



UNIVERSIDADE DA CORUÑA

Facultad de Economía y Empresa

Trabajo de
fin de grado

Análisis discriminante:
aplicación a la búsqueda
de variables relacionadas
con la capacidad de
exportar

Santiago Iñigo De Diego Lago

Tutor/a: Xosé Manuel Martínez
Filgueira

Grado en Economía

Año 2015

Resumen

El objetivo marcado en el presente trabajo es mostrar la utilidad que la aplicación de la Estadística tiene en áreas de conocimiento de la titulación de Grado en Economía.

Para desarrollar este propósito se estudia uno de los métodos de Análisis multivariante, concretamente el Análisis discriminante. Esta técnica estadística permite examinar las distinciones entre grupos existentes mediante el análisis de variables y su contribución a la diferenciación de ellos; a su vez también es posible predecir la pertenencia de nuevas observaciones a esos grupos. Las aplicaciones de este método en la economía son numerosas y de todas ellas se trata de hallar una posible explicación al comportamiento de empresas manufactureras españolas a la hora de realizar o no exportaciones en un determinado año.

Debido a los abundantes enfoques de esa explicación, el estudio se centra en la “Teoría de los Recursos y Capacidades de la empresa”, incluyendo en el análisis distintas variables directamente relacionadas con esa teoría y dando respuesta a qué recursos internos empresariales tienen una mayor influencia para la exportación.

Las variables empleadas en el análisis proceden de la “Encuesta sobre Estrategias Empresariales” realizada en dos años distintos, 2006 y 2010, con la finalidad de intentar observar cambios experimentados tras la crisis económica.

Palabras clave: Análisis discriminante, funciones discriminantes, centroides, Teoría de los Recursos y Capacidades, empresas manufactureras, exportación.

Número de palabras: 13.561

Abstract

The main objective of this thesis is to show the usefulness of applied Statistics in knowledge areas related to our degree in Economics.

In order to carry out this purpose, a technique from Multivariate analysis is studied, namely the Discriminant analysis; explaining its characteristics, criteria and its decision process. This statistical method allows for direct examination of different existing groups through analysing variables and their contribution to the differentiation of the groups themselves. Furthermore, it also renders possible the prediction of new observations belonging to these groups. The functions of this method in the economy are numerous and from all of them, a way to explain the behavior of manufacturing companies whether to export or not in a certain year, is selected.

Due to the abundant approaches of that explanation, the study centers around the “Resource-based Theory of Competitive Advantage”, including different variables in the analysis, directly connected to this theory and responding to the question of which internal resources of a business have a bigger impact on the export.

The variables applied for the analysis are derived from the “Encuesta sobre Estrategias Empresariales” realized in two different years, 2006 and 2010, with the aim of trying to observe possible changes after the economic crisis.

Key words: Discriminant analysis, discriminant functions, centroids, Resource-based Theory of Competitive Advantage, manufacturing companies, exportation.

Number of words: 13.561

Índice

Introducción	7
1. Análisis multivariante	9
2. Análisis discriminante	12
2.1 Definición y objetivos	12
2.2 Variables y tipos	13
2.3 Clasificación	14
2.4 Otros criterios de clasificación	18
2.5 Proceso de decisión	19
3. La Teoría de los Recursos y Capacidades de la empresa	27
4. Aplicación	30
4.1 Muestra	30
4.2 Variables	31
4.2.1 Discusión de variables a emplear	31
4.2.2 Especificación de las variables	33
4.3 Metodología	34
4.4 Análisis año 2006	34
4.4.1 Análisis descriptivo previo	34
4.4.2 Análisis discriminante simple.....	40
4.5 Análisis año 2010.....	48
4.5.1 Análisis descriptivo previo	48
4.5.2 Análisis discriminante simple.....	52
Conclusiones	59
Bibliografía	62
Anexo 1. Resumen de contenidos del cuestionario de la ESEE	62
Anexo 2. Tabla de las variables a emplear en el análisis	68
Anexo 3. Diagramas de caja y matriz de dispersión (2006)	69
Anexo 4. Diagramas de caja y matriz de dispersión (2010)	70

Índice de figuras

Gráfico 1. Funciones de distribución de frecuencias hipotéticas de dos grupos	16
Gráfico 2. Funciones de densidad para el año de constitución de la empresa (2006)	35
Gráfico 3. Funciones de densidad para el porcentaje de empleados con contrato indefinido (2006)	35
Gráfico 4. Funciones de densidad para el porcentaje de empleados con contrato temporal (2006)	36
Gráfico 5. Funciones de densidad para el total de trabajadores (2006).....	37
Gráfico 6. Funciones de densidad para el porcentaje de asalariados con estudios superiores (2006)	37
Gráfico 7. Funciones de densidad para el porcentaje de asalariados con estudios medios (2006)	38
Gráfico 8. Proceso de selección de variables en método LDA (2006)	43
Gráfico 9. Predicciones acertadas y erróneas en método LDA (2006).....	46
Gráfico 10. Proceso de selección de variables en método QDA (2006).....	47
Gráfico 11. Funciones de densidad para el porcentaje de empleados con contrato indefinido (2010)	49
Gráfico 12. Funciones de densidad para el porcentaje de empleados con contrato temporal (2010)	49
Gráfico 13. Funciones de densidad para el total de trabajadores (2010).....	50
Gráfico 14. Funciones de densidad para el porcentaje de asalariados con estudios superiores (2010)	51
Gráfico 15. Proceso de selección de variables en método LDA (2010).....	53
Gráfico 16. Predicciones acertadas y erróneas en método LDA (2010).....	56
Gráfico 17. Proceso de selección de variables en método QDA (2010).....	57

Índice de tablas

Tabla 1. Cuartiles y medias por variable independiente (2006)	39
Tabla 2. Test de normalidad Shapiro-Wilk (2006).....	40
Tabla 3. Prueba de homogeneidad de matrices varianza-covarianza M de Box (2006)	41
Tabla 4. Centroides de la función discriminante (2006)	43
Tabla 5. Coeficientes estandarizados (2006)	44
Tabla 6. Matriz de estructura (2006).....	44
Tabla 7. Matriz de confusión con todos los datos (2006)	45
Tabla 8. Matriz de validación cruzada (2006)	45
Tabla 9. Cuartiles y medias por variable independiente (2010)	51
Tabla 10. Test de normalidad Shapiro-Wilk (2010).....	52
Tabla 11. Prueba de homogeneidad de matrices varianza-covarianza M de Box (2010)	53
Tabla 12. Centroides de la función discriminante (2010)	54
Tabla 13. Coeficientes estandarizados (2010)	54
Tabla 14. Matriz de estructura (2010).....	54
Tabla 15. Matriz de confusión con todos los datos (2010)	55
Tabla 16. Matriz de validación cruzada (2010)	55

Introducción

La Sociedad Científica continúa evolucionando con la finalidad de descubrir nuevos procedimientos a través de la experiencia acumulada, en un proceso de constante formación y aprendizaje para la investigación, considerando importante por parte de Caro y García (2011) el uso de técnicas de análisis de datos. Estos mismos autores definen la Estadística como “ciencia cuya finalidad es estudiar los procedimientos destinados a la recogida, resumen, análisis e interpretación de un conjunto de datos, así como la obtención de inferencias científicas a partir de ellos”. Tras los grandes avances informáticos vividos en la actualidad, el uso de técnicas estadísticas es de gran utilidad en numerosos campos para obtener información y analizar resultados.

El objetivo del presente trabajo es esclarecer la utilización de una técnica estadística de manera útil en alguna área de conocimiento relacionada con la titulación. La técnica en cuestión es el Análisis discriminante, perteneciente a las técnicas incluídas en el Análisis multivariante. Éste método tiene dos finalidades: una explicada mediante un análisis descriptivo sobre distintas variables independientes en base a su influencia en la diferenciación de categorías de una variable dependiente y la otra, a través de un análisis predictivo con principios matemáticos que permite pronosticar la pertenencia de nuevos casos a grupos diferenciados que son conocidos.

Para lograr este objetivo se describe este método y se aplica en un análisis de la propensión a exportar de un grupo de empresas desde una perspectiva de la “Teoría de los Recursos y Capacidades de la empresa” intentando a responder a una serie de preguntas: ¿Qué variables relacionadas con los recursos internos de las empresas manufactureras españolas influyen en éstas a la hora de exportar o no sus productos?, ¿de qué manera contribuyen y cuáles en mayor cantidad?, ¿es el análisis satisfactorio en el cálculo de probabilidades de pertenencia en los grupos conocidos (exporta y no exporta) para nuevas observaciones?, ¿qué método de clasificación del Análisis

discriminante es más satisfactorio? y, ¿hay cambios trascendentales tras la crisis económica española?

La estructura del trabajo está constituida por una parte teórica y otra parte práctica. El marco teórico engloba los tres primeros apartados; el primero plantea brevemente en qué consiste el Análisis multivariante, el segundo profundiza en el Análisis discriminante explicando sus características, criterios y su proceso de decisión, el tercero comenta conceptos de la “Teoría de los Recursos y Capacidades de la empresa”, especificando su utilidad para realizar un enfoque de Análisis discriminante relacionado con ese campo.

El cuarto apartado abarca el marco práctico, en éste se efectúa la aplicación de dos Análisis discriminantes para conocer el comportamiento de empresas manufactureras españolas a la hora de realizar exportaciones, uno con datos procedentes del año 2006 y otro con datos del 2010.

Finalmente, se describen las conclusiones sintetizando los objetivos iniciales y los resultados obtenidos en el trabajo.

1. Análisis multivariante

“El Análisis multivariante es la rama de la Estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de $n > 1$ variables, que pueden ser cuantitativas, cualitativas o una mezcla.” (Cuadras, 1981)

Siguiendo esta definición inicial, hay una clasificación general de las distintas técnicas multivariantes. Tal como señala Pérez (2004) estas técnicas pueden, o enfocar todas las variables como independientes y simplemente tratar de averiguar cómo se organiza la relación entre ellas (Análisis de interdependencia) o por el contrario, basar su análisis en considerar una o más variables como dependientes y tratar de explicar su comportamiento con la ayuda de variables independientes (Análisis de dependencia).

Entre los métodos de Análisis de interdependencia más importantes hay varias distinciones:

- Análisis factorial y Análisis por componentes principales: ambos análisis poseen en sus variables naturaleza métrica y el objetivo es encontrar un modo de condensar la información contenida en un número alto de variables originales en un conjunto más pequeño de variables con la menor pérdida de información posible. (Hair et al, 1999)
- Análisis de correspondencias: su objetivo es el mismo que el Análisis factorial y el Análisis por componentes principales, con la importante diferencia de que las variables tienen carácter no métrico, es decir, son cualitativas. Resumiendo, se obtiene un mapa perceptual que recoja la asociación entre categorías de estas variables en modo gráfico.
- Análisis de conglomerados o Análisis cluster: a partir de una tabla de casos-variables trata de situar todos los casos en grupos homogéneos (llamados clusters) no conocidos de antemano pero sugeridos por la propia esencia de los datos, de manera que individuos que puedan ser considerados similares sean asignados a un mismo cluster, mientras que individuos diferentes se sitúen en clusters distintos. (Pérez, 2004)

- Escalamiento multidimensional: coincide con en el Análisis cluster en emplear tanto variables métricas como no métricas y su propósito es transformar las opiniones o preferencias de una serie de individuos sobre un conjunto de objetos en distancias capaces de ser representadas en un espacio multidimensional. (Arce, 1993)

En base a los métodos de Análisis de dependencia los más destacables son:

- Análisis de la varianza y covarianza: el Análisis de la varianza puede ser univariante o multivariante. El primero se define como un método estadístico que se usa para determinar si varios conjuntos de muestras aleatorias de una variable concreta proceden de una misma población o bien de poblaciones distintas, mientras que el Análisis multivariante de la varianza es una generalización del primero, donde se consideran simultáneamente varias variables dependientes que supuestamente están relacionadas entre sí, en lugar de una sola variable dependiente. (Uriel y Aldás, 2005)

En el Análisis de la covarianza, tanto simple como múltiple, las variables métricas independientes tienen como objetivo eliminar determinados efectos que puedan sesgar los resultados incrementando la varianza dentro de los grupos. (Pérez, 2004)

- Análisis conjunto: se usa con variables explicativas cualitativas y es definido como “técnica que se emplea para entender cómo conforman los individuos sus preferencias a los objetos”. (Iglesias y Sulé, 2003; p-21)
- Segmentación jerárquica: tanto la variable dependiente como las variables independientes puede ser cualitativas o cuantitativas y Díaz (2002) señala que:
Este análisis busca delimitar e identificar subconjuntos homogéneos de la población. El fin de esta técnica es polarizar las frecuencias de las categorías de una variable dependiente seleccionando los sujetos que cumplan determinadas características, identificando segmentos de población dentro de los cuales la variación de la variable dependiente sea significativamente menor que en el conjunto de la muestra. (p-20)
- Regresión lineal múltiple: conforme a Iglesias y Sulé (2003), esta técnica permite analizar la relación existente entre una variable dependiente métrica y varias variables independientes también métricas, pretendiendo determinar la combinación lineal de éstas cuyos cambios son los mejores predictores de los cambios experimentados por la variable dependiente.

- Análisis discriminante: “técnica estadística utilizada para analizar la relación entre una variable dependiente (o endógena) no métrica (categórica) y varias variables independientes (o exógenas) métricas”. El objetivo esencial del Análisis discriminante es utilizar los valores conocidos de las variables independientes para predecir con qué categoría de la variable dependiente se corresponde. (Pérez, 2004; p-5)
- Regresión logística: Esta técnica permite estudiar si una variable binomial (o dicotómica, con la que se mide la ausencia o presencia de lo que se quiere medir) depende o no de una o más variables métricas. En otras palabras son esquemas de regresión que pretenden modelar la influencia de la probabilidad de aparición de un suceso cuando se encuentra en presencia o ausencia de diversos factores. (De la Garza et al, 2013; p-16)

Englobando todos estos métodos, las funciones de las técnicas multivariantes, a parte del Análisis de dependencia o interdependencia de las variables, pueden sintetizarse en: clasificación de objetos, elaboración de hipótesis, simplificación y agrupación de variables.

Para toda realización de un correcto y exitoso Análisis multivariante, Pérez (2004) define una serie de fases que son necesarias:

- Primera fase: se lleva a cabo la definición del problema de la investigación, así como la elección de la técnica multivariante conveniente y los objetivos de nuestro análisis.
- Segunda fase: un segundo paso es desarrollar el proyecto de análisis poniendo en práctica la técnica multivariante seleccionada.
- Tercera fase: la tercera parte del proceso se centra en que tras la recogida de los datos, es preciso evaluar que se cumplan supuestos básicos tales como normalidad, linealidad, correlación... dependiendo de la técnica multivariante escogida.
- Cuarta fase: tiene lugar la estimación efectiva del modelo multivariante y la valoración global de su ajuste, como por ejemplo la capacidad de predicción del modelo o la significatividad de los parámetros.
- Quinta fase: se procede a la interpretación de los valores obtenidos del modelo estimado de acuerdo al problema inicialmente especificado. Si la interpretación no se ajusta coherentemente a este problema, debería reespecificarse el modelo con una nueva estimación.

- Sexta fase: por último se realiza la validación del modelo multivariante mediante la comprobación de los resultados con el mayor grado de fiabilidad posible, a través de contrastes específicos.

En relación con una aproximación a la actualidad, Peña (2002) señala que el avance en la tecnología informática, junto a la capacidad de adquirir y procesar grandes bases de datos en todas las ciencias, ha estimulado el desarrollo y utilización del análisis estadístico multivariante en muchas disciplinas.

Dillon y Goldstein (1984) añaden que en la mayoría de las investigaciones científicas es necesario analizar relaciones simultáneas entre tres o más variables y la complejidad de los fenómenos analizados hace que sean muchas las variables implicadas, y por ello, las investigaciones sean necesariamente multivariantes

Específicamente, en la Economía, estos métodos estadísticos multivariantes pueden usarse, entre otras cosas, para comprender el comportamiento de consumidores o cuantificar el desarrollo de un país.

2. Análisis discriminante

2.1 Definición y objetivos

De la Garza et al (2013) definen el Análisis discriminante como:

Técnica funcional que permite explicar las diferencias que existen entre los grupos de personas u objetos, por medio de variables independientes, con el objetivo de pronosticar o de clasificar nuevas observaciones en los grupos formados a priori a partir de determinados atributos que se conocen. (p-14)

Como indican Gil et al (2001) hay dos grandes objetivos que se persiguen en el uso de esta técnica:

- La descripción de diferencias entre grupos. Responde al objetivo de considerar en qué grado un conjunto de variables independientes contribuye a la diferenciación entre los grupos, es el llamado Análisis discriminante descriptivo.
- La predicción de pertenencia a los grupos. Este objetivo se basa en el uso de una o más ecuaciones matemáticas, llamadas funciones discriminantes, que permitan la clasificación de nuevos casos a partir de la información que poseemos sobre ellos a determinados grupos diferenciados conocidos, es el denominado Análisis discriminante predictivo.

Stevens (1986; p-286) destaca como rasgos apreciables de esta técnica la claridad en la interpretación que puede hacerse de los resultados, en el sentido de que un número amplio de variables puede ser reducido a un pequeño número de dimensiones capaces de explicar las diferencias fundamentales observadas entre los grupos.

2.2 Variables y tipos

La expresión funcional del Análisis discriminante es la siguiente:

$$y = F(x_1, x_2, \dots, x_m)$$

La variable dependiente y es no métrica y las variables independientes son métricas.

Como se observa en la fórmula, se trabaja con M número de variables independientes, en las que todas ellas son cuantitativas (métricas) y con una variable dependiente cualitativa (no métrica), esta última puede ser dicotómica, es decir formada por solo dos grupos (Análisis discriminante simple) o puede utilizar una escala nominal que tenga más de dos grupos, entonces se estaría ante Análisis discriminante múltiple.

Un claro ejemplo de Análisis discriminante simple sería los individuos que deciden comprar o no un producto, si a este análisis se une otro grupo como aquellos que se muestran indecisos en la compra, se considera Análisis discriminante múltiple de tres grupos.

2.3 Clasificación

El enfoque más común para este objetivo del modelo discriminante tiene sus bases en la generación de una ecuación que definirá las diferencias entre distintos grupos y permitirá la clasificación en ellos. Señala Pérez (2004) que su fundamento matemático está en conseguir, a partir de las variables explicativas, unas funciones lineales de éstas con capacidad para clasificar otros individuos. A cada nuevo caso se aplican dichas ecuaciones; y la función de mayor valor define el grupo al que pertenece.

Para conocer el proceso de clasificación de individuos en los grupos, inicialmente es necesario realizar una distinción del tipo de análisis que se va a realizar, si la clasificación será con dos grupos o con más de dos grupos.

I. Análisis discriminante simple (clasificación con dos grupos)

La metodología anteriormente mencionada, llamada ecuación discriminante de Fisher, tiene sus inicios en 1930 y como su nombre indica fue propuesta por el estadístico R. A. Fisher.

La expresión lineal de la ecuación de Fisher, que representa la combinación lineal de las variables discriminantes (independientes) es la siguiente:

$$Z = u_1X_1 + u_2X_2 + uX_3 + \dots u_mX_m$$

Z = punto discriminante

u_i = factor de ponderación o importancia de la variable i para discriminar

X_i = variables independientes

Establecen De la Garza et al (2013) que en la derivación del modelo se pretende encontrar los valores de los pesos que tienen las variables para discriminar, es decir, los valores de las constantes (u_i) siempre bajo el criterio de maximizar la relación entre las diferencias de los grupos con respecto a la variabilidad en los datos.

A su vez, a partir de demostraciones matemáticas, Fisher concluyó con un criterio para determinar los coeficientes (u_i) y la obtención de esta función discriminante:

$$\text{Maximización de} = \frac{\text{Variabilidad entre - grupos}}{\text{Variabilidad intra - grupos}}$$

Como acertadamente señalan Uriel y Aldás (2005) este criterio trata de precisar el eje discriminante de forma que las distribuciones proyectadas sobre el mismo estén lo más separadas posible entre sí (mayor variabilidad entre grupos) y, al mismo tiempo,

que cada una de las distribuciones esté lo menos dispersa (menor variabilidad dentro de los grupos). (p-288)

Analíticamente, este mismo criterio puede expresarse de esta forma:

$$\text{Maximización de } (\lambda) = \frac{u'Fu}{u'Wu}$$

$F = \text{matriz de covarianzas entre grupos}$ $W = \text{matriz}$

El cálculo de F será un cálculo equivalente a considerar que en cada grupo los valores de cada variable son iguales a la media en ese grupo y W calcula la matriz de la suma de cuadrados en cada grupo, sumándolos y dividiendo entre los grados de libertad.

La solución a este problema se consigue derivando (λ) respecto de (u) e igualando a cero; tras los cálculos se finaliza con la obtención de un vector propio (u) asociado a la matriz no simétrica $(W^{-1}F)$. De los valores propios (λ) obtenidos se retiene el mayor y (u) es el vector característico asociado al mayor valor propio de esa matriz. (Pérez, 2013)

Una vez calculada la función discriminante, si se tiene la observación de un nuevo individuo es necesario averiguar a qué grupo se le asignará y para ello hay que sustituir los valores de las variables para ese individuo en la función discriminante. El valor obtenido será la “puntuación discriminante”, que en este caso es única al estar ante Análisis discriminante simple con dos grupos.

La “puntuación discriminante” es pues el resultado que se obtiene al sustituir el valor de las variables (X_m) para ese individuo en la función discriminante:

$$D = u_1X_1 + u_2X_2 + \dots + u_mX_m$$

Para la decisión de esta asignación es necesario tener una frontera que defina hasta qué punto discriminante se asignará al primer grupo y a partir de ella al segundo grupo, es el llamado “punto de corte discriminante”.

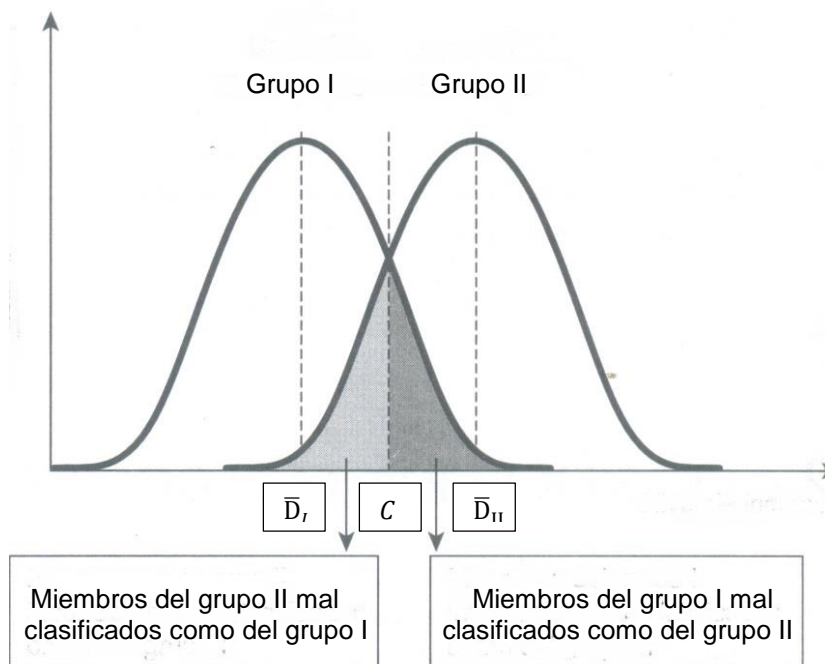
El “punto de corte discriminante” basa su cálculo en el promedio de las puntuaciones discriminantes de los denominados “centroides” de cada grupo. Los “centroides” son los estadísticos básicos que resumen la información sobre los grupos. Se calculan promediando las puntuaciones discriminantes para todos los individuos dentro del grupo particular, obteniendo la media. En este caso de Análisis discriminante simple, se obtienen dos “centroides”:

$$\bar{X}_I = \begin{pmatrix} \bar{X}_{1,I} \\ \bar{X}_{2,I} \\ \dots \\ \bar{X}_{m,I} \end{pmatrix} \quad \bar{X}_{II} = \begin{pmatrix} \bar{X}_{1,II} \\ \bar{X}_{2,II} \\ \dots \\ \bar{X}_{m,II} \end{pmatrix}$$

Con lo que, para los grupos I y II se obtiene lo siguiente, siendo C el “punto de corte discriminante”:

$$\begin{aligned} \bar{D}_I &= u_1 \bar{X}_{1,I} + u_2 \bar{X}_{2,I} + \dots + u_m \bar{X}_{m,I} \\ \bar{D}_{II} &= u_1 \bar{X}_{1,II} + u_2 \bar{X}_{2,II} + \dots + u_m \bar{X}_{m,II} \\ C &= \frac{\bar{D}_I + \bar{D}_{II}}{2} \end{aligned}$$

Gráfico 1. Funciones de distribución de frecuencias hipotéticas de dos grupos.



Fuente: Uriel y Aldás (2005)

Conforme a la Figura 1, el criterio para clasificar el individuo es el siguiente:

Si $D_i < C$, se clasifica al individuo i en el grupo I

Si $D_i > C$, se clasifica al individuo i en el grupo II

Aplicando este criterio se cometen errores de clasificación, como puede comprobarse al ver la Figura 1. Así, el área existente a la derecha de C recoge individuos

pertenecientes al grupo I pero en los que $D_i > C$, es decir, son individuos del grupo I incorrectamente clasificados en el grupo II. Recíprocamente, el área a la izquierda de C recoge individuos pertenecientes al grupo II pero en los que $D_i < C$, es decir, son individuos del grupo II incorrectamente clasificados en el grupo I. (Uriel y Aldás, 2005)

II. Análisis discriminante múltiple (clasificación con más de dos grupos)

Este tipo de análisis ocurre con el caso de G grupos ($G > 2$) y siguiendo la definición de Uriel y Aldás (2005) el número máximo de ejes discriminantes viene dado por $\min(G-1, m)$. De acuerdo con esta expresión, se puede obtener hasta $G-1$ ejes discriminantes, si el número de variables explicativas m es mayor o igual que $G-1$, hecho que suele ser siempre cierto.

La explicación descrita por Pérez (2013) es un proceso muy similar al explicado anteriormente con únicamente dos grupos, cada función discriminante D_i se obtiene como función lineal de las m variables explicativas X , es decir:

$$D_i = u_{i1}X_1 + u_{i2}X_2 + \dots + u_{im}X_m \quad i = 1, 2, \dots, G - 1$$

Los $G-1$ ejes discriminantes vienen definidos respectivamente por los vectores u_1, u_2, \dots, u_{G-1} y éstos mediante las siguientes expresiones:

$$u_1 = \begin{pmatrix} u_{11} \\ u_{12} \\ \dots \\ u_{1m} \end{pmatrix} \quad u_2 = \begin{pmatrix} u_{21} \\ u_{22} \\ \dots \\ u_{2m} \end{pmatrix} \quad \dots \quad u_{G-1} = \begin{pmatrix} u_{G-11} \\ u_{G-12} \\ \dots \\ u_{G-1m} \end{pmatrix}$$

Y al igual que en el caso de análisis simple, para la obtención del primer "eje discriminante" se maximiza (λ_1) donde:

$$\lambda_1 = \frac{u_1' F u_1}{u_1' W u_1}$$

El comienzo del proceso es como anteriormente, la ecuación para la obtención del primer eje discriminante se traduce en la obtención de un vector propio (u_1) asociado a la matriz no simétrica ($W^{-1}F$) y también de los valores propios (λ_i) obtenidos se retiene el mayor, ya que precisamente (λ_1) es la ratio que queremos maximizar. Por último, (u_1) es el vector característico asociado al mayor valor propio de esa matriz.

El resto de los ejes discriminantes vendrán dados por los vectores propios asociados a los valores propios de la matriz $(W^{-1}F)$, los cuales están ordenados de mayor a menor. De esta manera, el segundo eje discriminante tendrá menos poder discriminatorio que el primero, pero más que cualquiera de los restantes. (Pérez, 2013; p-387). La importancia de los valores (λ_i) se resume en el cociente $\frac{\lambda_i}{\sum_k \lambda_k}$ que divide un valor entre la suma de todos, indicando la proporción de discriminación que le corresponde a la función discriminante asociada a ese valor.

2.4 Otros criterios de clasificación

Anteriormente se ha considerado el criterio de Fisher como planteamiento principal del Análisis discriminante, interpretando los significados de las funciones discriminantes y los centroides de cada grupo, permitiendo clasificar a los individuos en un grupo u otro, pero en muchas ocasiones es necesario tener información complementaria a las funciones discriminantes. (De la Fuente, 2011)

Para este enfoque se puede emplear el denominado “Análisis discriminante Bayesiano”¹ que como su nombre indica tiene sus bases en el “Teorema de Bayes”.

Como Uriel y Aldás (2005) indican, la aplicación del “Teorema de Bayes” permite el cálculo de las probabilidades a priori y de la información muestral contenida en las puntuaciones discriminantes.

De la Fuente (2011) define que en el caso general de G grupos, el “Teorema de Bayes” establece que la probabilidad a posteriori de pertenencia a un grupo g con una puntuación discriminante D, con probabilidades a priori π_g es:

$$Prob\left(\frac{g}{D}\right) = \frac{\pi_g + Prob\left(\frac{D}{g}\right)}{\sum_{i=1}^G \pi_i + Prob\left(\frac{D}{\bar{i}}\right)}$$

La probabilidad condicionada $Prob(D/g)$ se obtiene calculando la probabilidad de la puntuación observada suponiendo la pertenencia a un grupo g. Dado que el denominador es una constante, se utiliza también la forma equivalente:

¹ Considerar este criterio supone disponer de información sobre la probabilidad de pertenencia a cada grupo, realizando análisis más matizados, incluyendo informaciones tales como la información a priori o los costes que implica una información errónea. (De la Fuente, 2011)

$$\text{Prob} \left(\frac{g}{D} \right) \propto \pi_g + \text{Prob} \left(\frac{D}{g} \right)$$

\propto = proporcionalidad

Este mismo autor dice que la clasificación de cada individuo se puede realizar mediante la comparación de las probabilidades a posteriori. Así, se asignará un individuo al grupo para el cual sea mayor esta probabilidad. El cálculo de probabilidades se puede realizar bajo tres supuestos diferentes: cálculo de probabilidades sin información a priori, cálculo de probabilidades con información a priori y cálculo de probabilidades con información a priori considerando los costes.

Existen distintos métodos de clasificación a mayores, entre los que destacan:

- Análisis de la regresión: la relación entre el Análisis discriminante y el Análisis de regresión es muy estrecha, si se realiza un ajuste por mínimos cuadrados, tomando como variable dependiente la variable dicotómica que define la pertenencia a uno u otro grupo y como variables explicativas a las variables clasificadoras. La obtención final serán unos coeficientes proporcionales a la función discriminante de Fisher. (De la Fuente, 2011)
- Distancia de Mahalanobis: clasifica a los individuos de la misma manera que lo hace la función de Fisher pero la diferencia reside en que la Distancia de Mahalanobis se calcula en el espacio de las variables originales y en el criterio de Fisher se sintetizan todas las variables en la función discriminante, que es la utilizada para hacer la clasificación. (Pérez, 2004)

Con este criterio se calculan para un punto i , las siguientes dos distancias:²

$$DM_{i,I}^2 = (x_i - x_I)' V_w^{-1} (x_i - x_I)$$
$$DM_{i,II}^2 = (x_i - x_{II})' V_w^{-1} (x_i - x_{II})$$

2.5 Proceso de decisión

Como cualquier técnica estadística, el Análisis discriminante implica un proceso de decisión dividido en varios pasos, los cuales presentan subdivisiones. Bajo los criterios de Hair et al (1999) se constituyen seis etapas diferenciables:

² La aplicación de este criterio consiste en asignar cada individuo al grupo para el que la distancia de Mahalanobis es menor. (De la Fuente, 2011)

Primer paso: Objetivos del análisis discriminante

La primera decisión llevada a cabo debe ser el objetivo de la investigación, si toma un enfoque de análisis de perfil mediante un estudio de las diferencias de los grupos a través de la determinación de qué variables independientes indican mejor las diferencias entre ellos o si por otro lado, se basa en una técnica analítica predictiva mediante la clasificación correcta de objetos en grupos.³

Segundo paso: Diseño de la investigación

Para una correcta aplicación del Análisis discriminante son necesarias una serie de cuestiones:

1. Selección de las variables independientes y dependientes

El investigador debe decidir inicialmente cuál va a ser la variable dependiente y cuáles las independientes, con la condición ya expuesta previamente de que deben existir al menos dos grupos en la dependiente. A esta necesidad se le añade el hecho de que los grupos que la forman tienen que ser excluyentes y exhaustivos, en otras palabras, que los grupos elegidos no contengan información encontrada en otros grupos y que se abarquen la mayor parte de grupos posibles.

Siguiendo las consideraciones de Hair et al (1999), tras la decisión sobre la variable dependiente, es el momento de hacer la selección de las variables independientes, esta elección puede ser de dos formas. La primera de ellas implica identificar las variables tanto en la investigación previa como desde el modelo teórico que sirve de base para la pregunta de la investigación. La segunda manera es intuitiva, con un enfoque dirigido a variables que no tengan investigación previa o teoría y que podrían relacionarse para predecir los grupos de la variable dependiente.

2. Tamaño de la muestra

La muestra mínima sería de cinco observaciones por variable independiente, aunque es recomendable el uso de veinte por cada una de ellas. Lógicamente, la muestra debe ser representativa de la población y el tamaño del grupo más pequeño debe ser mayor que el número de variables independientes.⁴

³ Para fines de clasificación, proporciona una base no solo para clasificar la muestra utilizada para estimar la llamada función discriminante, sino también otras observaciones que puedan tener valores para todas las variables independientes. (Hair et al, 1999; p-261)

⁴ Se requiere que la muestra sea aleatoria simple.

3. División de la muestra

La muestra deberá separarse en dos submuestras o etapas, la primera es considerada etapa de derivación, en la que se utilizarán el 60 o 70% de los datos (muestra de análisis), y el restante 40 o 30% se dejará para la segunda etapa de validación del modelo (llamada ampliación de la muestra). En esta división de la muestra es preciso conservar las proporciones de la población para mantener su representatividad.

Tercer paso: Cumplimiento de supuestos

Toda técnica estadística necesita el cumplimiento de supuestos para su perfecta aplicación, citando a Gil et al (2001) en el caso del Análisis discriminante para que sea óptimo tiene que satisfacer:

- a) Distribución normal multivariante: este supuesto ha de cumplirse para que se pueda obtener con óptima precisión tanto para las probabilidades de pertenencia a un grupo, como para el desarrollo de diferentes pruebas de significación implicadas en el Análisis discriminante. Para comprobar tanto la normalidad univariable como multivariante existen numerosas técnicas tanto gráficas como no gráficas.
- b) Igualdad de matrices de varianza-covarianza: la violación de este supuesto es menos importante en muestras grandes o iguales pero es más sensible ante muestras pequeñas o desiguales, en este caso, ante los procedimientos de clasificación, los individuos tenderían a ser clasificados en los grupos que tienen mayor dispersión. De cualquier manera, su cumplimiento es muy severo, raramente se satisface y cuando el Análisis discriminante se va a utilizar con finalidad predictiva podría ser innecesario (Huberty, 1984), ya que la función de clasificación lineal puede ser sustituida por la función de clasificación cuadrática, que no requiere de este supuesto. Entre las pruebas para su comprobación destaca la prueba M de Box.
- c) Linealidad: el supuesto de linealidad implica que existen relaciones lineales entre las variables dentro de cada grupo. Entre los métodos para la comprobación de la existencia de linealidad se encuentran la realización de diagramas de dispersión y el cálculo del coeficiente de correlación de Pearson.
- d) Ausencia de multicolinealidad y singularidad: la multicolinealidad ocurre cuando dos variables, en este caso independientes, tienen una correlación alta o

perfecta y se comportan del mismo modo ya que muestran el mismo patrón de correlaciones con las restantes variables, por lo tanto, aportan una información redundante. La singularidad supone que las puntuaciones alcanzadas por una variable son aproximadamente una combinación lineal de otras, se da cuando existe una correlación múltiple perfecta entre una variable y las restantes.⁵ Este supuesto deja de ser un problema cuando se utiliza el método de selección de variables “stepwise”, que será explicado a continuación.

Cuarto paso: Estimación del modelo discriminante y valoración del ajuste global

- Método de cálculo y criterio de selección de variables

En esta fase se elige el método de cálculo para la función discriminante, la estimación puede ser simultánea (todas las variables independientes son consideradas al mismo tiempo) o estimación por etapas, donde se escogen aquellas variables que proporcionan una mayor discriminación.

Bajo las instrucciones de Gil et al (2001; p-43), la estimación por etapas es necesaria para la selección de variables a incluir en la función discriminante porque hay situaciones en las que algunas habrían de ser eliminadas, dada su baja contribución a la discriminación de los grupos. Es el caso de variables para las cuales los grupos alcanzan valores medios similares, y por tanto no permiten establecer una diferenciación entre ellos.

Hay distintos métodos de selección de variables automáticos de los que se destacan tres:

- a) Método de selección hacia delante (forward): la primera variable que entra a formar parte del análisis es la que maximiza la separación entre grupos. A continuación, se forman parejas entre esta variable y las restantes, con la finalidad de encontrar aquella que provoca la mayor discriminación. La variable que contribuye a la mejor pareja es seleccionada en segundo lugar, este proceso se repite hasta que todas las variables hayan sido seleccionadas o las variables restantes no supongan un suficiente incremento en la capacidad de discriminación.

⁵ El motivo de la necesidad de ausencia de multicolinealidad y singularidad reside en que impiden la inversión de matrices ya que el rango de la matriz desciende, resultando un determinante cero. Esta inversión de matrices es necesario para el cálculo de los coeficientes de la función discriminante.

- b) Método de selección hacia atrás (backward): todas las variables son consideradas inicialmente, y van siendo excluidas una a una, eliminándolas del modelo hasta quedarse únicamente con las más significativas.
- c) Método de selección paso a paso (stepwise): es el más utilizado y es la combinación de los dos métodos anteriores. Empieza con una sola variable clasificadora, la más significativa, y se va introduciendo o retirando una variable en cada interacción, hasta que se establezca un criterio de detención.

- Criterios y condiciones en la selección de las variables:

Si se ha procedido con un método de selección de variables “stepwise”, hay una implicación de una regla de decisión, por lo que es preciso utilizar alguna medida de discriminación para la óptima selección de variables.

De todos los criterios posibles mencionados por Gil et al (2001), se destaca el criterio basado en la minimización de la “lambda de Wilks”. Estos autores lo definen como un criterio en el que se elige en cada paso aquella variable que una vez incorporada a la función discriminante produce el valor de lambda más pequeño para el conjunto de las variables incluidas en la función. Este proceso hacia delante, se repetirá continuamente hasta que todas las variables hayan entrado en el modelo o bien ninguna de las restantes cumpla las condiciones mínimas para la selección (condiciones que se analizan a continuación). El estadístico “lambda de Wilks”⁶ se calcula como el escalar resultante de dividir los determinantes de dos matrices:

$$\Lambda \text{ de Wilks} = \frac{|W|}{|B + W|}$$

|W| = desviaciones intergrupos

|B+W| = desviaciones globales

Como señalan estos autores, el valor de lambda puede transformarse en un estadístico multivariante general F, que permite contrastar la existencia de diferencias significativas entre los grupos. También podría usarse un F parcial, calculado como el estadístico F de entrada que se verá más adelante. Cualquiera de los tres estadísticos (lambda de Wilks, F general y F de entrada) mostrará resultados similares.

⁶ “Medida que evalúa la capacidad discriminante de la función y expresa la proporción de la varianza total explicada por las puntuaciones discriminantes”. (De la Garza et al, 2013)

Sea cual sea el criterio seleccionado para la inclusión de las variables en el método “stepwise”, éstas han de ser revisadas, de manera que cumplan una serie de condiciones; tanto antes de introducir una variable (tolerancia y F de entrada), como después de hacerlo, y comprobar que todas las seleccionadas hasta ese momento satisfacen la condición fijada para el estadístico F de salida.⁷

- a) Tolerancia: “la tolerancia de una variable independiente es la proporción de varianza de esa variable que no está asociada (que no depende) del resto de variables independientes incluidas en la ecuación” (Pardo, s.f). Es decir es una medida de asociación lineal entre ellas.⁸
- b) F de entrada: Pérez (2013) dice que:
Representa el incremento producido en la discriminación tras la incorporación de una variable respecto al total de discriminación tras la incorporación alcanzado por las variables ya introducidas. Una F pequeña aconsejaría no seleccionar la variable, pues su aporte a la discriminación de los grupos no sería importante.⁹ (p-391)
- c) F de salida: “permite valorar el descenso en la discriminación si una variable fuera extraída del conjunto de las ya seleccionadas. Aquellas variables para las cuales el valor de F es bajo, podrían ser descartas antes de proceder a un nuevo paso”¹⁰ (Gil et al, 2001; p-50). Una vez acabado el proceso de “stepwise” es posible ordenar todas las variables seleccionadas para la diferenciación de grupos y aquellas con mayor F serán las que más aportan a la discriminación.

Quinto paso: Interpretación de los resultados

Como anteriormente se señaló, además de la función predictiva del análisis discriminante, también tiene una función explicativa que informa sobre la actuación de las variables para permitir interpretar las diferencias entre los grupos.

Para explicar estas diferencias se puede recurrir a distintos resultados como son:

⁷ Una variable inicialmente seleccionada puede resultar después inadecuada si otras variables introducidas posteriormente aportan la misma contribución a la separación de grupos. (Gil et al, 2001)

⁸ $1 - R^2$, donde R es el coeficiente de determinación entre esa variable y todas las variables ya incluidas. Generalmente se fija un nivel mínimo de tolerancia del 0,001. (Pérez, 2013)

⁹ Se distribuye según F con (g-1) y (n-s-g+1) grados de libertad. n= número de individuos, g= número de grupos y s=número de variables discriminantes.

¹⁰ Los grados de libertad de la F son (g-1) y (n-s-g). n= número de individuos, g= número de grupos y s=número de variables discriminantes.

- a) Representaciones gráficas: las puntuaciones discriminantes pueden representarse gráficamente mediante histogramas (unigrupales o totales) o diagramas de dispersión.
- b) Matriz de estructura: formada por los llamados coeficientes de correlación entre cada variable y la función de discriminante, indicando las variables que aportan más a la hora de hacer la diferenciación entre grupos.¹¹
- c) Coefficientes de las funciones discriminantes estandarizados: son los coeficientes de la función discriminante (u_i), pero calculados con valores tipificados. De esta manera se homogeniza las diferencias de escala entre las variables, lo que hace comparable los respectivos coeficientes.
- d) Funciones de los centroides de los grupos: Se procede a la sustitución de los valores de los centroides de cada grupo en las funciones discriminantes y se comparan los resultados.

Sexto paso: Validación de los resultados

Para validar los resultados es útil comprobar la bondad del ajuste y ver cómo queda explicada la diferencia entre grupos mediante las funciones discriminantes.

En el caso de Análisis discriminante simple, como medida de bondad de ajuste para explicar la capacidad explicativa de la función discriminante se usa la raíz cuadrada del coeficiente eta cuadrado (η^2), el llamado “coeficiente de correlación canónica” que equivale al coeficiente de determinación de una regresión que usase una variable dicotómica como variable dependiente para indicar el grupo y la función determinante como función de regresión para las variables independientes.¹²

$$\eta = \sqrt{\frac{\lambda}{1 + \lambda}}$$

Por otro lado, para el Análisis discriminante múltiple se usa un contraste de significación para comprobar si los valores (λ_i) son estadísticamente significativos, es decir, para conocer si ayuda a la discriminación entre los diferentes grupos se usa un estadístico llamado “V de Barlett”.¹³ Las hipótesis y expresión son las siguientes:

¹¹ Valores entre -1 y 1 en valor absoluto.

¹² Un valor alto de este coeficiente indicaría que existe una relación entre el grupo de pertenencia y la función discriminante, es decir, se discrimina satisfactoriamente.

¹³ Es una función de la Λ de Wilks y se aproxima a una Chi-Cuadrado. (Uriel y Aldás, 2005)

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_G \quad H_1: \text{No todos } \lambda_g \text{ son iguales}^{14}$$

$$V = - \left\{ n - 1 - \frac{K + G}{2} \right\} \ln \Lambda \quad V \sim X_{K(G-1)}^2$$

Se contrasta secuencialmente hasta que algún valor no rechace la hipótesis nula y el número de (λ_i) significativamente distinto de 0, indica el número de funciones discriminantes a estimar. Como medida descriptiva complementaria de este contraste se suele calcular el porcentaje acumulativo de la varianza después de la incorporación de cada nueva función discriminante. (Pérez, 2013)

En este último paso también se analizan los resultados para saber su capacidad de clasificación, mediante un ratio de aciertos a través de la matriz de confusión (que incluye la proporción de aciertos y la situación de los fallos) o, en el caso de comprobar si tienen validación tanto externa como interna, es recomendado realizar un procedimiento de validación cruzada con una ampliación de la muestra, ya que el ratio de aciertos si se evaluase únicamente la muestra de análisis sería mucho mayor.

El dividir los grupos aleatoriamente en la muestra de análisis y en una ampliación de la muestra implica validar la función discriminante construida en la primera muestra en la segunda, porque como anteriormente se ha mencionado, la precisión clasificatoria será más alta.

Aunque Hair et al (1999) defienden el dividir el total de la muestra no solo una vez, si no varias veces, comprobando cada vez la validez de la función a través de la construcción de matrices de clasificación y de un ratio de aciertos.¹⁵

¹⁴ La hipótesis ha de ser rechazada para que pueda seguir utilizándose el análisis discriminante múltiple ya que sino las variables no tendrían carácter discriminatorio.

¹⁵ A partir de los diferentes ratios de aciertos se haría un promedio de ellos para obtener una media única.

3. La Teoría de los Recursos y Capacidades de la empresa

Antes de realizar la aplicación del Análisis discriminante es preciso conocer las bases que serán proporcionadas a este análisis.

La finalidad de este trabajo es conocer las variables que más discriminan a la hora de que una empresa realice una decisión estratégica, en este caso exportar, y como ya se ha especificado, antes de llevar a cabo la selección de las variables es necesario hacer una investigación previa sobre el marco teórico para que el estudio estadístico sea más preciso.

El enfoque será basado sobre la “Teoría de los Recursos y Capacidades de la empresa” que como su nombre indica refleja el papel que juegan los recursos y capacidades de la empresa a la hora de decidir su estrategia empresarial.

Durante las décadas de los 70 y 80, la dirección estratégica de la empresa solía orientarse básicamente en el análisis sectorial y de su posición competitiva en relación a la competencia. Como señala Huerta (1993), el pionero de ésta idea fue Porter (1980) que consideraba que la estrategia competitiva de la empresa significa que ésta, de la observación del entorno y de la dinámica de sus competidores, reconoce un conjunto de oportunidades, que debe aprovechar, y amenazas, que debe evitar. (p-88)

Fue a partir de los años 90 cuando los investigadores comienzan a centrar el análisis de la ventaja competitiva en los aspectos internos de la empresa, siendo éstos explicativos de su comportamiento y, haciendo referencia a Suárez e Ibarra (2002), fueron los estudios de la concepción de la empresa basada en los recursos y de las competencias y capacidades organizativas los que provocaron este hecho, surgiendo así, esta teoría. Cuervo (1993) afirma que “es la heterogeneidad de las empresas la explicación [...] de las ventajas competitivas sostenibles y de los resultados”.

La “Teoría de los Recursos y Capacidades de la empresa” debe su existencia a unas aportaciones previas. Una de ellas de la mano de Penrose (1959) con su teoría propia,

considerando que la dimensión de una empresa viene explicada por el conjunto de recursos de los que disponga, y su crecimiento será posible gracias a la explotación de éstos, generando servicios directivos que configurarán las competencias empresariales. Esta economista también precisó que la principal función de la empresa es “adquirir y organizar personas y otros recursos siendo rentable para suministrar bienes y servicios a un determinado mercado”. A continuación Andrews (1977) afirmó que el crear una actividad empresarial diferente mediante “competencias distintivas” es la clave del éxito de la empresa.

De esta manera, esta teoría se centra principalmente en el análisis interno de la empresa (sus recursos) como las características relevantes que explican sus resultados, distinguiéndose del enfoque tradicional de considerar las características de la industria como la ventaja clave (López, 2004), permitiendo definir a la empresa como elemento significativo y alejándose del concepto de igualdad entre empresas en términos de recursos y productos controlados por ellas. (Knudsen, 1995)

Los autores que siguieron a estos pioneros definen los recursos como la totalidad de activos tangibles e intangibles disponibles y/o controlables por parte de la empresa para desarrollar una estrategia competitiva. Dentro del grupo de los tangibles se encuentran los físicos y financieros, mientras que en el grupo de los intangibles están los tecnológicos y humanos (Suárez e Ibarra, 2002), ampliando así la concepción clásica de trabajo, tierra y capital.

Castillo y Portela (2002) incluyen como recurso de vital importancia la “organización”, relacionándola con la estructura organizativa y aspectos organizativos, como pueden ser la propia cultura interna de la empresa y los sistemas de planificación y control, con la finalidad de explotar de la manera más eficiente todos los recursos y capacidades en un enfoque competitivo. Desde el punto de vista de Cuervo (1993), el hecho de que el capital humano de la empresa intercambie información, proporciona la creación de dichas rutinas organizativas, las cuales dependen de un sistema de incentivos y de la integración del personal.

Barney (1991) sugiere que para desarrollar ventajas competitivas sostenibles a lo largo del tiempo, los recursos de la empresa deben ser escasos, inimitables, valiosos y no sustituibles.

Las capacidades también conocidas como competencias, son para Cuervo (1993), el conjunto de conocimientos y habilidades que aparecen con un aprendizaje colectivo

en ésta. A lo que Castillo y Portela (2002) añaden: las capacidades vienen definidas como la forma en que la empresa despliega sus recursos de forma combinada para generar alguna actividad¹⁶.

En definitiva, como afirma López (2004), esta teoría pretende aprovechar las complementariedades existentes entre la economía y la dirección estratégica para explicar el comportamiento y resultados de la empresa, con la implicación de encontrarse ante una competencia caracterizada por la incertidumbre, la asimetría de información y racionalidad limitada, y que, mediante la implantación de una estrategia, se puede alterar la estructura de la competencia en un sector.

Suárez e Ibarra (2002) concluyen que el hecho de que los recursos y capacidades se hayan convertido en la base para la estrategia a largo plazo de la empresa es debido a dos aspectos: primero, los recursos y capacidades internos proveen la dirección básica para la estrategia de la empresa, y segundo, los recursos y capacidades son la fuente principal de rentabilidad para la empresa. (p-68)

Medina (1998) apunta una posible crítica a esta teoría por parte de algunos economistas; el hecho de ser introspectiva al no tener en cuenta el entorno como aspecto también clave para la ventaja competitiva, y afirma que los defensores de la “Teoría de los Recursos y Capacidades de la empresa” piensan que los recursos dependen en gran medida de las características del entorno en el que opera la empresa.

¹⁶ “Los recursos originan las capacidades y las capacidades son el origen de la ventaja competitiva”. (Grant, 1991)

4. Aplicación

Tras la revisión teórica fundamental para el posterior estudio estadístico, se tratará de analizar en qué medida los recursos internos de la empresa influyen sobre la conducta exportadora de las mismas, concretamente, la variación en el comportamiento exportador entre distintas empresas pertenecientes a una misma categoría.

En este apartado se detallarán inicialmente la descripción técnica de la muestra de la que proceden los datos, la metodología a seguir y las variables que se usarán en el Análisis discriminante, finalizando con el análisis propiamente dicho aplicado a un grupo de empresas en dos años diferentes: 2006 y 2010.

4.1 Muestra

Los datos provienen de una muestra donde el ámbito geográfico de referencia es el conjunto del territorio español, la llamada “Encuesta sobre Estrategias Empresariales” o también conocida como ESEE, que afronta el estudio de diferentes aspectos de la empresa industrial manufacturera española.

La ESEE¹⁷ tiene su origen en un acuerdo suscrito en 1990 entre el Ministerio de Industria y la Fundación Empresa Pública para que esta última, a través de su Programa de Investigaciones Económicas, diseñara y realizara una investigación estadística anual con carácter de panel a empresas industriales. El objetivo era construir una fuente que permitiera desarrollar caracterizaciones y análisis de la realidad industrial española de corte microeconómico, situando el acento en los comportamientos estratégicos de las empresas. (Fariñas y Jaumandreu, 1994)

Entre las características de la ESEE¹⁸, estos mismos autores destacan tres:

¹⁷ La población de referencia de la ESEE son las empresas con 10 o más trabajadores de lo que se conoce habitualmente como industria manufacturera y las variables tienen dimensión temporal anual.

¹⁸ Ver Anexo 1 para conocer los contenidos de la encuesta.

- 1- La representatividad de la muestra de las empresas encuestadas, donde estas unidades fueron seleccionadas mediante criterios de muestreo aleatorio y exhaustividad, dependiendo del empleo de las empresas.
- 2- La orientación de la encuesta basada en obtener conocimiento y cuantificación de variables empresariales que den información sobre las estrategias de las empresas, considerando como estratégicas todas las decisiones adoptadas por la empresa sobre variables que le sirven como instrumento de competencia, tanto en el corto plazo como en el largo plazo.
- 3- El enfoque de tener observaciones a lo largo del tiempo de la muestra de empresas, adecuado para análisis de evaluación de las decisiones de las empresas.

En relación al tamaño de la muestra, consiste en más de 1300 observaciones por variable independiente, por lo que se está a una muestra de tamaño muy grande que puede provocar problemas en el cumplimiento de los supuestos pero que se ha escogido por su alta representatividad.

4.2 Variables

4.2.1 Discusión de variables a emplear

Como se ha visto anteriormente desde el punto de vista teórico hay numerosos recursos internos en la empresa que pueden influenciar en la potencialidad de la misma a la hora de exportar. De la división entre recursos tangibles e intangibles, el análisis se realizará sobre los últimos ya que son los más exclusivos de la empresa y los que, en mi opinión, pueden presentar unas conclusiones más interesantes y menos predecibles a priori.

Ya nombrado previamente, los recursos intangibles se dividen en recursos tecnológicos, recursos organizacionales y recursos humanos. Asimismo, se ha decidido prescindir de los primeros porque la obviedad de la importancia de éstos a la hora de exportar es conocida; que una empresa tenga disponibilidad de recursos tecnológicos proporciona una capacidad de innovar mayor, creando así nuevos productos y pudiendo mejorar la productividad de la propia empresa, con la posterior posición privilegiada en el mercado si consiguen una diferenciación significativa con sus competidores. En definitiva, variables como el número de innovaciones en los productos o la intensidad

de gastos en I+D+i son para Durán y Úbeda (2001) claves para la expansión e internacionalización de la empresa, debido a la generación de ventajas competitivas que producen.

Otro de los motivos para la decisión de no incluirlos en el estudio reside en el importante sesgo que aparecería en los resultados si no se hace una distinción previa de la muestra en función del tipo de empresa manufacturera, ya que en caso de no hacerla se daría una misma importancia a los recursos tecnológicos pertenecientes, por ejemplo, a una industria química y a una industria de alimentos, ambas incluidas en el grupo de empresas recogidas en la ESEE.

De esta manera, la aplicación estadística se centrará en factores que pueden parecer menos determinantes a la hora de exportar como son los recursos humanos y los recursos organizacionales, pero que numerosos autores han defendido como fundamentales.

La discusión sobre los recursos humanos puede resumirse en que son difíciles de imitar, debido al tiempo en obtener un recurso similar y Bloodgood et al (1996) afirman que las empresas que dispongan de estos recursos tendrán una mayor tendencia a exportar. El estudio de López (2004) sobre la exportación de las empresas pone énfasis en la importancia de estos recursos y hace una distinción de ellos, hablando de capital humano genérico y capital humano específico, donde el primero es en relación al nivel y tipo de estudios que tienen los individuos y el segundo en base a la formación y experiencia que los empleados adquieren trabajando en la empresa. Este mismo autor apoya sus premisas diciendo:

El principal activo de las empresas son las personas [...] de la calidad de los recursos humanos de que dispone una empresa va a depender su eficiencia y capacidad competitiva [...] una mayor cantidad de capital humano dotará a la empresa de una mayor capacidad para afrontar los retos competitivos que supone la entrada en los mercados internacionales.

Centrándose en los recursos organizacionales, Barney (1986) considera importante una cultura organizativa exclusiva como recurso de carácter estratégico para la empresa, pudiendo influir en la capacidad exportadora. Ampliando esta información, no tiene valor disponer de una buena dotación de recursos tecnológicos, de capital humano... si la empresa carece de una apropiada coordinación dentro de la misma y de ésta con el entorno, puesto que la importancia estratégica de los recursos organizacionales descansa en su capacidad para coordinar eficientemente las

actividades de la empresa, lo cual se consigue mediante la generación de buenas rutinas organizativas. (López, 2004; p-154)

4.2.2 Especificación de las variables

Como introducción es preciso mencionar que como se ha indicado en el anterior subapartado, las variables a especificar están relacionadas con los recursos humanos y los recursos organizacionales pero, es sabido que el Análisis discriminante solo puede emplear variables numéricas por lo que hay una considerable limitación en su elección, ya que la mayoría de las relacionadas con la organización de la empresa son cualitativas. De esta manera, las escogidas para el análisis no son las más óptimas debido a esta restricción y también por la necesidad de ajustarse a los datos disponibles de la muestra.¹⁹

Las variables con sus distintas funciones se clasifican de la siguiente manera:

- 1) Variable dependiente: es una variable dicotómica que indica si la empresa realiza ventas en el extranjero, es decir, si exporta (valor 1) o no (valor 0).
- 2) Variables independientes:
 - Recursos humanos: dentro de éstos se pretende analizar el capital humano genérico con dos variables; una definida como el porcentaje que el personal con títulos medios representa sobre el total del personal de la empresa y otra como la proporción de asalariados con estudios superiores sobre el total del personal.

En relación al capital humano específico se usan las variables que indican las proporciones de asalariados con contratos fijos y la de asalariados con contrato temporal sobre el total de asalariados, la elección de estas variables se justifican porque este cociente puede indicar la acumulación de experiencia específica adquirida con la permanencia de un empleado en la empresa.

Por último se decide añadir el tamaño de la empresa en función al total de empleados, la razón de incluirla en el análisis reside en que numerosos autores no ven una relación directa con el hecho de exportar y otros sí.

- Recursos organizacionales: las limitaciones de variables numéricas en esta categoría hace necesario formular una hipótesis: a una mayor antigüedad de la empresa, se asumen unas mejores rutinas organizacionales, como reflejo de habilidades organizativas que se acumulan en el tiempo y la consiguiente

¹⁹ Ver Anexo 2 para una mayor explicación de las variables empleadas.

estructura organizativa afianzada. Por lo que la variable a utilizar será la que recoge el año de constitución de la empresa.

4.3 Metodología

El objetivo clave de este trabajo es mostrar la utilidad que tiene el Análisis discriminante para un análisis cuantitativo de relaciones entre una variable dependiente con otras variables independientes (análisis de perfil) y para la predicción de pertenencias a los dos grupos, exportación y no exportación (análisis predictivo), de esta manera la metodología a usar será la utilización del Análisis discriminante en un programa informático llamado R, que centra su función en el análisis estadístico y gráfico.

Dentro de los períodos de tiempo disponibles en la base de datos a usar se escogen dos años diferentes, el 2006 y el 2010, para una posible comparación en el comportamiento de las empresas vivido en un momento previo a la crisis económica y en otro posterior. Por lo que se realizan dos Análisis discriminantes simples, uno para cada año.

4.4 Análisis año 2006

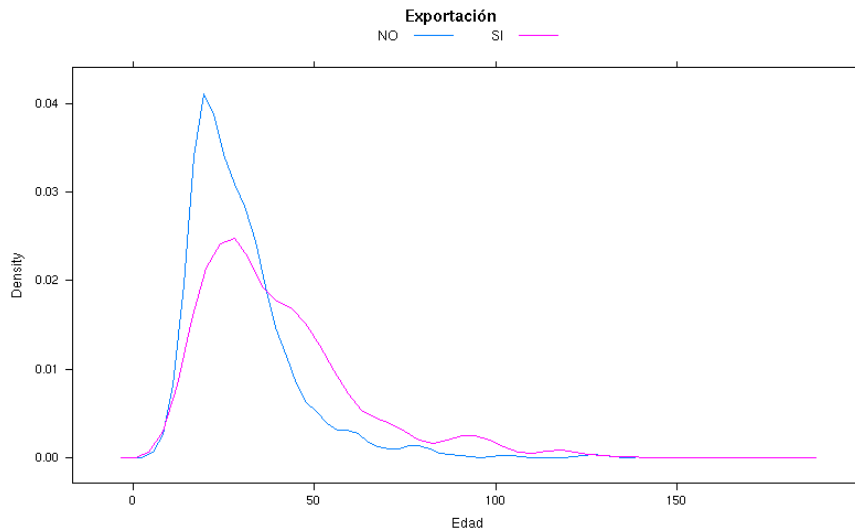
4.4.1 Análisis descriptivo previo

Antes de comenzar con la aplicación del Análisis discriminante se procede con un análisis descriptivo previo para poder observar inicialmente las diferencias en el comportamiento de las variables independientes en relación a la variable dependiente.

Se ha decidido realizar funciones de densidad de las variables individualmente para este análisis menos exhaustivo debido a la pobre visualización de las representaciones gráficas de diagramas de caja y matriz de dispersión.²⁰

²⁰ Ver Anexos 3.

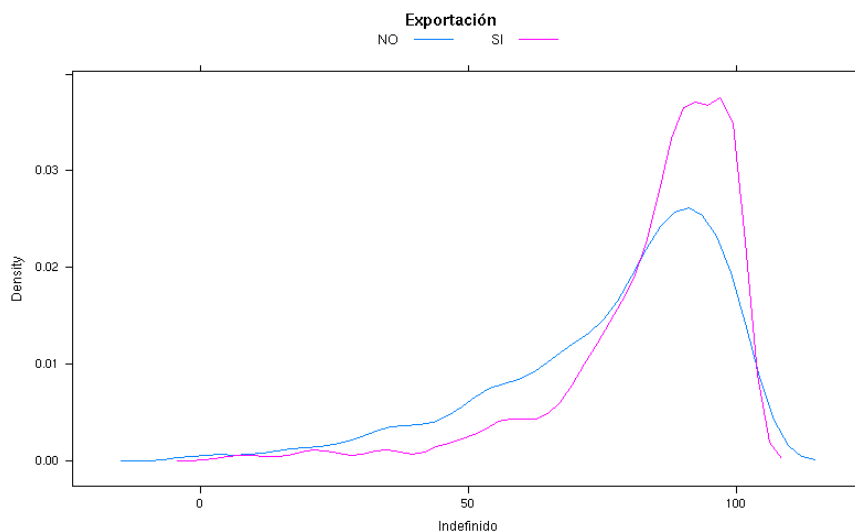
Gráfico 2. Funciones de densidad para el año de constitución de la empresa (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

Se puede apreciar claramente una diferencia en la distribución de la edad de las empresas, ya que la mayor parte de las no exportadoras se concentran en una zona próxima a los 20 años. Por otro lado, las empresas exportadoras se distribuyen más uniformemente y a partir de una edad cercana a los 50 años, el número de empresas que exportan es considerablemente mayor que las que no lo realizan.

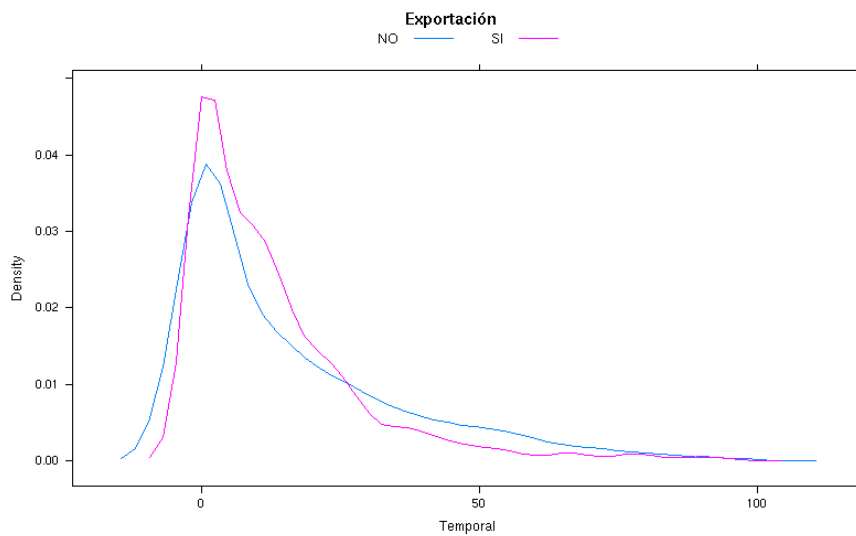
Gráfico 3. Funciones de densidad para el porcentaje de empleados con contrato indefinido (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

Aquellas empresas con un mayor porcentaje de empleados con contrato indefinido sobre el total de asalariados tienden a realizar exportaciones, pero aunque esta representación gráfica permite visualizar el grado de importancia de esta variable, la relevancia se nota en la proximidad del 80%, ya que cuando la empresa supera ese porcentaje de asalariados fijos se produce un gran incremento en el comportamiento de exportar. Para valores menores a ese porcentaje hay una mayor propensión a no exportar.

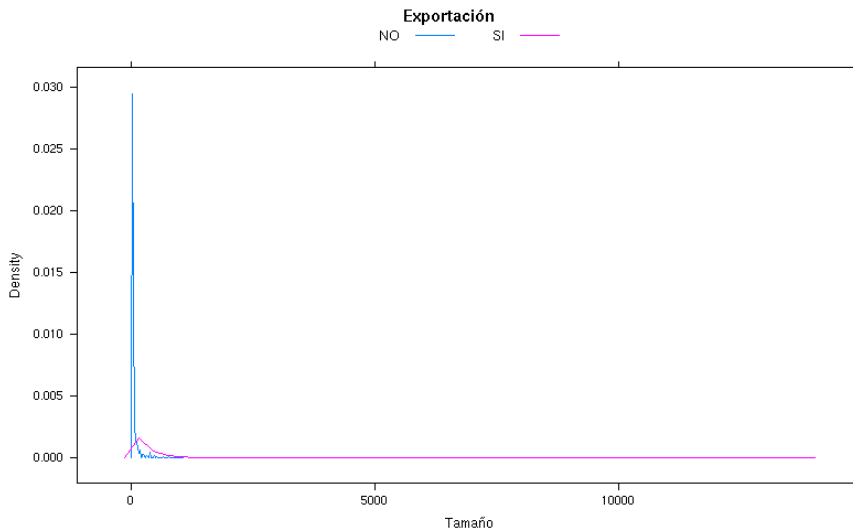
Gráfico 4. Funciones de densidad para el porcentaje de empleados con contrato temporal (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

En este caso, ante la variable del porcentaje de empleados con contrato temporal sobre el total del personal las diferencias no son tan marcadas, aunque se aprecia algo más de temporalidad en las empresas exportadoras. Esta diferencia reducida podría indicar que esta variable no es determinante en el análisis.

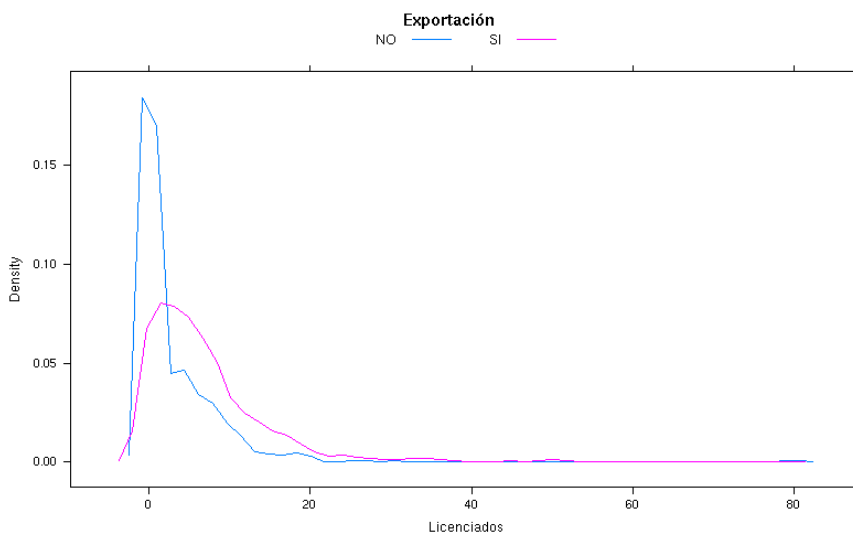
Gráfico 5. Funciones de densidad para el total de trabajadores (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

Este gráfico visiblemente revela una gran concentración de empresas no exportadoras cuando éstas tienen un tamaño pequeño, en otros términos, poco personal total. Mientras que las empresas exportadoras se distribuyen más ampliamente, tomando tanto valores altos como valores bajos. Es necesario el análisis más exhaustivo que tendrá lugar posteriormente para entender su contribución a la discriminación de los grupos.

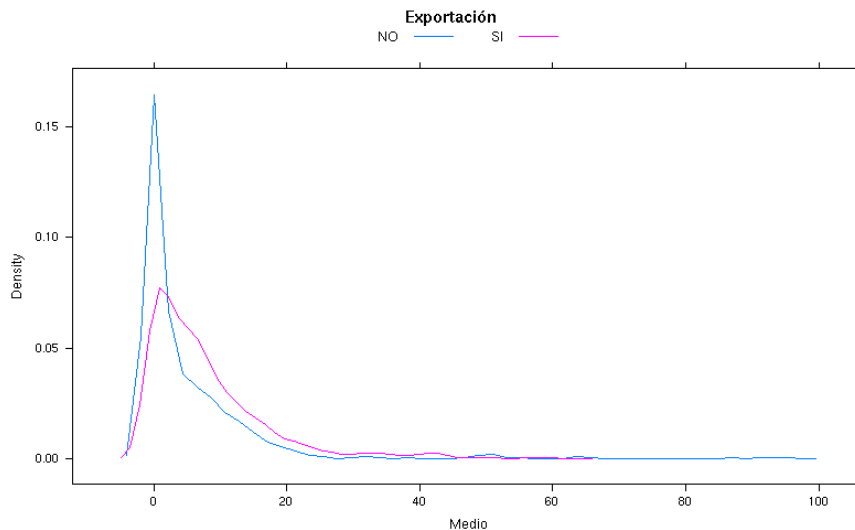
Gráfico 6. Funciones de densidad para el porcentaje de asalariados con estudios superiores (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

Se deduce claramente que las empresas no exportadoras tienen un porcentaje de recursos humanos con estudios superiores en la empresa muy inferior a las exportadoras, es decir, hay una relación directa positiva entre tener un mayor personal con estudios de ingenierías superiores o licenciaturas y realizar ventas en el extranjero.

Gráfico 7. Funciones de densidad para el porcentaje de asalariados con estudios medios (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

La relación entre exportación y titulados de estudios medios tiene un perfil cercano al de los licenciados pero la diferencia no está tan marcada, lo que podría indicar una menor relevancia de esta variable.

Para completar este análisis, se amplía la información de las representaciones de funciones de densidad previamente expuestas con la siguiente tabla que muestra los datos de la media y de los cuartiles (primero y tercero) de cada variable independiente en cada grupo clasificatorio.

Tabla 1. Cuartiles y medias por variable independiente (2006)

Variable: Edad	Variable: Indefinido
Factor: 0 (no exporta)	Factor: 0 (no exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
19.00 35.00 29.74	66.7 92.30 77.10
Factor: 1 (exporta)	Factor: 1 (exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
24.00 50.00 40.21	79.20 95.70 84.63

Variable: Temporal	Variable: Tamaño
Factor: 0 (no exporta)	Factor: 0 (no exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 25.00 15.83	14.00 39.00 50.06
Factor: 1 (exporta)	Factor: 1 (exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 17.90 12.68	36.00 308.50 350.50

Variable: Licenciados	Variable: Medio
Factor: 0 (no exporta)	Factor: 0 (no exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 4.30 2.83	0.00 7.05 5.28
Factor: 1 (exporta)	Factor: 1 (exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
1.90 9.10 6.99	1.50 11.10 8.03

Fuente: Elaboración propia a partir de datos de la ESEE.

Entre las conclusiones que se pueden extraer destaca la inferioridad del valor numérico de la media en el grupo 0 (no exporta) en comparación con el valor del primer cuartil en el grupo 1 (exporta) para la variable “indefinido”, y la gran igualdad de ambos valores para la variable “edad”. Esto se traduce en una gran diferencia entre los dos grupos en relación a los valores que estas variables toman en cada uno de ellos y previamente estas variables podrían diferenciar satisfactoriamente ambos grupos. A ambas se podría añadir “tamaño” que presenta igualmente grandes diferencias.

A su vez, hay diferencias en “licenciados” y “medio” pero en una menor proporción, y con la información que aporta la variable “temporal” se puede intuir un significado contrario debido a que la media es mayor en el grupo de no exportar que en el de exportar, este hecho de posible correlación negativa se comprobará más adelante.

Con este análisis descriptivo previo se pueden extrapolar unas posibles deducciones iniciales que se sintetizan en:

- Aparentemente las variables “tamaño”, “indefinido” y “edad” son importantes para la diferenciación de grupos.
- Las variables “licenciados” y “medio” parecen tener una moderada significación.
- La única correlación negativa podría provenir de la variable “temporal”.

Estas conclusiones a priori obtenidas del Análisis univariante se pueden ver afectadas por interrelaciones entre estas variables, por lo que ninguna de estas

afirmaciones se puede considerar definitiva hasta realizar el Análisis discriminante, donde si se tiene en cuenta el comportamiento conjunto de las variables

4.4.2 Análisis discriminante simple

Una vez detallado el análisis descriptivo previo se profundiza en el primordial, el Análisis discriminante simple. Las primeras dos etapas del proceso ya han sido puntualizadas (objetivos y diseño de la investigación) y por ello se continúa con la tercera fase conocida como la confirmación de los supuestos.

Para comprobar el supuesto de normalidad multivariante se ha usado el Test de Shapiro-Wilk²¹, donde se comprueba que en todos los casos las variables tienen un p-valor menor que un nivel de significación del 0.05, por lo que se rechaza la hipótesis de que los datos sigan una distribución normal. La importancia de este incumplimiento se ve reducida por trabajar con una muestra muy grande Tabachnick y Fidell (1983), si bien es necesario ser cuidadoso en los aspectos a los que podría influir la falta de normalidad, como son las probabilidades de pertenencia a los grupos y la significatividad de los parámetros de la función discriminante.

Tabla 2. Test de normalidad Shapiro-Wilk (2006)

Shapiro-Wilk normality test	
Variable: Edad	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.8082, p-value < 2.2e-16	W = 0.8695, p-value < 2.2e-16

Variable: Indefinido	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.8841, p-value < 2.2e-16	W = 0.8016, p-value < 2.2e-16

Variable: Temporal	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.7898, p-value < 2.2e-16	W = 0.7639, p-value < 2.2e-16

Variable: Tamaño	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.391, p-value < 2.2e-16	W = 0.2949, p-value < 2.2e-16

Variable: Licenciados	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.5164, p-value < 2.2e-16	W = 0.7286, p-value < 2.2e-16

Variable: Medio	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.4994, p-value < 2.2e-16	W = 0.7698, p-value < 2.2e-16

Fuente: Elaboración propia a partir de datos de la ESEE.

²¹ H_0 : p-valor > nivel de significación ($\alpha = 0.05$). Se cumple normalidad.

El incumplimiento de la normalidad afecta a la calidad de la prueba M de Box²², con la que se contrasta la homogeneidad de las matrices de varianza-covarianza. Además, como ya ha sido citado en el marco teórico del trabajo, es un supuesto muy difícil de cumplir, con menor importancia en las muestras grandes y superable con el uso del modelo en forma cuadrática a la hora de estimarlo.

Tabla 3. Prueba de homogeneidad de matrices varianza-covarianza M de Box (2006)

Box's M-test for Homogeneity of Covariance Matrices
Chi-Sq. (approx.) = 2048.045, df = 21, p-value < 2.2e-16

Fuente: Elaboración propia a partir de datos de la ESEE.

Dado que el objetivo de mi aplicación es estudiar la capacidad de discriminación de un grupo de variables empleando el Análisis discriminante, tiene más sentido valorar la linealidad a posteriori, antes que usarla como decisión de descarte de la aplicación del método. Además, con más de tres variables independientes es difícil observar la linealidad, y su ausencia se relacionaría solamente con una menor eficacia a la hora de separar el grupo de empresas exportadoras del grupo de las que no lo son.

El supuesto de ausencia de multicolinealidad y singularidad se verá superado, como se mencionó en la parte teórica, por el uso de un método de selección de variables para incluir en el Análisis discriminante.

Como resumen del análisis de los supuestos previos nos encontramos en una situación que no es recomendable el uso de Análisis discriminante, por lo menos en una versión lineal que es la que se está estudiando pero para enfrentarse a ello se van a seguir dos estrategias. La más importante es limitarse a una interpretación descriptiva en la que la tasa de acierto será el criterio principal para decidir la importancia de las variables en el proceso de discriminar el comportamiento exportador, renunciando a cualquier tipo de inferencia relacionada con el tema.

La segunda estrategia será estimar un Análisis discriminante cuadrático (QDA), que no se ve afectado por el incumplimiento de los supuestos como se comentó en el marco teórico, comparando los resultados obtenidos con los dos métodos.

²² Tiene una misma hipótesis nula que el Test Shapiro-Wilk, pero en este caso se cumple la igualdad de matrices.

Una vez desarrollada la fase de los supuestos tiene lugar la estimación del modelo (cuarta fase), ésta no tiene carácter simultáneo puesto que se realiza una estimación por etapas, seleccionando las variables más óptimas para incluir en la función discriminante.

En la parte teórica del trabajo se definió el método “stepwise” como el más correcto para la selección de las variables. Este método se emplea como forma abreviada de escoger un grupo de variables óptimo entre todos los casos posibles, sin embargo, el hecho de trabajar con plantillas de código de R y la capacidad de los ordenadores actuales permite la comprobación de todos los casos de forma más rápida, por lo que se ha preferido esta opción, haciendo los cálculos para modelos con cada vez un mayor número de variables hasta que el incremento en la tasa de acierto no supere un determinado nivel de tolerancia, del 0.001.²³

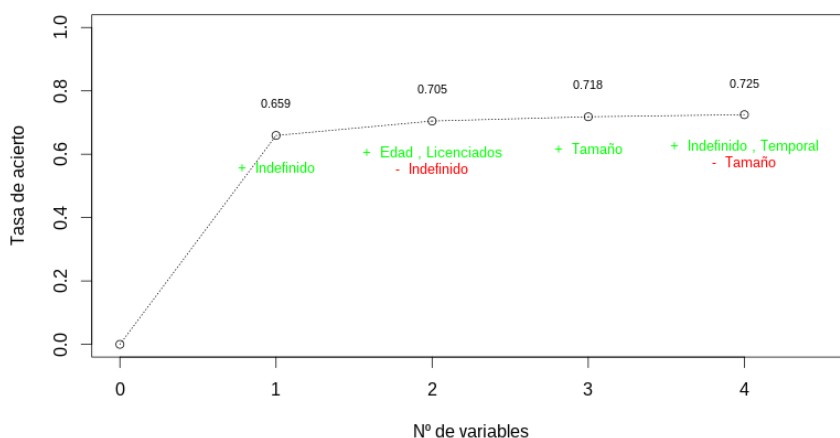
A continuación, se describirán los diferentes resultados obtenidos, que se corresponderán con la quinta y sexta etapa de interpretación y validación de resultados, separando el análisis en Análisis discriminante lineal y en Análisis discriminante cuadrático.

➤ Método LDA (Análisis discriminante lineal)

La relación de variables para este método puede seguirse en el siguiente gráfico, en el que se indican la entrada y la salida de variables, comenzando con una variable y acabando con cuatro, ya que la inclusión de más variables no mejora la tasa de aciertos del modelo.

²³ El empleo de este método también garantiza la ausencia de multicolinealidad y singularidad.

Gráfico 8. Proceso de selección de variables en método LDA (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

En el modelo con una variable la más importante es “indefinido”, que llega a una tasa del 65%. Sin embargo esta variable es mejorada por la combinación de variables “edad” y “licenciados”, o esas dos y “tamaño” para el modelo de tres. Finalmente se vuelve a recuperar “indefinido” y entra “temporal” para el modelo de cuatro variables que es el que se retiene finalmente con una tasa de acierto del 72%.

Usando las cuatro variables “licenciados”, “indefinidos”, “temporal” y “edad” se realizan los cálculos definitivos con los resultados que se van describiendo a continuación.

Con relación a la única puntuación discriminante, ésta tiene un valor de -0.1332, lograda a partir de los dos centroides que se muestran en esta tabla:

Tabla 4. Centroides de la función discriminante (2006)

Grupo	Función
0	- 0.6073600
1	0.3408004

Fuente: Elaboración propia a partir de datos de la ESEE.

Por lo que se percibe, hay una considerable distancia entre ambos grupos y supuestamente la diferencia entre ellos está bien explicada (en la sexta etapa se analizará en mayor profundidad mediante un estadístico).

De cara a la interpretación de la función discriminante a partir de las variables independientes incluídas en la discriminación se recoge en las siguientes tablas los coeficientes estandarizados y la matriz de estructura.

Tabla 5. Coeficientes estandarizados (2006)

Variable	Función
Edad	0.4106945
Indefinido	1.8222092
Temporal	1.5821483
Licenciados	0.5065895

Fuente: Elaboración propia a partir de datos de la ESEE.

Tabla 6. Matriz de estructura (2006)

Variable	Función
Edad	0.5894891
Indefinido	0.4723047
Temporal	-0.2062682
Licenciados	0.6355635

Fuente: Elaboración propia a partir de datos de la ESEE.

Los primeros permiten una comparación de los coeficientes independientemente de la dimensión de las variables, mientras que la matriz incluye las correlaciones entre las variables y la función discriminante, indicando qué variables están más relacionadas con ella.

De esta manera, la cantidad de empleados con estudios de ingeniería superior o licenciaturas son el tipo de capital humano que tiene mayor influencia a la hora de discriminar entre empresas que exportan o no, seguido de la antigüedad de la propia empresa y en tercer lugar, de los trabajadores que disponen de un contrato fijo. Todas estas variables tienen una influencia positiva, mientras que la variable que informa de los asalariados con contrato temporal actúa de forma contraria, a mayor cantidad de personal con contrato temporal, mayor tendencia a que la empresa no exporte. Aún así, la correlación negativa de ésta tiene un peso mucho menor que el resto de las otras tres y es menos influyente.

Para concluir el análisis se realiza la validación de los resultados (sexta y última fase). Se empieza con la obtención del valor de correlación canónica para conocer la bondad del ajuste, es decir, cómo viene explicada la diferencia entre los grupos. Este estadístico toma un valor medio de 0.4143, lo que sería una explicación baja de las diferencias. También es preciso mencionar que la exportación de una empresa manufacturera depende de más variables y que el propósito de este trabajo es conocer el efecto de aquellas que están relacionadas con los recursos internos de la empresa, más concretamente los humanos y organizacionales.

A continuación, para saber el ratio de aciertos del análisis en relación a su capacidad de predicción para la clasificación se obtuvo la matriz de confusión, que nos da una tasa de acierto del 72%, valor semejante al obtenido con validación cruzada.

Tabla 7. Matriz de confusión con todos los datos (2006)

		Predicted	
		0	1
Original	0	195	287
	1	82	777
Tasa de acierto = 72.48%			

Fuente: Elaboración propia a partir de datos de la ESEE.

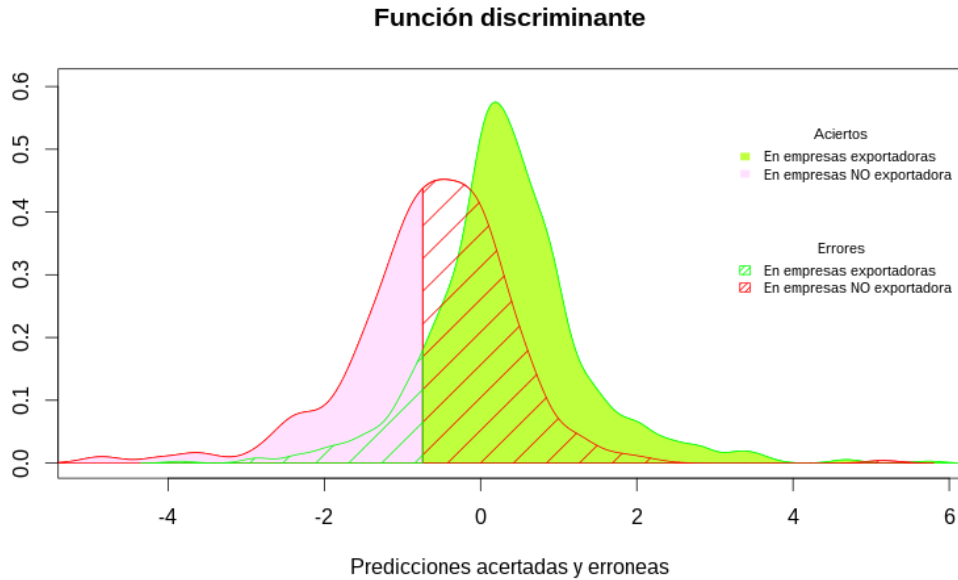
Tabla 8. Matriz de validación cruzada (2006)

		Predicted	
		0	1
Original	0	194	288
	1	84	775
Tasa de acierto = 72.25%			

Fuente: Elaboración propia a partir de datos de la ESEE.

En este gráfico se puede ver el grado de acierto según la empresa exporta o no:

Gráfico 9. Predicciones acertadas y erróneas en método LDA (2006)



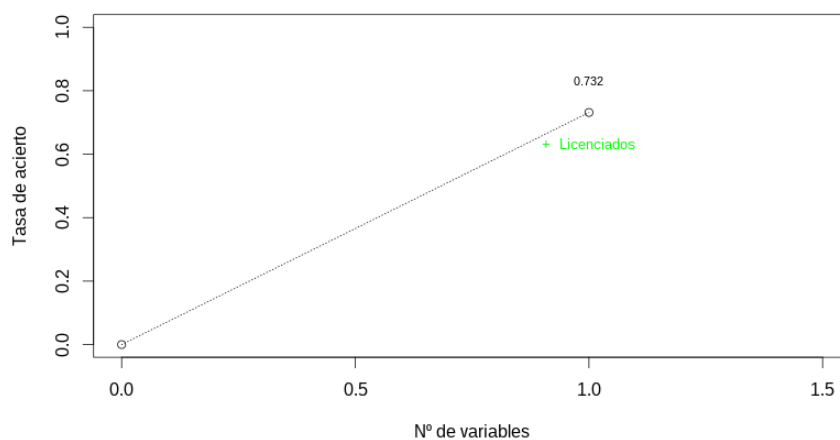
Fuente: Elaboración propia a partir de datos de la ESEE.

Es claro el gran acierto en la predicción en el grupo 1 (empresas exportadoras), llegando casi a su totalidad. Por otro lado hay una mayor limitación en el grupo 0, en el que se predice correctamente casi el 50% de las empresas no exportadoras.

➤ Método QDA (Análisis discriminante cuadrático)

Con este otro método se alcanza una situación muy diferente, incluyendo solamente una variable en el análisis se llega a prácticamente la misma tasa de acierto, siendo ésta incluso ligeramente superior (73%). Como se ve en el siguiente gráfico, la singular variable es “licenciados”.

Gráfico 10. Proceso de selección de variables en método QDA (2006)



Fuente: Elaboración propia a partir de datos de la ESEE.

Para concluir la última fase del análisis también se ha realizado la validación cruzada, dando ésta un valor de 73.07%

El hecho de llegar al mismo resultado con menos variables reside en la mejoría en la práctica de este método por la superación de los supuestos iniciales, pero a pesar de esta situación, en las conclusiones se tendrán en cuenta los dos métodos porque la validación final de los resultados es prácticamente la misma.

Extrapolando todo lo visto anteriormente, para el año 2006 se obtienen las siguientes conclusiones:

1. Las deducciones obtenidas en un análisis descriptivo previo a partir de funciones de densidad, medias y cuartiles de cada variable independiente en cada grupo de empresas permiten una ligera aproximación a la explicación, pero no son suficientes para conocer la influencia de cada una sobre el acto de exportar o no. Un claro ejemplo ha sido la aparente transcendencia del total de empleados de la empresa, que una vez habiendo profundizado en el Análisis discriminante se ha confirmado su falta de significación.
2. Los recursos humanos de la empresa basados en el capital genérico son más importantes que los basados en el capital específico a la hora de que la empresa tenga un comportamiento exportador. Concretamente, un alto nivel de estudios por parte del personal de la empresa es más valioso que una elevada

experiencia y formación de éstos tras años en la empresa, aún así, ambos son de muy notable importancia y necesarios para tener en cuenta.

3. La hipótesis con relación a los recursos organizacionales puntualizada en la especificación de las variables tiene una gran influencia a la hora de vender productos en el extranjero, teniendo incluso un mayor peso que el porcentaje de personal con contratos indefinidos o temporales. De esta manera, el que la empresa disponga de una estructura organizativa afianzada por el paso del tiempo y rutinas organizacionales superiores a las de empresas jóvenes, influye positivamente para poder exportar.
4. Se disponían de dos variables que podrían contribuir negativamente a la hora de exportar, ambas relacionadas con los recursos humanos, una de ellas medía el porcentaje de titulados medios en la empresa y la otra el porcentaje de asalariados con contrato temporal. Mientras que la primera relacionada con el capital humano genérico no se muestra relevante, el hecho de tener un alto porcentaje de personal con contrato temporal (falta de formación y experiencia dentro de la empresa) influye en que la empresa no realice exportaciones, aunque su contribución es en menor medida que el resto de variables.
5. El Análisis discriminante cuadrático no requiere del cumplimiento de los supuestos iniciales, llega a prácticamente una misma tasa de acierto en el objetivo de clasificar nuevos individuos a los dos grupos. Aún así, es importante destacar que el criterio de selección de variables en este método solo precisa del uso de una variable, mientras que el Análisis discriminante lineal son necesarias tres más.
6. La capacidad de predicción futura para nuevas observaciones será muy óptima si éstas se clasifican en empresas exportadoras. Por otro lado aunque siendo considerablemente buena, ésta capacidad será más limitada en el caso de incluirse en las no exportadoras porque la tasa de error es mayor.

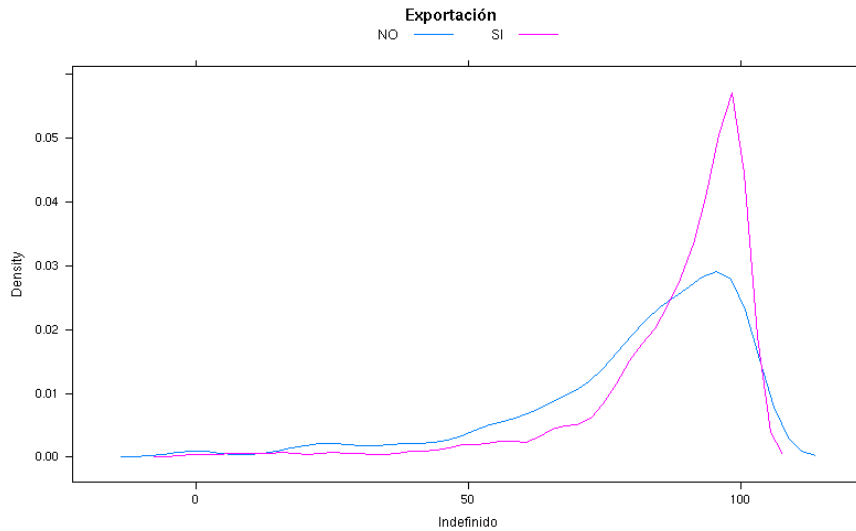
4.5 Análisis año 2010

4.5.1 Análisis descriptivo previo

Para una comparación con el año 2006 se incluyen las representaciones gráficas de aquellas variables que parecen haber sufrido más cambios, prescindiendo de las que

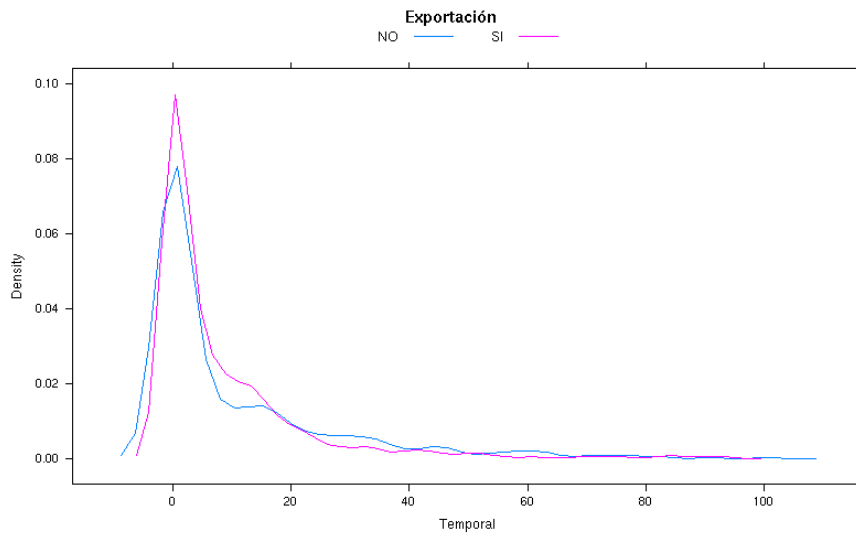
no han experimentado transformaciones importantes: la edad de la empresa y los empleados con titulaciones medias en estudios.²⁴

Gráfico 11. Funciones de densidad para el porcentaje de empleados con contrato indefinido (2010)



Fuente: Elaboración propia a partir de datos de la ESEE.

Gráfico 12. Funciones de densidad para el porcentaje de empleados con contrato temporal (2010)

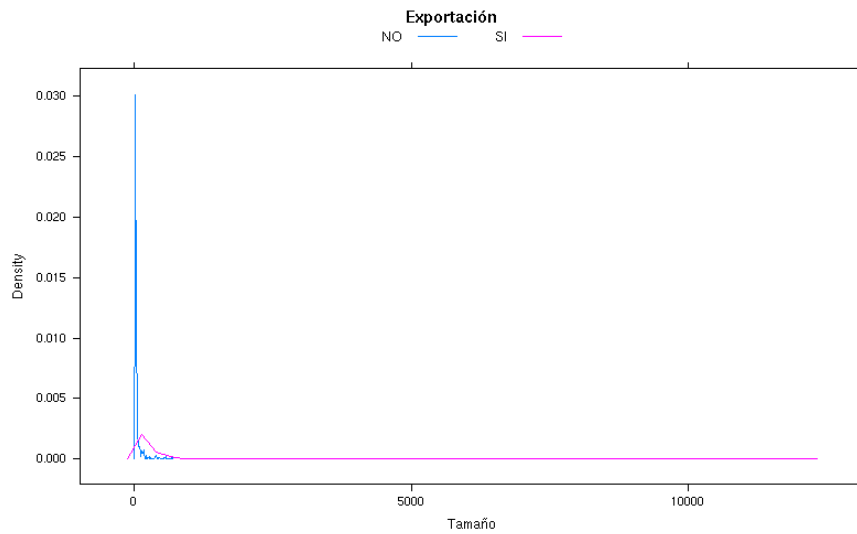


Fuente: Elaboración propia a partir de datos de la ESEE.

²⁴ Ver Anexo 4 para los diagramas de caja y matriz de dispersión.

Las variables relacionadas con el capital humano específico han vivido cambios sustanciales tras cuatro años, hay una menor dispersión de los empleados con contratos temporales e indefinidos en las empresas exportadoras, concentrándose los primeros en valores más bajos y los segundos en valores más altos. Esta situación se puede deducir que es causada por uno de los efectos de la crisis económica, el despido de personal con contrato temporal y el mantenimiento de aquellos trabajadores con más tiempo de experiencia en la propia empresa.

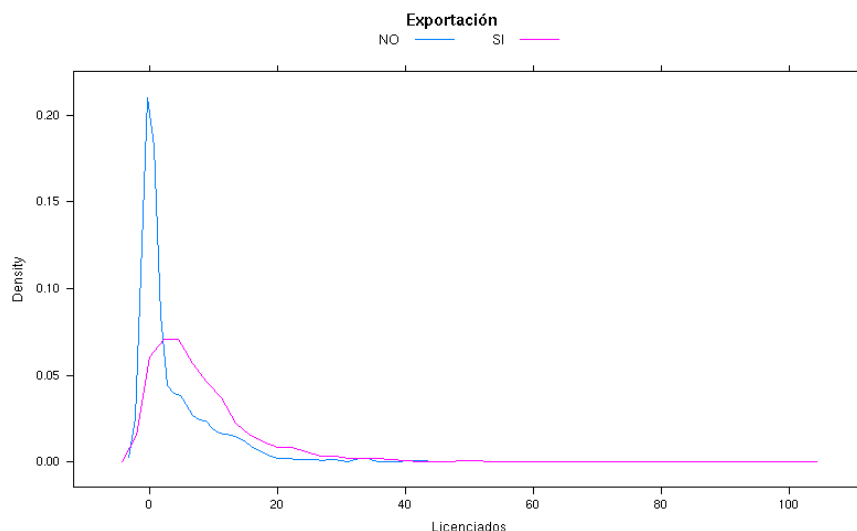
Gráfico 13. Funciones de densidad para el total de trabajadores (2010)



Fuente: Elaboración propia a partir de datos de la ESEE.

El cambio notable en relación al tamaño total tiene lugar en las empresas exportadoras, siendo esta función de densidad menos plana, una posible traducción de la pérdida de personal en la empresa.

Gráfico 14. Funciones de densidad para el porcentaje de asalariados con estudios superiores (2010)



Fuente: Elaboración propia a partir de datos de la ESEE.

Hay un ligero descenso de personal con estudios altos en las empresas exportadoras y una mayor concentración de éstos en valores bajos en las que no exportan.

Agregando la tabla de medias y cuartiles (primero y tercero) por variable independiente:

Tabla 9. Cuartiles y medias por variable independiente (2010)

Variable: Licenciados	Variable: Indefinido
Factor: 0 (no exporta)	Factor: 0 (no exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 5.60 3.64	72.10 95.70 80.60
Factor: 1 (exporta)	Factor: 1 (exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
2.47 10.82 8.07	83.38 98.40 87.89

Variable: Temporal	Variable: Tamaño
Factor: 0 (no exporta)	Factor: 0 (no exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 15.00 10.31	11.00 32.00 37.97
Factor: 1 (exporta)	Factor: 1 (exporta)
1st Qu. 3rd Qu. Mean	1st Qu. 3rd Qu. Mean
0.00 11.92 8.71	30.00 267.00 308.80

Fuente: Elaboración propia a partir de datos de la ESEE.

En esta tabla se puede ver que ha habido aumentos considerables en los valores de las variables “licenciados” e “indefinidos” tanto para el grupo 0 de empresas como para

el grupo 1. Se intuye que esta situación procede del descenso del personal de ambos grupos de empresas como indica la importante bajada de los valores de la variable “tamaño”, manteniendo los perfiles de trabajadores con estudios superiores y que llevan una larga duración trabajando en la propia empresa y eliminando aquellos con contratos temporales (como se aprecia en la bajada de sus valores) o de estudios medio.

Más que posibles deducciones con este análisis descriptivo previo se puede presentar una cuestión que se tratará de responder en el siguiente apartado: ¿el hecho de haber cambios tan significativos dentro de las empresas en relación con sus recursos internos se verá reflejado a la hora de influir en la clasificación de los grupos de manera distinta en el Análisis discriminante simple?

4.5.2 Análisis discriminante simple

La tercera y cuarta etapa tiene un igual procedimiento que el realizado en el 2006 por lo que para no reiterarse en la explicación simplemente se adjuntan las salidas del programa R.

Tabla 10. Test de normalidad Shapiro-Wilk (2010)

Shapiro-Wilk normality test	
Variable: Edad	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.8098, p-value < 2.2e-16	W = 0.8734, p-value < 2.2e-16

Variable: Indefinido	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.8303, p-value < 2.2e-16	W = 0.7222, p-value < 2.2e-16

Variable: Temporal	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.6645, p-value < 2.2e-16	W = 0.6281, p-value < 2.2e-16--

Variable: Tamaño	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.4045, p-value < 2.2e-16	W = 0.275, p-value < 2.2e-16

Variable: Licenciados	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.6413, p-value < 2.2e-16	W = 0.7747, p-value < 2.2e-16

Variable: Medio	
Factor: 0 (no exporta)	Factor: 1 (exporta)
W = 0.5146, p-value < 2.2e-16	W = 0.7661, p-value < 2.2e-16

Fuente: Elaboración propia a partir de datos de la ESEE.

Tabla 11. Prueba de homogeneidad de matrices varianza-covarianza M de Box (2010)

Box's M-test for Homogeneity of Covariance Matrices
 Chi-Sq (approx.) = 2201.769, df = 21, p-value < 2.2e-16

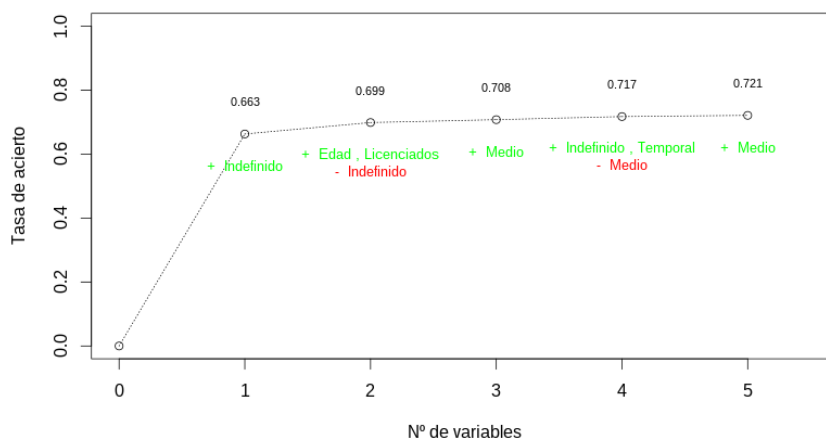
Fuente: Elaboración propia a partir de datos de la ESEE.

En la quinta y sexta fase de interpretación y validación de resultados se efectúa la misma división anterior: Análisis discriminante lineal y Análisis discriminante cuadrático.

➤ Método LDA (Análisis discriminante lineal)

Las variables a incluir en la función discriminante se incrementan en una unidad, siendo éstas: “edad”, “indefinido”, “temporal”, “licenciados” y “medio”. El método de selección de variables llegó hasta el modelo con cinco variables, obteniendo una tasa de acierto próxima al 2006:

Gráfico 15. Proceso de selección de variables en método LDA (2010)



Fuente: Elaboración propia a partir de datos de la ESEE.

La similitud en comparación con el método de selección de las variables para el año 2006 es que “edad”, “indefinido”, “temporal” y “licenciados” siguen incorporándose en la función discriminante. La diferencia es que se incluye “medio”.

La puntuación discriminante toma un valor de -0.1381, y viendo la tabla de los dos centroides se aprecia una menor lejanía entre ellos que en el 2006.

Tabla 12. Centroides de la función discriminante (2010)

Grupo	Función
0	- 0.5886903
1	0.3124897

Fuente: Elaboración propia a partir de datos de la ESEE.

Examinando los coeficientes y estandarizados y la matriz de estructura:

Tabla 13. Coeficientes estandarizados (2010)

Variable	Función
Edad	0.3928227
Indefinido	1.2110674
Temporal	0.9876959
Licenciados	0.4991973
Medio	0.1822151

Fuente: Elaboración propia a partir de datos de la ESEE.

Tabla 14. Matriz de estructura (2010)

Variable	Función
Edad	0.5673192
Indefinido	0.4913561
Temporal	-0.1229984
Licenciados	0.6509100
Medio	0.3646819

Fuente: Elaboración propia a partir de datos de la ESEE.

El orden de las tres variables que más discriminan a la hora de exportar siguen siendo las mismas: la que más lo hace es la cantidad de empleados con estudios de ingeniería superior o licenciaturas, seguida por la edad de la empresa y por el personal con contrato indefinido. Todas estas variables siguen teniendo una influencia positiva, pero en estos cuatros años ha habido cambios en los pesos de sus relevancias, aunque

bastante insignificantes: tanto la primera como la tercera amplían en un 0.02, mismo valor en el que mengua la segunda.

La distinción surge con la aparición de los titulados medios trabajando en la empresa (antes no significativa), superando incluso a aquellos con contrato temporal. Llama la atención que los primeros no presentan un efecto contrario, es decir, influyen a la hora de que la empresa exporte, mientras que los segundos continúan con una correlación negativa aunque en menor proporción que en el 2006.

Para concluir el análisis se realiza la validación de los resultados (sexta y última fase). El valor de correlación canónica en este caso es de 0.3944, por lo que hay una disminución en su capacidad para explicar la clasificación de los grupos si se compara con el obtenido en el 2006, siendo así menos satisfactoria

Una bajada de las tasas de aciertos en comparación con el 2006 también ocurre tanto en la matriz de confusión, con un resultado del 72.11%, como en la validación cruzada, que se sitúa en el 71.43%.

Tabla 15. Matriz de confusión con todos los datos (2010)

Original	Predicted	
	0	1
0	164	301
1	73	803
Tasa de acierto = 72.11%		

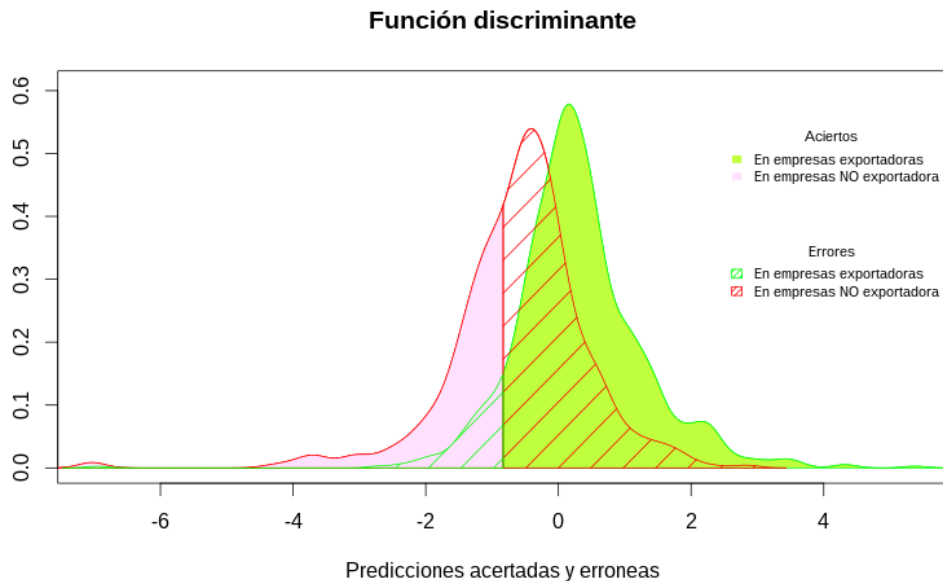
Fuente: Elaboración propia a partir de datos de la ESEE.

Tabla 16. Matriz de validación cruzada (2010)

Original	Predicted	
	0	1
0	127	338
1	45	831
Tasa de acierto = 71.43%		

Fuente: Elaboración propia a partir de datos de la ESEE.

Gráfico 16. Predicciones acertadas y erróneas en método LDA (2010)



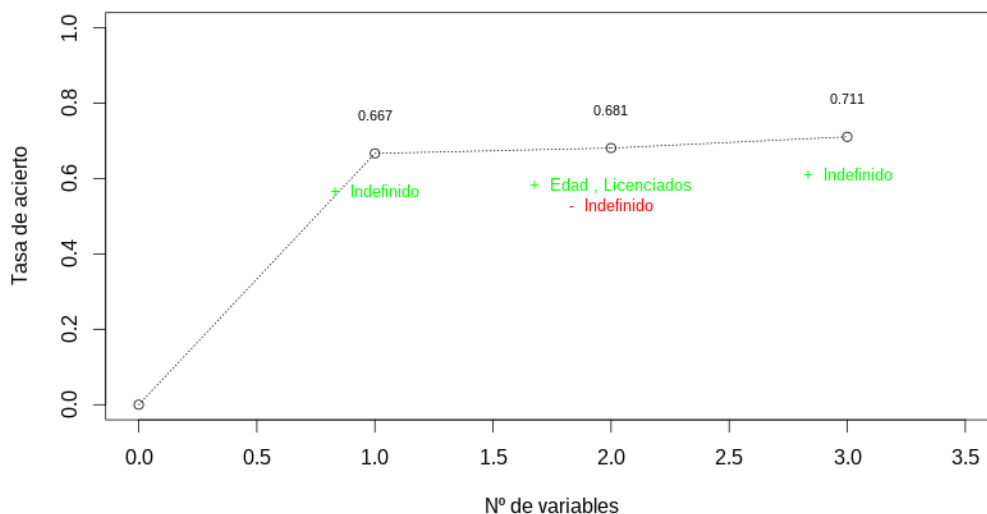
Fuente: Elaboración propia a partir de datos de la ESEE.

En el gráfico se ve un mayor solapamiento de ambos grupos por lo que se puede ver un empeoramiento en la capacidad predictiva del análisis. Aunque siga siendo alta para el grupo de empresas exportadoras, se ha experimentado un descenso importante para el otro grupo (no exportadoras).

➤ Método QDA (Análisis discriminante cuadrático)

Con este otro método se incluyen dos variables más que el experimentado en el 2006, eliminando así la explicación con solamente una variable. Esta variable se conserva (“licenciados”), y se incorporan “edad” e “indefinido”.

Gráfico 17. Proceso de selección de variables en método QDA (2010)



Fuente: Elaboración propia a partir de datos de la ESEE.

Concluyendo la última fase del análisis se ha hecho la validación cruzada, dando ésta un valor de 68.9%, siendo tanto menor que la obtenida en el método QDA del año 2006, como en la obtenida en el método LDA de este año 2010.

Extrapolando todo lo comentado precedentemente, se logran las siguientes conclusiones para el año 2010:

1. En el análisis descriptivo previo se puede percibir a partir de las funciones de densidad, medias y cuartiles de cada variable independiente un considerable descenso del total de empleados en cada grupo de empresa tras la crisis económica. Esta situación de importante pérdida de recursos humanos no se traduce como relevante a la hora de influenciar sobre el acto de exportar o no, ya que tanto en el Análisis discriminante lineal como en el Análisis discriminante cuadrático se ha confirmado su falta de significación al no ser introducida en ninguno.
2. El porcentaje de trabajadores con alto nivel de estudios, junto con el año de constitución de ésta y el porcentaje de empleados con contratos indefinidos, siguen siendo las tres variables más discriminantes ya que son las únicas incluidas en el Análisis discriminante cuadrático y las que presentan mayores correlaciones en el Análisis discriminante simple. El paso de estos cuatro años

no ha variado el orden de importancia de ellas pero sí su correlación, las dos variables relacionadas con los recursos humanos la han aumentado, mientras que la vinculada con los recursos organizacionales ha decrecido. Tras esto se sospecha que puede estar el hecho de que se decida mantener tanto el capital humano genérico con mayor nivel de cualificación como el capital humano específico con superior experiencia en la propia empresa, con un mayor porcentaje de ambos dentro de ella y una mayor influencia de su presencia a la hora de exportar. De todas maneras, los cambios en las correlaciones no son muy significantes.

3. El mayor cambio si se compara el análisis con el del 2006 tiene lugar en el Análisis discriminante lineal, donde aparece el porcentaje de titulados medios en la empresa, antes inexistente en la selección final de variables llevado a cabo por el método correspondiente. Esta variable no contribuye negativamente a la hora de exportar como se podría pensar a priori, ya que su correlación es positiva. Por último, la variable de porcentaje de empleados con contrato temporal sigue manteniendo la correlación negativa pero descendiendo en 0.08 puntos porcentuales.
4. El Análisis discriminante simple nos muestra una tasa de acierto prácticamente igual al del 2006. Por el contrario, la tasa de acierto del Análisis discriminante cuadrático (68.9%) no solo disminuye en relación al año 2006, sino que también lo hace en comparación al LDA del propio 2010 (71.43%). De esta manera, se concluye que el análisis de ambos métodos es útil y conveniente ante posibles sesgos.
5. La capacidad de predicción futura para nuevas observaciones seguirá siendo mejor si nos interesan las empresas exportadoras y mucho más limitada en el caso de interesar las no exportadoras. Comparando la función discriminante con el año 2006 es primordial destacar una mayor proximidad entre ambos centroides en el 2010, aumentando así la tasa de error en la predicción.

Conclusiones

En la introducción del presente trabajo se propuso mostrar la utilidad de la técnica estadística llamada Análisis discriminante en algún área de conocimiento relacionada con la titulación. Las finalidades de este método son dos: la descripción de la influencia de diversas variables independientes sobre una variable explicada categórica y la predicción en clasificar individuos en los distintos grupos que forman esa variable dependiente.

Para manifestar esa utilidad, se ha incluido un análisis práctico enfocado desde una perspectiva de la “Teoría de los Recursos y Capacidades de la empresa”, sintetizando sus principios en que los recursos internos son claves a la hora de llevar a cabo una estrategia empresarial, la cual se ha definido en este caso como comportamiento exportador. De esta manera, las variables a incluir en el análisis son la variable dependiente que contiene solamente dos categorías (exportar o no exportar) y un conjunto de variables independientes relacionadas con los recursos humanos y los recursos organizacionales (proporción de empleados que son ingenieros superiores o licenciados, proporción de trabajadores con títulos medios, porcentaje de personal con contrato indefinido, porcentaje de asalariados con contrato temporal, total de empleados y año de constitución de la empresa).

Las dos primeras variables informan sobre el capital humano genérico (nivel de estudios), las dos siguientes sobre el capital humano específico (experiencia y formación en la propia empresa), el total de asalariados sobre el capital humano total y sobre el año de constitución de la empresa se realizó una hipótesis para relacionarla con los recursos organizacionales, el hecho de una mayor antigüedad supone una estructura organizativa más consolidada.

Debido al amplio tipo de empresas en la sociedad, se encuadra el análisis en las empresas industriales manufactureras españolas, con una muestra representativa de éstas (“Encuesta sobre Estrategias Empresariales”).

Se efectúan dos Análisis discriminantes simples, donde la justificación de denominarse simple reside en el hecho de que la variable explicada es dicotómica, y también se ejecutan distintos métodos de clasificación dentro de los análisis para dos años distintos, 2006 y 2010.

Respondiendo a las preguntas planteadas en la introducción sobre la función descriptiva del análisis se descomponen de dos maneras, analizando los años en conjunto y por separado:

Por un lado, analizando los años de forma conjunta se confirma que los recursos humanos de la empresa basados en el capital genérico son más significativos que los basados en el capital específico en relación con el comportamiento exportador de las empresas industriales manufactureras españolas, en concreto una de cada tipo: un alto nivel de estudios por parte del personal de la empresa tiene una mayor validez que una alta experiencia y formación tras años en ella. Aún así, las dos son las variables que más contribuyen en el análisis.

La hipótesis planteada inicialmente sobre la antigüedad de la empresa se cumple, el disponer de una estructura organizativa afianzada por el paso de los años y rutinas organizacionales superiores a las de empresas jóvenes influyen positivamente a la hora de vender productos en el extranjero.

La única variable que tiene una correlación negativa es la definida por el personal con contrato temporal, en otras palabras, el hecho de tener un elevado porcentaje de personal con este tipo de perfil influye en que la empresa no realice exportaciones, aunque su relevancia es pequeña ya que está muy poco correlacionada con la función discriminante.

Por el otro lado, analizando los años por separado no se observan cambios muy significativos ya que la transformación más importante tras la crisis económica ha sido la reducción de empleados (variable definida por el total de asalariados en la empresa), la cual nunca se ha mostrado relevante en ningún análisis y método. Hay ligeros cambios en la proporción de influir por parte de algunas variables, pero ninguno revelador.

Sobre la función predictiva del análisis se concluye que es muy satisfactoria para las empresas exportadoras y más limitada en las no exportadoras tanto en el año 2006 como en el 2010, aunque el número de errores es mayor en este último.

Respondiendo a la última pregunta de qué método de clasificación en el Análisis discriminante es más satisfactorio no se llega a una conformidad. Se sabe por los resultados del 2006 que el incumplimiento de los supuestos de la variables hace el Análisis discriminante lineal más condicionado, superando el cuadrático esa limitación y llegando a una misma tasa de acierto con un menor uso de variables. Sin embargo, en el año 2010 la tasa de acierto del LDA es considerablemente superior al QDA, siendo necesario un uso de ambos para hacer comparaciones porque pueden aparecer sesgos tanto en uno como en otro.

Bibliografía

Andrews, K. R. (1977). *El concepto de estrategia de la empresa*. Pamplona: Universidad de Navarra.

Arce, C. (1993). *Escalamiento Multidimensional, una técnica Multivariante para el Análisis de datos de proximidad y preferencia*. Barcelona: PPU

Balsalobre, C.; Del Río L.; y Maurandi A. (s.f.). *Fundamentos estadísticos para investigación. Introducción a R*. Recuperado de: <http://www.bubok.es/libros/223207/Fundamentos-estadisticos-para-investigacionIntroduccion-a-R>

Barney, J.B. (1986). *Strategic factor markets: Expectations, luck and business strategy* en *Management Science*, 32, pp. 1231-1241.

Barney, J. B. (1991). *Firm resources and sustained competitive advantage* en *Journal of Management*, 17, pp. 99-120.

Bloodgood, J.M.; Sapienza, H.J; y Almeida, J.G. (1996). *The internationalization of new high-potencial U.S. ventures: antecedents and outcomes* en *Entrepreneurship Research: Global Perspectives*, pp. 207-247.

Boqué, R.; Maroto, A. (s.f.). *El Análisis de la varianza (ANOVA)*. Recuperado de <http://www.quimica.urv.es/quimio/general/anovacast.pdf>

Caro, R.; García, F. (2011). *Historias de Matemáticas - ¡Qué Historia esto de la Estadística!* en *Pensamiento Matemático*, pp. 1-9.

Castillo I.; Portela M. (2002). *Tecnología y competitividad en la Teoría de los Recursos y Capacidades* en Informa Economía, 308, pp. 35-42.

Cea, M. A. (2002). *Análisis Multivariable. Teoría y práctica en la Investigación Social*. Madrid: Editorial Síntesis.

Cuadras, C.M. (1981). *Métodos de Análisis Multivariante*. Barcelona: Eunibar.

Cuervo, A. (1993). *El papel de la empresa en la competitividad* en Papeles de Economía Española, 56, pp. 363-367.

De la Fuente, S. (2011). *Análisis discriminante*. Recuperado de <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/DISCRIMINANTE/analisis-discriminante.pdf>

De la Gaza, J.; Morales, B.; y González B. (2013). *Análisis estadístico Multivariante. Un enfoque teórico y práctico*. México D.F: McGraw-Hill.

Díaz, V. (2002). *Técnicas de Análisis Multivariante para Investigación Social y Comercial*. Madrid: Ra-Ma.

Dillon W.R.; Goldstein M. (1984). *Multivariate Analysis. Methods and Applications*. New York: Wiley & Sons.

Durán, J.; Ubeda, F. (2001). *Fuente de ventaja competitiva y análisis geográfico de la empresa multinacional española*. Revista ICADE, 54, septiembre -diciembre.

Fariñas, J; Jaumandreu J. (1994). *La Encuesta sobre Estrategias Empresariales: Características y usos* en Economía Industrial. Septiembre-octubre, pp. 109-119.

Flury, B. (1997). *A First Course in Multivariate Statistics*. Bloomington: Editorial Board.

Gil, J.; García, E.; Rodríguez, G. (2001). *Análisis discriminante*. Madrid: La Muralla S.A.

Grant, R.M. (1991). *The Resource-Based Theory of Competitive Advantage: Implications for Strategy Formulation*. California Managem. Rev., pp. 114-135.

Hair, J.F.; Anderson, R.E.; Tatham, R.L.; y Black, W.C. (1999). *Análisis Multivariante*. Madrid: Prentice Hall Iberia.

Huerta, E. (1993). *Contribuciones de la organización industrial a la división estratégica de la empresa* en Proyecto social: Revista de relaciones laborales, 1, pp. 85-108.

Huberty, C. J. (1984) *Issues in the use and interpretation of discriminant analysis*, en Psychological Bulletin, 95, pp. 156-171.

Iglesias, S.; Sulé, M.A. (2003). *Introducción al Análisis Multivariable*. En Lévy, J.P.; Varela, J. (2003). *Análisis Multivariable para las Ciencias Sociales*. (pp. 1-40). Madrid: Pearson Prentice Hall.

Knudsen, C. (1995). *Theories of the firms, strategic management, and leadership*, en Resource-based Evolutionary Theories of the Firm. Boston: Kluwer Academic Publishers, pp. 179-217.

Lévy, J.P.; Varela, J. (2003). *Análisis multivariable para las Ciencias Sociales*. Madrid: Pearson Prentice Hall.

López, J. (2004). *Análisis de la actividad exportadora de la empresa: una aproximación desde la Teoría de Recursos y Capacidades*. A Coruña. Recuperado de: http://ruc.udc.es/bitstream/2183/1161/1/LopezRodriguezJose_opt.pdf

Medina, D.R. (1998). *Una visión integral de la empresa basada en los recursos, el conocimiento y el aprendizaje* en Investigaciones Europeas de Dirección y Economía de la Empresa, 2, pp.77-90.

Pardo, A. (s.f). *Guía de SPSS*. Recuperado de <http://www.listinet.com/bibliografia-comuna/Cdu311-F36A.pdf>

Penrose, E. (1959). *The Theory of the Growth of the Firm*. London: Basil Blackwell.

Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.

Pérez, C. (2004). *Técnicas de Análisis Multivariante de datos*. Madrid: Pearson Educación.

Pérez, C. (2013). *Análisis Multivariante de datos. Aplicaciones con IBM SPSS, SAS Y STATGRAPHICS*. Madrid: Garceta Grupo Editorial.

Porter, M. (1980). *Competitive Strategy. Techniques for analyzing industries and competitors*. New York: Free Press.

Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. New York: Harper & Row, Publishers.

Suárez, J.; Ibarra, S. (2002). *La teoría de los recursos y las capacidades. Un enfoque actual de la estrategia empresarial* en Anales de estudios económicos y empresariales, 15, pp. 63-89.

Tabachnick, B.G.; Fidell, L.S. (1983). *Using multivariate statistics*. Nueva York: Harper & Row Publishers.

Tacq, J. (1997). *Multivariate analysis techniques in Social Science Research from problem to analysis*. London: Sage Publications.

Uriel, E.; Aldás, J. (2005). *Análisis Multivariante aplicado. Aplicaciones al Marketing, Investigación de Mercados, Economía, Dirección de Empresas y Turismo*. Madrid: Thomson.

Anexo 1. Resumen de contenidos del cuestionario de la ESEE

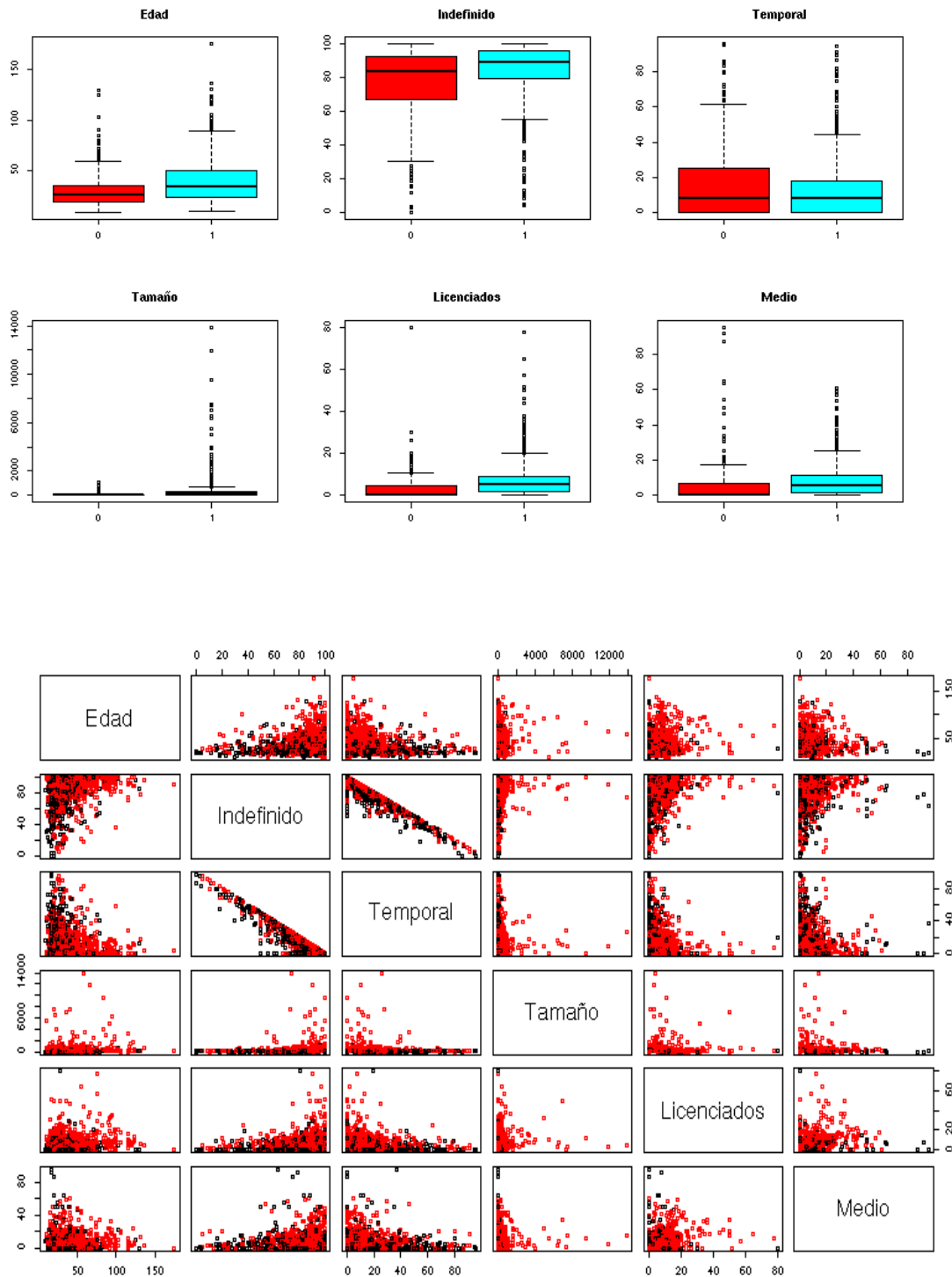
- a) **ACTIVIDAD, PRODUCTOS Y PROCESOS DE FABRICACIÓN:** incluye algunas características de la empresa y de sus operaciones, como son sus establecimientos industriales y no industriales, su forma jurídica y algunas participaciones significativas en su capital social, actividad y características de los productos manufacturados, tecnología utilizada.
- b) **CLIENTES Y PROVEEDORES:** recoge información relacionada con el tipo de clientes de la empresa, el destino final de los productos que manufactura, canales de distribución empleados, actividades de promoción comercial, características de los proveedores y contratación de servicios.
- c) **COSTES Y PRECIOS:** aporta información sobre los precios pagados por la empresa y sobre la política de precios de venta.
- d) **MERCADOS SERVIDOS:** recoge información relacionada con los mercados servidos por la empresa, de forma que supongan en conjunto, al menos, el 50% de sus ventas totales y queden identificados por líneas de productos, tipo de clientes, ámbito geográfico u otras características. La información se refiere a la cuota de mercado de la empresa, al número de competidores y la cuota de los principales, a la variación experimentada por los precios durante el año y los motivos de dicho cambio.
- e) **ACTIVIDADES TECNOLÓGICAS:** recoge preguntas relacionadas con actividades de I+D, registro de patentes, innovaciones de producto y de proceso y pagos e ingresos por licencias y asistencia técnica.

- f) **COMERCIO EXTERIOR:** incluye exportaciones e importaciones, distribución por áreas geográficas y vías.
- g) **EMPLEO:** recoge personal ocupado en la empresa, composición según tipos de contrato, categorías y titulación y otros datos dirigidos a determinar la jornada efectiva de trabajo durante el año.
- h) **DATOS CONTABLES:** esta última parte incorpora tres bloques de información. El primero es un resumen de partidas de la cuenta de pérdidas y ganancias. El segundo recoge el valor de las inversiones en inmovilizado material. El tercero es un resumen de las partidas más importantes del balance de la empresa.

Anexo 2. Tabla de las variables a emplear en el análisis

Identificador	Nombre de la variable	Definición de la variable
Licenciados	Proporción de licenciados e ingenieros	Porcentaje que los ingenieros superiores y licenciados representan sobre el total del personal a 31 de diciembre
Medio	Proporción de titulados medios	Porcentaje que los empleados con títulos medios representan sobre el total del personal de la empresa a 31 de diciembre
Indefinido	Personal asalariado fijo	Porcentaje que el personal con contrato indefinido representada sobre el total de los empleados de la empresa a 31 de diciembre
Temporal	Personal asalariado temporal	Porcentaje que el personal con contrato temporal representada sobre el total de los empleados de la empresa a 31 de diciembre
Tamaño	Personal total	Total de empleados que trabajan en la empresa a 31 de diciembre
Edad	Año de constitución de la empresa	Recoge el año de constitución de la empresa

Anexo 3. Diagramas de caja y matriz de dispersión (2006)



Anexo 4. Diagramas de caja y matriz de dispersión (2010)

