministraron sin transformación alguna al paquete estadístico Fuzme 3.5b y así detectar las relaciones directas entre ellos y conseguir una agrupación más óptima. Este procedimiento fue el siguiente:

- Se introdujo la base de datos, conformada por los 19 *ratios* financieros sin transformación alguna para con ello detectar las relaciones directas entre ellos, al paquete estadístico Fuzme 3.5b.

<sup>5</sup> Los grupos iniciales son; 1= micro empresa; 2= pequeña empresa y 3= mediana empresa.

b) Posteriormente se asignaron los atributos para conformar los *clusters*.

c) Se generaron las 5 clases, de las cuáles se tomó en consideración la clase 2, es decir, la que generó los dos *fuzzyclusters*.

d) Se integraron los resultados para cada *ratio*.

Cabe mencionar que este análisis se realizó para los siguientes exponentes borrosos: 1.1, 1.3, 1.5, 1.7 y 2. En el contexto de un exponente borroso, para  $m = 1$  se dice que la partición es 'hard' y, en la medida en que aumenta  $m$  entonces la partición se dice que es más difusa (borrosa). La función objetivo será mínima cuando los prototipos estén situados de tal manera que representen agrupaciones de elementos alrededor de los mismos. Para  $m=1$  Dunn (1974) encuentra que el mínimo sólo puede existir si  $\mu_{ik}$

$= \{0,1\}$ , para  $m=2$ , imponiendo la condición  $\sum_{i=1}^c \mu_{ik} = 1$  encuentra que la función objetivo  $\bar{J}$  está generalmente aceptada, aunque presenta una clara contradicción, ya que si aumentamos el valor de  $m$ , como las probabilidades son menores de la unidad, el valor de la función objetivo es cada vez menor y esto nos lleva a que, para el mismo número de grupos, una partición más difusa tiene un error menor que una partición menos difusa. Si se asume este último hecho, entonces los resultados obtenidos para los centroides de los dos *fuzzyclusters* se presentan en la tabla 7.

Con base en los centroides calculados y haciendo uso de un diferente exponente borroso se procedió a calcular la probabilidad de pertenencia máxima a cada *fuzzyclustering*. Obteniéndose un resultado como el que se muestra en la tabla 8.

Así, el nivel de atracción para cada *fuzzyclustering* y para cada exponente borroso se presenta en la tabla 9. Por su parte, el error de pertenencia para cada *fuzzyclustering* se aprecia en la tabla 10.

Para concluir este apartado, el nivel de pertenencia utilizando un exponente borroso de 1.7, con respecto al grupo inicial, es el que se presenta en la tabla 11.

Por tanto, el algoritmo borroso permite observar que un gran porcentaje (76%) de las empresas consideradas en este análisis serían clasificadas en el conglomerado 2, es decir, como empresas medianas. Sorprende el resultado de que el 90% de las denominadas pequeñas empresas bien pudiesen ser clasificadas como empresas medianas.

**Tabla 8. Nivel máximo de pertenencia al conglomerado estimado**

ID	MaxCls	CI	2a	2b
1	2b	0.57341	0.28670	0.71330
2	2b	0.12247	0.06123	0.93877
3	2b	0.02700	0.01350	0.98650
4	2a	0.09610	0.95195	0.04805
5	2b	0.86077	0.43038	0.56962
6	2a	0.06432	0.96784	0.03216
7	2b	0.03820	0.01910	0.98090
8	2b	0.00068	0.00034	0.99966
9	2b	0.06690	0.03345	0.96655
10	2a	0.57495	0.71252	0.28748
11	2b	0.02811	0.01405	0.98595
12	2a	0.06277	0.96862	0.03138
13	2b	0.58408	0.29204	0.70796
14	2b	0.00079	0.00039	0.99961
15	2b	0.04624	0.02312	0.97688
16	2b	0.09119	0.04559	0.95441
17	2a	0.14700	0.92650	0.07350
18	2b	0.11140	0.05570	0.94430
19	2a	0.00894	0.99553	0.00447
20	2b	0.08576	0.04288	0.95712
21	2a	0.09154	0.95423	0.04577
22	2b	0.02129	0.01064	0.98936
23	2b	0.02476	0.01238	0.98762
24	2b	0.00089	0.00045	0.99955
25	2b	0.01538	0.00769	0.99231
26	2b	0.03548	0.01774	0.98226
27	2b	0.01862	0.00931	0.99069
28	2b	0.00383	0.00191	0.99809
29	2b	0.02163	0.01082	0.98918
30	2b	0.00151	0.00076	0.99924
31	2b	0.02702	0.01351	0.98649
32	2a	0.06241	0.96879	0.03121
33	2b	0.00724	0.00362	0.99638
34	2b	0.00300	0.00150	0.99850
35	2b	0.15537	0.07769	0.92231
36	2b	0.03259	0.01629	0.98371
37	2b	0.35283	0.17642	0.82358
38	2b	0.02296	0.01148	0.98852
39	2b	0.05838	0.02919	0.97081
40	2a	0.00301	0.99850	0.00150
41	2b	0.01246	0.00623	0.99377
42	2a	0.92282	0.53859	0.46141
43	2b	0.00592	0.00296	0.99704
44	2b	0.01791	0.00895	0.99105
45	2b	0.02719	0.01360	0.98640
46	2b	0.00756	0.00378	0.99622
47	2a	0.14319	0.92841	0.07159
48	2a	0.01476	0.99262	0.00738
49	2b	0.01962	0.00981	0.99019
50	2b	0.00691	0.00346	0.99654
51	2a	0.94047	0.52976	0.47024
52	2b	0.00368	0.00184	0.99816
53	2b	0.02600	0.01300	0.98700
54	2b	0.00414	0.00207	0.99793
55	2b	0.01182	0.00591	0.99409
56	2b	0.01016	0.00508	0.99492
57	2b	0.02368	0.01184	0.98816
58	2b	0.00321	0.00161	0.99839
59	2b	0.00096	0.00048	0.99952
60	2b	0.01184	0.00592	0.99408
61	2b	0.01955	0.00977	0.99023
62	2b	0.00968	0.00484	0.99516

Los resultados asumen un exponente borroso de 1.7.

Fuente: Elaboración de los autores mediante Fuzme



**Tabla 9. Nivel de pertenencia con diferente exponente borroso**

Clúster	Fuzzy clustering (2)	%	Fuzzy clustering (1.7)	%	Fuzzy clustering (1.5)	%	Fuzzy clustering (1.3)	%	Fuzzy clustering (1.1)	%
1	3 163	25.0	3 037	24.0	2 972	23.5	2 933	23.2	2 931	23.2
2	9 495	75.0	9 621	76.0	9 686	76.5	9 725	76.8	9 727	76.8
TOTAL	12 658	100.0	12 658	100.0	12 658	100.0	12 658	100.0	12 658	100.0

Fuente: Elaboración de los autores con resultados obtenidos mediante Fuzme.

**Tabla 10. Error de pertenencia**

Clúster	Exponente borroso					Promedio	%	Pertenencia (frecuencias)
	1.1	1.3	1.5	1.7	2			
1	2 931	2 933	2 972	3 037	3 163	3 007	24	3 007
2	9 727	9 725	9 686	9 621	9 495	9 650	76	9 650
Total	12 658	12 658	12 658	12 658	12 658	12 658	100	12 658

Fuente: Elaboración de los autores con resultados obtenidos mediante Fuzme.

**Tabla 11. Nivel de pertenencia**

Grupo inicial	Fuzzyclustering				Total
	1	%	2	%	
1	747	9.8	6 895	90.2	7 642
2	2 207	45.0	2 694	55.0	4 901
3	83	72.2	32	27.8	115
Total	3 037	24.0	9 621	76.0	12 658

Los resultados asumen un exponente borroso de 1.7.

Fuente: Elaboración de los autores con resultados obtenidos mediante Fuzme.

**Tabla 12. Información sobre los datos**

		Nº de casos
Rechazados	Válidos	6,318
	Fuera de rango (a)	0
	Perdidos	0
	Número de respuestas > Número de sujetos	6,340
Grupo control		938
Tamany	1	3,093
	2	3,134
	3	91

Fuente: Elaboración de los autores mediante SPSS

### 3.3. Análisis Probit

Como ya se ha mencionado, este procedimiento mide la relación entre la intensidad de un estímulo y la proporción de casos que presentan una cierta respuesta a dicho estímulo. Es útil para las situaciones en las que se dispone de una respuesta dicotómica que se piensa puede estar influenciada o causada por los niveles de alguna o algunas variables independientes y es particularmente adecuada para datos experimentales. Este tipo de análisis permite estimar la intensidad necesaria para que un estímulo llegue a inducir una determinada proporción de respuestas. Los resultados para la distribución de los datos, que se usaron para la probabilidad de pertenencia a cada conglomerado, se muestran en la tabla 12.

Cabe mencionar que con este tipo de análisis se encontró una respuesta óptima sobre la convergencia del modelo a partir de 89 iteraciones. A manera de resultado, y debido a que el análisis contable financiero no es el alcance de este estudio empírico, la estimación de los parámetros se muestra en la tabla 13. De éstos puede apreciarse que doce son estadísticamente significativos al 95% de confianza, lo cual sugiere un refinamiento del modelo.

La tabla 14 muestra el resultado donde se contrasta la correcta especificación del modelo, obteniéndose para este caso

un nivel de significancia de 1, lo cual además de confirmar la validez del modelo, ratifica un refinamiento del mismo.

Así, en la tabla 15 se presenta la probabilidad promedio de pertenencia para cada grupo pronosticado con base en los tres grupos iniciales.

**Tabla 14. Contrastes de Chi-cuadrado**

		Chi-cuadrado	gl(a)	Sig.
PROBIT	Contraste de la bondad de ajuste de Pearson	3,744.519	6,295	1.000

a. Los estadísticos basados en casos individuales difieren de los estadísticos basados en casos agregados.

Fuente: Elaboración de los autores mediante SPSS

**Tabla 15. Probabilidad máxima (promedio) de pertenencia**

Grupo inicial	Grupo pronosticado	Probabilidad
1	1	0.69
	2	0.99
2	1	0.67
	2	0.99
3	1	0.65
	2	1.00

Fuente: Elaboración de los autores mediante resultados obtenidos con SPSS

Tabla 13. Estimaciones de los parámetros

	Parámetro	Estimación	Error típico	Z	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
PROBIT(a)	Rendibilitatfinancera	-1.100	1.081	-1.018	0.309	-3.218	1.018
	Rendibilitateconómica	-15.817	6.072	-2.605	0.009	-27.718	-3.916
	Marge	70.910	27.598	2.569	0.010	16.818	125.002
	Rotación	0.990	0.368	2.694	0.007	0.270	1.711
	Fons propis / passiumunerat	-1.523	0.866	-1.760	0.078	-3.220	0.173
	Creditors c/t / Passiumunerat	2.624	0.942	2.786	0.005	0.778	4.470
	Immob net explot / Actiu net	-4.272	2.933	-1.456	0.145	-10.020	1.477
	Act circexplot / Actiu net	-1.025	2.731	-0.375	0.708	-6.378	4.328
	Consumsexplot / Ingexplot	0.394	0.992	0.397	0.691	-1.550	2.338
	VAB / Ingexplot	-0.017	1.904	-0.009	0.993	-3.748	3.714
	VAB / Immob net explot	0.002	0.070	0.034	0.973	-0.136	0.140
	VAB / Act net explot	-1.781	0.861	-2.067	0.039	-3.469	-0.092
	Desp personal / n treballadors	-0.040	0.034	-1.196	0.232	-0.107	0.026
	VAB/ Nombre treb	-0.003	0.024	-0.125	0.900	-0.051	0.045
	Tresoreria/Ingressosexplot	-2.650	1.480	-1.790	0.073	-5.551	0.252
	Recursos Generats	-0.226	0.043	-5.262	0.000	-0.310	-0.142
	Cash Flow / creditors a c/t	0.504	1.105	0.457	0.648	-1.661	2.670
	Rene / Ingexplot	-57.014	26.457	-2.155	0.031	-108.870	-5.159
	Rene / Despfinanceres	-0.014	0.037	-0.376	0.707	-0.087	0.059
Intersección(b)	1	14.818	4.415	3.356	0.001	10.403	19.233
	2	15.208	4.495	3.383	0.001	10.713	19.703
	3	16.043	4.761	3.370	0.001	11.282	20.804

a. Modelo PROBIT: PROBIT(p) = Intersección + BX

b. Corresponde a la variable de agrupación tamany

Fuente: Elaboración de los autores mediante SPSS

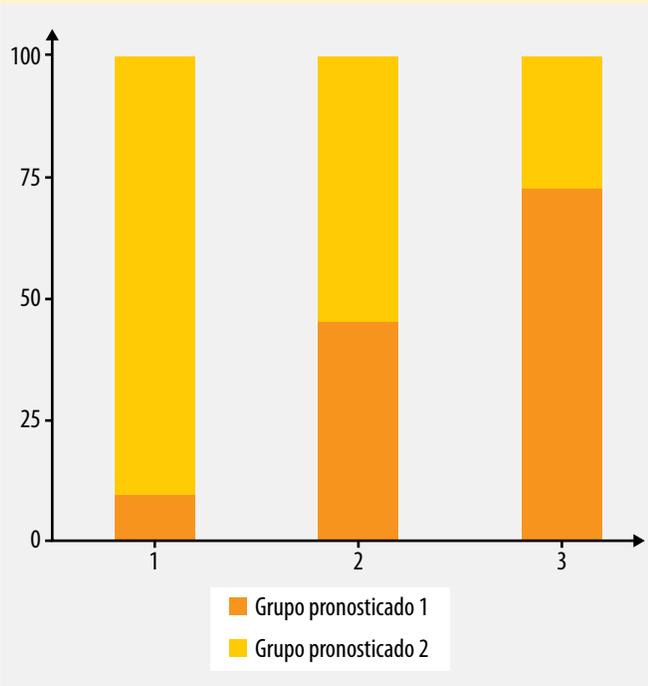
Es claro que el conglomerado 2 absorbe la mayoría de los casos que sufrieron un cambio de grupo. Sin embargo, esta probabilidad no debe de entenderse como la probabilidad de "atracción" a dicho grupo, sino más bien como la probabilidad de acierto en dicho grupo. Por último, en el gráfico 2

se presenta la distribución de los grupos pronosticados con respecto a los grupos iniciales.

El modelo *probit* confirma el resultado obtenido con el algoritmo borroso y el cual hace énfasis a que un porcentaje

considerable de las denominadas pequeñas empresas presentan información contable-financiera que bien pudiesen ser clasificadas como empresas medianas.

**Gráfico 2. Distribución de los grupos pronosticados**



Fuente: Elaboración de los autores mediante resultados obtenidos con SPSS

#### 4. Conclusiones

Los resultados obtenidos en este trabajo empírico demuestran que la clasificación de una empresa, a través de su información contable-financiera, es posible. Con base en 19 ratios financieros seleccionados, existen evidencias estadísticas de que los grupos iniciales no deberían ser tres sino dos para las 12,658 PyMEs de Cataluña. Sobresale el hecho de que una gran parte de las denominadas pequeñas empresas presentan información financiera que bien pudiesen clasificarlas como empresas medianas. Aunque dichos resultados de ninguna manera pueden ser considerados como concluyentes para la extrapolación a otra muestra de empresas. No obstante, se presentan resultados robustos que bien pueden ser la base para llevar a cabo una mejor clasificación en cuanto al tamaño empresarial se refiere. Ello, sin duda, podría hacer más eficiente y representativa la captación de impuestos a través de un tamaño empresarial más acorde a la realidad financiera de una empresa. Con este trabajo empírico también se desprende lo siguiente: si una empresa dispone de información contable y financiera que la posicionara en un tamaño empresarial superior bien pudiese ser motivada a generar un mayor número de empleos.

La medida sintética obtenida para explicar el tamaño empresarial mediante los ratios seleccionados podría ser sesgada, debido a la naturaleza de los datos principal-

mente. Sin duda, ello sugiere una transformación de los mismos o, en su caso, la recolección de una base de datos más enriquecida. Una alternativa sería la selección conveniente de otros ratios financieros más significativos para un mismo sector y, con ello, validar los resultados obtenidos. Estarvisión del tamaño empresarial proporciona una nueva perspectiva ya que los resultados encontrados no se desprenden del enfoque clásico, es decir, no se basan totalmente en el número de empleados o del total de ventas en un determinado periodo. Sin embargo, la principal limitante a la cual se enfrenta este estudio empírico es la información. Es decir, convencer a las empresas, cualquiera que sea su tamaño y giro comercial, de que integren y presenten sus estados financieros y contables no es tan sencillo como parece ser. Por tanto, en la medida en que se pueda recabar información con calidad, consistente y representativa entonces los resultados serán más acordes a su realidad. Con miras a trabajos futuros, la importancia de los resultados obtenidos en las últimas décadas para las PyMEs bien podría ser el justificante que detone el crecimiento económico y el potencial innovador de un determinado país o región.

En este trabajo se procedió a la integración de dos conglomerados mediante la técnica de *Fuzzy C-Means*, obteniéndose para dichos conglomerados borrosos una importancia promedio de los atributos de 76.24% para el conglomerado borroso 2 y 23.76% para el conglomerado borroso uno. Se encontró una respuesta óptima, a partir de 89 iteraciones, mediante un análisis *Probit* para estimar la probabilidad de pertenencia a cada grupo pronosticado. El grupo con mayor probabilidad promedio de acierto fue el denominado fuzzy-clustering dos. En otras palabras, el modelo *probit* confirma el resultado obtenido con el algoritmo borroso y el cual hace énfasis a que un porcentaje considerable de las denominadas pequeñas empresas presentan información contable-financiera que bien pudiesen ser clasificadas como empresas medianas.

#### 5. Recomendaciones

Existen evidencias estadísticas para un refinamiento del modelo e involucrar un análisis contable financiero más profundo y, con ello, enfatizar la coherencia de los resultados. Sin embargo, los resultados obtenidos con esta metodología fueron más que aceptables. La búsqueda de un concepto o medida empresarial a partir de datos numéricos por procedimientos estadísticos tradicionales de ninguna manera limitó la finalidad del presente estudio empírico. Sin embargo, y debido a la naturaleza de los datos, resulta oportuna la aplicación de otras técnicas para su análisis, como por ejemplo un análisis de redes neuronales. Éste serviría para fortalecer la relación de los ratios o para facilitar la construcción de otros ratios sintéticos que proporcionen mayor información respecto al tamaño empresarial.

## Referencias

- Aragón, J.J., Hurtado, N., Sharma, S. y García V. (2008). Environmental strategy and performance in small firms. *Journal of environmental management*, 86(1) 88-113. DOI <http://dx.doi.org/10.1016/j.jenvman.2006.11.022>
- Avci, U., Madanoglu, M., & Okumus, F. (2011). Strategic orientation and performance of tourism firms. Evidence from a developing country. *Tourism Management*, 32(1) 147-157.
- Bierens, H. (2007). *Introduction to the mathematical and statistical foundations of econometrics*. Cambridge, USA: Cambridge University Press.
- Boedo, V.L. (2001). Incidencia del tamaño sobre el comportamiento financiero de la empresa. Un análisis empírico con PyMEs Gallegas. *Revista Galega de Economía*, 10 (2) 1-23.
- Castelleti, O. (2005). Experiencia Italiana: Los Distritos Industriales. Ponencia presentada en el *Seminario Iberoamericano sobre Cooperación Empresarial*. Santiago de Chile, Chile. 26 y 27 de Mayo de 2005. Disponible en: <http://iberpyme.sela.org/Chile0505/ProgramaTrabajoChile.htm>
- Deakin, E. (1976). Distributions of Financial Accounting Ratios: Some Empirical Evidences. *The Accounting Review*, 51 (1), 90-96.
- Díez, J. L., Navarro, J. L., y Sala, A. (2004). Algoritmos de agrupamiento en la identificación de modelos borrosos. *Revista Iberoamericana de Automática e Informática Industrial*, 1 (2), 32-41.
- Dunn, J. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well separated cluster. *Journal of Cybernetics*, 3 (3), 32-57.
- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1998). A density-based algorithm for discovering clusters in large spatial databases. *Data Mining and Knowledge Discovery Journal*, 2(2), 169-194.
- European Commission (EC). (2012). *Comisión europea. Las pequeñas empresas crean el 85% de los empleos nuevos*. Comunicado de prensa. Documento disponible en [http://europa.eu/rapid/press-release\\_IP-12-20\\_es.htm](http://europa.eu/rapid/press-release_IP-12-20_es.htm)
- Galindo, A. (2005). *El tamaño empresarial como factor de diversidad*. Cádiz, España: Universidad de Cádiz y el Grupo de Investigación SEJ-366 del Plan Andaluz de Investigación.
- Guaipatín, C. (2003). *Observatorio MIPYME: Compilación estadística para 12 países de la Región*. Informe de Trabajo. Washington, USA: Banco Interamericano de Desarrollo, BID.
- Guha, S., Rastogi, R., & Shim, K. (2001). CURE: an efficient clustering algorithm for large databases. *Information Systems*, 26 (1), 35-58.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17 (2-3), 107-145.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, Second Edition*. Waltham, MA, USA: Morgan Kaufmann Publishers Inc. Elsevier.
- Kabanoff, B., & Brown, S. (2008). Knowledge structures of prospectors, analyzers and defenders: content structure, stability and performance. *Strategic Management Journal* 29(2), 149-17. DOI: [1http://dx.doi.org/10.1002/smj.644](http://dx.doi.org/10.1002/smj.644)
- Labatut Serer, G., Pozuelo, J., y Veres , E. J. (2009). Modelización temporal de los ratios contables en la detección del fracaso empresarial de la PYME española. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, 38(143), 423-447.
- Lev, B. (1978). *Análisis de los Estados Financieros: Un nuevo Enfoque*. Madrid, España: ESIC.
- Longenecker, J. (2001). *Administración de pequeñas empresas: enfoque emprendedor*. México, Ciudad de México: McGraw Hill Interamericana.
- Mitra, S., & Acharya, T. (2003). *Data Mining: Multimedia. Soft Computing and Bioinformatics*. New York, USA: Wiley.
- Miles, R.E., & Snow, Ch. (1978). *Organizational Strategy, Structure and Process*. New York, USA: Mc. Graw-Hill.
- Mongrut, S., Alberti, F., Fuenzalida, D., y Akamine, M. (2011). Determinantes de la Insolvencia Empresarial en el Perú. *Academia Latinoamericana de Administración*, 47,126-139.
- Mosqueda, R. (2004). Propuesta de un ratio ponderado de valoración empresarial. *IV Workshop de Investigación Empírica en Contabilidad Financiera*, Universidad de Sevilla. Recuperado de: <https://www.redalyc.org/pdf/461/46125172003>.
- Mosqueda, R. (2010). Fiabilidad del Método Rough Set en la conformación de modelos índice de riesgo dinámico en la predicción del fracaso. *Journal of Economics Finance and Administrative Science*, 15(28), 65-88.
- Pérez, S., Puldón, J., y Espín, R. (2012). Modelo clustering para el análisis en la ejecución de procesos de negocio. *Revista investigación operacional*, 33 (3), 210-221.

- Rencher, A., & Christensen, W. (2012). *Methods of the multivariate analysis*, 3 ed. Wiley series in probability and statistics. New York, USA: Wiley.
- Sabatino, T., y Smith D.M. (2004). *Modelo financiero para evaluar pequeñas y medianas empresas (PYMES)*. Trabajo final de especialización. Universidad Simón Bolívar. Caracas, Venezuela. Recuperado 01/07/2016 de <http://159.90.80.55/tesis/000129322.pdf>
- Sánchez, M.G., Mosqueda, R., y García, M.L. (2004). Modelo de predicción de quiebra en micro y pequeñas empresas. *FAEDPYME International Review*, 3 (5). Recuperado 01/08/2016 <http://www.faedpyme.upct.es/fir/index.php/revista1/article/view/43>
- Schmitz, H. (1982). Growth Constraints on Small-scale Manufacturing in Developing Countries: A Critical Review. *World Development*, 10(6), 429-450.
- Schmitz, H. (1995). Collective Efficiency: Growth Path for Small-scale Industry. *Journal of Development Studies*, 31(4), 529-566.
- Soto, J., Flores, A., y Vigo, M. (2004). Marco formal para una nueva función objetivo en agrupación difusa. *Revista Iberoamericana de Inteligencia Artificial*, 8 (23), 35-41.
- Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson Addison New York, USA: Wesley.
- Tugrul, D., Rueda, G., Martin, H., & Gerdstri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73 (8), 981-1012.
- Valda, J. (2012). *Las PyMEs europeas*. Recuperado 01/14/2016 de <https://jcvalda.wordpress.com>.
- Zevallos, E. (2003, Abr.). Micro, Pequeñas y Medianas empresas en América Latina. *Revista de la CEPAL*, 79, 53-70.
- Zevallos, E. (2007). *Restricciones del Entorno a la Competitividad Empresarial en América Latina*. La Paz, Bolivia: Fundación para el Desarrollo Sostenible, FUNDES.
- Wooldridge, J. (2012). *Introductory Econometrics, A Modern Approach*, 5 ed. Mason, (OH) USA: Cengage Learning.

#### ¿Cómo citar este artículo? / How to quote this article?:

Reyes-Ruiz, G., y García-Vargas, M. de L. E. (2016). Medida sintética de asociación para determinar el tamaño empresarial de las PyMEs. *FAEDPYME INTERNATIONAL REVIEW*, 5(8), 1-12. doi:10.15558/FIR.V5I8.109