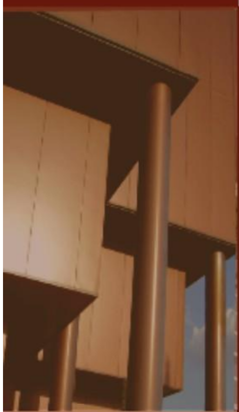
Proyecto Fin de Carrera

Ingeniería en Organización Industrial

Un estudio de la clasificación de deuda pública en las CCAA españolas mediante análisis discriminante

Autor: Pablo Antonio Galván Delgado

Tutor: Ester Gutiérrez Moya



Dep. Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2016

Proyecto Fin de Carrera

Ingeniería en Organización Industrial

Un estudio de la calificación de deuda pública en las CCAA españolas mediante análisis discriminante

Autor:

Pablo Antonio Galván Delgado

Tutor:

Ester Gutiérrez Moya

Dep. Organización Industrial y Gestión de Empresas I

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2017

Proyecto Fin de Carrera: Aplicación del Método Análisis Discriminante a las Calificaciones de deuda pública de las CCAA españolas

Autor: Pablo Antonio Galván Delgado

Tutor: Ester Gutiérrez Moya

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla
, 2017

El Secretario del Tribunal

Resumen

Este trabajo persigue dos objetivos, en primer lugar, conocer la influencia de veinte variables económico-financieras sobre las calificaciones de deuda pública de las CCAA españolas, y, en segundo lugar, predecir cuál serían estas calificaciones en años venideros.

Las calificaciones se distribuyeron en cuatro grupos (categorías de calificación) en los que se encuadraron las distintas CCAA según su año de la calificación y se obtuvo una muestra con un total de 66 casos para analizar. Se utilizó la técnica estadística de clasificación supervisada Análisis Discriminante Múltiple.

El conjunto de las variables ‘Deuda/PIB’, ‘Carga Financiera / Ingresos Corrientes’ y ‘Importaciones % PIB’ resultaron ser las variables con un poder discriminante mayor de un total de ocho variables a las que se le aplicó la técnica multivariante. Relativo a la predicción, el resultado fue muy satisfactorio para CCAA clasificadas en los grupos 1 y 4 ya que el porcentaje de acierto de los casos analizados fue elevado. No obstante, los errores encontrados para los grupos 2 y 3 fueron mayores. De esta manera se consiguió conocer las variables con una mayor influencia en la calificación de las CCAA y se obtuvieron funciones capaces de predecir las calificaciones futuras a pesar de que el número de variables se redujo de veinte a ocho y también a que no se logró toda la información anual de las distintas CCAA.

Palabras clave: Calificación, Deuda, Comunidades Autónomas, Análisis Discriminante.

Abstract

The purpose of this paper was, firstly to determine the influence of twenty economic and financial variables on the public debt ratings of the Spanish Autonomous Communities and secondly, to predict what these ratings would be in the following years.

The ratings were divided into four groups (rating categories) in which the different Spanish Autonomous Communities were framed according to their year of qualification and a sample was obtained with a total of 66 cases to analyze. It was used the statistical technique supervised Multiple Discriminate Analysis.

The set of variables 'Debt / GDP', 'Financial Burden / Current Income' and 'Imports% GDP' proved to be the variables with a greater discriminating power of a total of eight variables to which the multivariate technique was applied. Regarding the prediction, the result was very satisfactory for ACs classified in groups 1 and 4 since the percentage of correctness of the cases analyzed was high. However, the errors found for groups 2 and 3 were higher. In this way it was possible to know the variables with a greater influence on the qualification of the Autonomous Communities and obtained functions capable of predicting the future qualifications even though the number of variables was reduced from 20 to 8 and also to that it was not achieved at all the annual information of the different regions.

Key words: Ratings, Debt, Autonomous Communities, Discriminant Analysis.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. Introducción

Las administraciones públicas recurren a la emisión de títulos de renta fija para la financiación de sus proyectos. Desde hace unos años, esta actividad ha ido en aumento y ha favorecido al desarrollo de las calificaciones crediticias realizadas por agencias especializadas como Standard & Poor´s, Moody´s o Fitch, siendo estas tres agencias las más importantes a nivel internacional de las 74 que existen en todo el mundo, controlando aproximadamente el 95% del mercado.

La misión que tienen las agencias de calificación es informar con sus calificaciones, de forma que predicen si una inversión de un producto financiero (letras del tesoro, bonos, acciones, etc.) es arriesgada o no, o bien, analizan si es probable que un inversor cobre los cupones, intereses o dividendos y que recupere el dinero cuando el producto se ha amortizado o liquidado. Es muy común que gobiernos, bancos o empresas soliciten a una agencia una evaluación cuando van a emitir deuda o van a solicitar financiación. Estas calificaciones van a servir a inversores y deudores para orientarles en los tipos de interés al que concederían la financiación. Normalmente coinciden los anuncios de calificación cuando los países u otras instituciones solicitan deuda.

El presente Proyecto Fin de Carrera tiene como objetivo conocer qué indicadores de los comúnmente utilizados en análisis financiero son los que poseen un poder discriminante mayor para el cálculo de las calificaciones de las distintas Comunidades Autónomas españolas realizadas por las agencias de calificación Standard & Poor´s, Moody´s y Fitch. Asimismo, la obtención de una función que pueda clasificar o predecir la calificación de una Comunidad Autónoma. En conclusión, se aborda el problema de clasificación o discriminación de las calificaciones crediticias (también comúnmente expresada en castellano mediante el anglicismo rating) de las CCAA españolas.

De esta manera, se pretende identificar qué indicadores presentan una mayor capacidad para discriminar entre las distintas puntuaciones asignadas por las agencias de calificación para las Comunidades Autónomas españolas, así como el poder predecir qué calificación van a tener dichas Comunidades Autonómicas en años venideros. Para ello, se utiliza el método estadístico Análisis Discriminante Múltiple que será expuesto y desarrollado en loscapítulos4 y 6 de este Proyecto Fin de Carrera, respectivamente.

Las Comunidades Autónomas tienen distintas posibilidades de endeudamiento siendo una de ellas la emisión de títulos de renta fija y su colocación en el mercado de capitales. Hace algunos años, esta posibilidad no era utilizada frecuentemente por las administraciones autonómicas y locales, aunque sí por la Administración del Estado español. En años posteriores ha comenzado a ser más habitual la emisión de valores de renta fija y en los últimos años las Comunidades Autónomas recurren a la emisión de títulos como vía de financiación externa.

Esta forma de financiación es bien conocida por administraciones públicas de otros países, como es el caso de EEUU. Todo esto potenció el desarrollo de las agencias de calificación ya comentadas anteriormente (Standard &Poor´s, Moody´s y Fitch).

El objeto de este trabajo es el análisis de la capacidad de un conjunto representativo de variables económica y financieras para predecir las calificaciones asignadas por las agencias de calificación y para ello, se ha recopilado las calificaciones realizadas por Standard &Poor´s, Moody´s, Fitch para todas las Comunidades Autónomas españolas con la excepción de Ceuta y Melilla entre los años 2009 y 2014, así como la información económica-financiera de dichos años. De esta manera se han elaborado un conjunto de indicadores que en un principio son candidatos para explicar la calificación. Mediante la aplicación de la técnica estadística multivariante, Análisis Discriminante Múltiple se pretende determinar qué variables del conjunto seleccionado tienen alto poder discriminante, relevancia e importancia hoy en día siendo determinantes para la calificación final realizada por las agencias.

En el **capítulo 2** se definirá el objeto del Proyecto.

A continuación, en el **capítulo 3** se explicará la importancia de las empresas de calificación y cómo países (y en el caso concreto las Comunidades Autónomas españolas) recurren a éstas para solicitar financiación.

Seguidamente, en el **capítulo 4** se desarrollará la técnica estadística multivariante, Análisis Discriminante.

Posteriormente, en el **capítulo 5** se definirán las variables económicas seleccionadas para el análisis y se realizará el diagnóstico de las hipótesis básicas.

Inmediatamente después, en el**capítulo6** se encontrará la aplicación del Análisis Discriminante Múltiple a los datos recogidos.

En el **capítulo 7** se indicarán las conclusiones tras la aplicación del método estadístico.

Antes de finalizar, en el **capítulo 8** se presentará la bibliografía utilizada en este Proyecto Fin de Carrera.

Y, por último, se presentarán los **anexos** con tablas y figuras no incluidas en los capítulos del Proyecto Fin de Carrera.

2. Objeto del Proyecto Fin de Carrera

Este Proyecto Fin de carrera tiene por objeto el de conocer aquellas variables económico-financieras de un total de 20 que tienen influencia (o una mayor influencia) en la calificación (media de las tres calificaciones) otorgada por las tres agencias de calificación más importantes a nivel mundial (Standard & Poor´s, Moody´s y Fitch) a las Comunidades Autónomas españolas.

Una segunda finalidad es servir como herramienta de clasificación a partir de un grupo teórico de variables medibles de carácter económico-financiero, proporcionando una o varias funciones que predigan la clasificación en cuatro grupos (determinados éstos por la variable dependiente) que tendrán las CCAA en años posteriores.

Para el cálculo de los estadísticos se utilizará la aplicación informática estadística IBM SPSS.

3. Los Índices de Calificación de Deuda Pública

Cuando una administración pública recurre a los mercados de capitales para financiar sus políticas de gasto, los acreedores querrán saber cuál es la situación financiera en la que se encuentra dicho organismo. Desgraciadamente, en la mayoría de los casos resulta imposible, sobre todo para inversores particulares e incluso institucionales, debido al gran número de instituciones públicas que existen y la dificultad de conocer la situación económico-financiera real de cada una de ellas.

Por este motivo, las agencias de calificación cumplen el papel de otorgar una calificación de la deuda de las Administraciones Públicas que acuden al mercado para su financiación, permitiendo de esta forma que los futuros inversores conozcan su solvencia, que se traducirá en una prima de riesgo incorporada al tipo de interés de cada caso. La relación entre el tipo de interés y la solvencia se establece con la calificación. Una calificación elevada significará un menor tipo de interés de la deuda y viceversa, una calificación baja conducirá a un mayor tipo de interés de la deuda. Así mismo, una nota de calificación por alguna agencia tendrá que ver con la obtención de préstamos y la colocación de títulos entre los posibles inversores.

Estas agencias de calificación tienen una implicación directa en las administraciones públicas ya que la cantidad de inversión externa dependerá en gran medida las calificaciones proporcionados por estas agencias.

No obstante, existen también estudios que afirman que estas agencias no son del todo neutrales. Broto y Molina (2014) señalan que “la agencia Standard & Poor´s baja las calificaciones con un ritmo superior al que las sube. Existen fuertes asimetrías, es decir, las caídas suelen ser muy breves y salvajes mientras que las subidas son procesos muchos más lentos y que además no siempre se vuelve a la calificación inicial.”

En definitiva, las agencias de calificación realizan un juicio a las instituciones para informar a los inversores sobre si estas instituciones van a ser capaces de hacer frente a su deuda y a los intereses devengados periódicamente devolviéndola, junto con los intereses una vez cumpla el crédito o préstamo. La información que analizan las agencias comprende normalmente un período de cinco a ocho años.

Las empresas de calificación utilizadas en el presente Proyecto Fin de Carrera son:

- **Standard & Poor´s:**

  Comienza en 1860 cuando publica *Historia de Ferrocarriles y canales en los Estados Unidos* donde ofrecía la situación financiera y operativa de las compañías de ferrocarriles en Estados Unidos. Su fundador fue Henry Varnum Poor que junto a su hijo ocho años después fundaron H.V. y H. W. Pobre Co. que publicaba dos manuales actualizados anualmente.

En 1906 se funda Standard Statistics Bureau por parte de Luther Lee Blake donde se proporcionaba información financiera sobre aquellas empresas estadounidenses que no estaban relacionadas con el ferrocarril.
No fue hasta 1941 cuando se fusionan estas dos empresas naciendo así Standard & Poor´s.
Standard & Poor´s no ha podido evitar las críticas debido a su imparcialidad así como a algunas predicciones nefastas que han realizado en los últimos años. Por citar algún caso, cabe destacar el de Islandia en el año 2001. Se calificó AAA (máxima calificación, fiable y estable) por la nueva legislación bancaria liberal. Años más tarde, en 2008 el sistema bancario de Islandia entró en crisis. Otro caso reciente es el de Italia en 2012, donde se inició un proceso de investigación judicial debido a las acusaciones sobre S&P donde se le atribuía la especulación en bolsa para lucrarse a sí misma.

- **Moody´s:**

Fue fundada en 1909 por John Moody. En 1924, las calificaciones de Moody's cubrían casi el 100 por ciento del mercado de bonos de EE.UU. En 1970 empezó a trabajar con deuda comercial y a recibir beneficio a los emisores de bonos por la calificación que realizaba e igualmente a los inversores. Moody´s ha ido creciendo a medida que los años pasaban, en 1975 sólo estaban en tres países. Contaban con 33 países en 1990 y actualmente están presentes en más de 100 países.

Una de las más sonadas críticas sufridas por la compañía fue en 2011 con Portugal. La deuda soberana de este país fue calificada de “calidad de crédito cuestionable”. Esto molestó tanto a Portugal como a la Unión Europea y se le acusó de especular con los mercados y atacar a la economía europea. Algunas empresas públicas y de infraestructuras portuguesas fueron también rebajadas a pesar de poseer una situación financiera sólida e importantes beneficios.

- **Fitch:**

Se creó en 1913 en Nueva York por John Knowles. Ya en el año 1997 se combinó con IBCA Limited y sede en Londres, y es propiedad de una sociedad holding francesa. Fitch es la empresa más pequeña de las tres grandes de la calificación (Standard & Poor´s, Moody´s y Fitch).

Fitch tampoco se ha podido librar de críticas debido a grandes pérdidas en el mercado de obligaciones de deuda colateralizadas (producto financiero estructurado y respaldado por activos financieros) después de asignar la máxima calificación por la agencia.

La tabla 1 muestra las calificaciones realizadas en el presente año 2016 a las Comunidades Autónomas españolas. En dicha tabla se observa que todas las CCAA tienen similares calificaciones, clasificadas en el grupo 3. Tan sólo existen dos excepciones, Cataluña que tiene la peor calificación (grupo 4) y Navarra que posee la mejor (grupo 1).

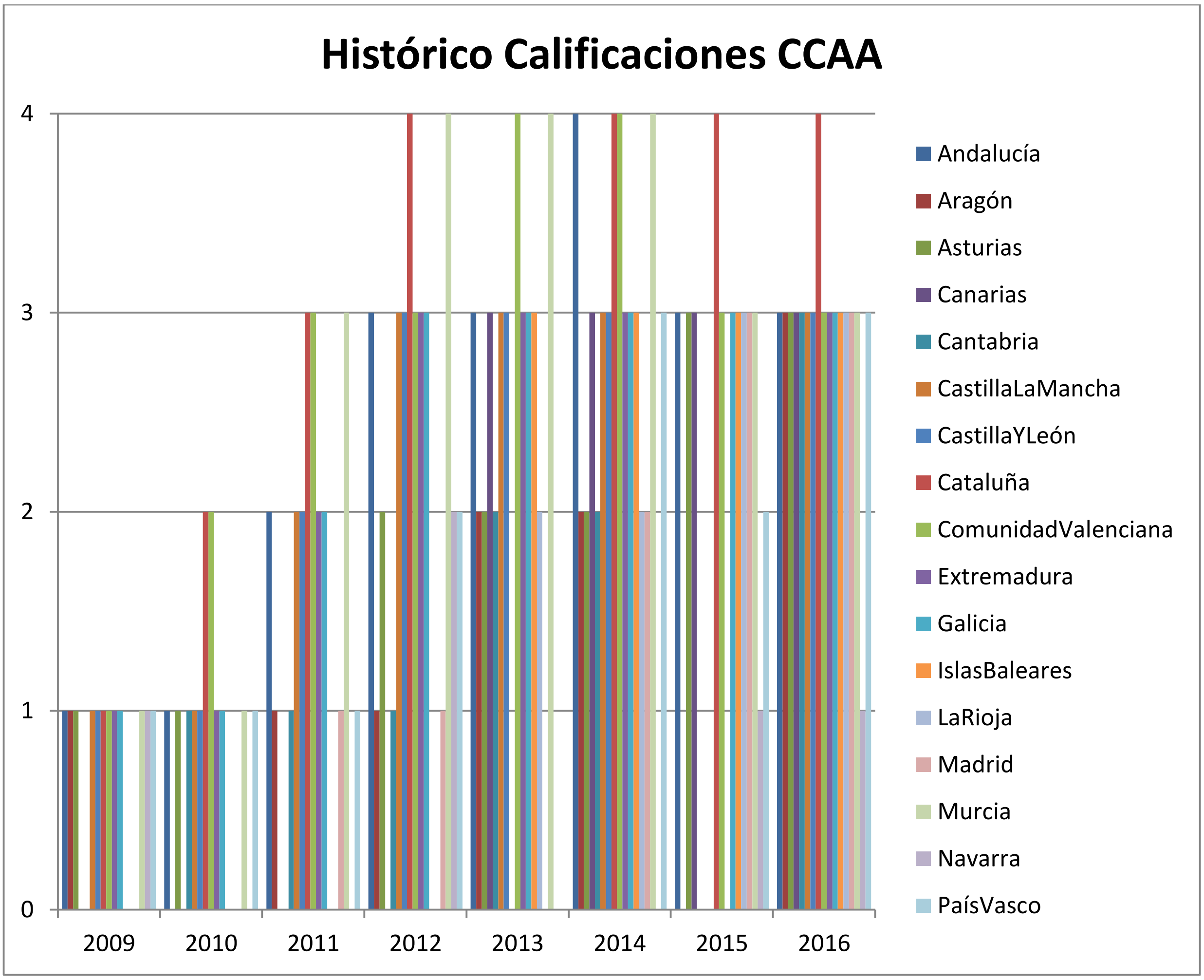
Tabla 1:
Calificaciones crediticias de las CCAA en el año 2016.

CCAA	Moody´s	S&P	Fitch	Grupo
Andalucía	Baa3	-	BBB-	3
Aragón	-	BBB-	-	3
Asturias	-	-	BBB	3
Canarias	-	-	BBB-	3
Cantabria	-	-	BBB	3
Castilla La Mancha	Baa2	-	-	3
Castilla y León	Baa2	-		3
Cataluña	Ba3	B+	BB/B	4
Comunidad Valenciana	Ba2	-	BBB-	3
Extremadura	Baa3	-	-	3
Galicia	Baa2	-	-	3
Islas Baleares	-	BBB-	-	3
La Rioja	-	-	BBB	3
Madrid	Baa2	-	BBB	3
Murcia	Ba2	-	BBB-	3
Navarra	-	A	-	1
País Vasco	Baa1	-	BBB+	3

Elaboración Propia.

En la figura 1 se muestra la evolución de las calificaciones en las distintas CCAA desde el año 2009 a 2016:

Figura 1:
Evolución de las calificaciones crediticias en las CCAA desde 2009 a 2016.



Se observa claramente que las calificaciones fueron aumentando progresivamente con los años, llegándose al año 2014 en el que más de la mitad de las CCAA fueron calificadas en el grupo 3 y 4 (esta nomenclatura se explicará y desarrollará en el capítulo 5 del presente Proyecto Fin de Carrera). Las calificaciones 3 y 4 representan una probabilidad de impago de las inversiones altas o muy altas, respectivamente.

4. Metodología: Análisis Discriminante

4.1 Introducción

El Análisis Discriminante forma parte del grupo de técnicas denominadas de clasificación supervisada. Estos métodos de clasificación supervisada es una técnica que deducirá a partir de unos datos previos una función. La salida de esta función puede ser un valor numérico, o como en el caso que se aborda, una etiqueta de clase, que serán los grupos de clasificación.

El Análisis Discriminante se utiliza para clasificar a distintos individuos en grupos o poblaciones a partir de un conjunto de variables sobre estos individuos a los que se pretende clasificar. Cada individuo solamente va a pertenecer a un grupo único. La pertenencia a uno u otro grupo se introduce en el Análisis Discriminante mediante una variable categórica que toma tantos valores como grupos existentes haya. La categórica se denomina también *variable dependiente*. Por otro lado, a las variables que se utilizan para realizar la clasificación se denominarán *variables clasificadoras* (se conocen también por variables criterio o predictoras). En el Análisis Discriminante, estas variables clasificadoras van a aparecer en las denominadas funciones discriminantes, que son las que realmente se utilizarán en el proceso de clasificación.

Como puntualizó anteriormente, el Análisis Discriminante es una herramienta que permite asignar o clasificar nuevos individuos dentro de grupos previamente reconocidos o definidos. Este análisis se aplica para fines tanto explicativos como predictivos. Cuando se utiliza el análisis para un fin explicativo, lo que se pretende es determinar cuál es la contribución que tiene cada variable en la clasificación de los individuos. En cambio, cuando su utilización es para un fin predictivo, se trata de establecer el grupo al que va a pertenecer un individuo únicamente con los valores de las variables clasificadoras.

El Análisis Discriminante está muy relacionado con el Análisis Multivariante de la varianza con un factor, aunque el papel que juegan los distintos tipos de variables está invertido en uno y otro método. Así, en el Análisis de la Varianza, la variable categórica (el factor) es la variable explicativa, mientras que en el Análisis Discriminante la variable categórica es precisamente la variable dependiente.

El análisis comienza con una tabla de datos de *n* individuos en que se han medido *p* variables cuantitativas independientes o denominadas también explicativas. Una variable cualitativa adicional dependiente, con dos o más categorías (para el presente proyecto cuatro categorías), ha definido por otros medios el grupo a que cada individuo pertenece. Se trata, pues, de una tabla de dimensión $n_x \cdot (p + 1)$ en que cada caso está configurado con un perfil y una asignación de grupo. A partir de ella se obtendrá un modelo matemático discriminante contra el cual será contrastado el perfil de un nuevo individuo cuyo grupo se desconoce para que, en función de un resultado numérico, sea asignado al grupo más probable. Cuanto mejor sea esta información inicial indudablemente será más fiable el resultado de las asignaciones posteriores.

El modelo discriminante también puede ser contrastado contra sí mismo, es decir, mediante su aplicación a los propios individuos de la tabla ignorando momentáneamente la clasificación que en ella figura. De esta forma puede establecerse la doble finalidad que posee el análisis discriminante. Por una parte, explica la pertenencia de cada individuo de la muestra inicial a uno u otro grupo, en función de las variables de su perfil, para comprobar su pertenencia o no al grupo preestablecido, además de cuantificar el peso de cada una de estas variables en la discriminación. Y, por otra parte, el análisis predice a qué grupo va a pertenecer con una mayor probabilidad un nuevo individuo del que únicamente se conocen sus variables.

Existen dos enfoques para alcanzar este objetivo de clasificación: Un primer enfoque, basado en la obtención de funciones discriminantes de cálculo similar a las ecuaciones de regresión lineal múltiple; y un segundo enfoque, que emplea técnicas de correlación canónica y de componentes principales, denominado análisis discriminante canónico. El primero es el más habitual y 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 el que pertenece. Éste será el seguido en el presente Proyecto Fin de Carrera.

4.2 Descripción del Método

Para desarrollar el análisis se requieren una serie de hipótesis básicas:

1. *Debe de haber una variable categórica y las demás variables son de intervalo o de razón y son independientes respecto de ella.*

   La variable categórica del presente Proyecto Fin de Carrera es la Variable Calificación

2. *Se necesitan al menos dos grupos (en el caso a analizar son cuatro), y para cada grupo se necesitan dos o más casos (el análisis se realizará sobre66 casos).*

3. *El número de variables discriminantes debe ser menor que el número de objetos menos 2, es decir,* $(x_1, x_2, \dots, x_p)$ *donde p < (n-2) siendo n: número de objetos.*

   P= 8; n=66 (ocho variables y 66 casos quedarán tras los tres supuestos).

4. *Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.*

   Las variables son independientes entre sí.

5. *El número máximo de funciones discriminantes es el mínimo [número de variables, número de grupos menos uno] – con q grupos, (q-1) funciones discriminantes.*

   Se obtendrán tres funciones discriminantes (por los cuatro grupos de calificación).

6. *Las matrices de covarianzas dentro de cada grupo deben de ser aproximadamente iguales.*

   Para ello se realiza el contraste de hipótesis de homocedasticidad de varianzas.

7. *Las variables continuas deben seguir una distribución normal multivariante.*

   Para ello se contrasta la hipótesis de normalidad.

Los pasos a seguir son los siguientes:

1. Definición del objetivo del Análisis Discriminante (en el capítulo 2: Objeto del Proyecto Fin de Carrera).
2. Selección de variables, dependiente e independientes.
3. Cumplimiento de los requisitos previos.
4. Estimación del modelo discriminante.
5. Interpretación de los resultados obtenidos
6. Conclusiones

Paso 2: Selección de variables

En las aplicaciones del Análisis Discriminante se dispone frecuentemente de observaciones de un número relativamente elevado de variables que serán potencialmente discriminantes sobre la variable dependiente. La variable dependiente formará dos o más grupos excluyentes. Por otro lado, las variables independientes serán escogidas según una investigación previa que muestren su posible predicción sobre la variable dependiente o bien, aunque no se disponga de dicha investigación, se intuye que podrían discriminar.

Paso 3: Las hipótesis sobre el proceso de obtención de la muestra

Con los contrastes de significación que se realizan en el análisis discriminante con dos grupos se trata de dar respuesta a tres tipos de cuestiones diferentes:

(a) ¿Difieren significativamente las medias poblacionales de los dos grupos?

(b) ¿Se cumple la hipótesis de homocedasticidad del modelo?

(c) ¿Se cumplen las hipótesis de normalidad?

El análisis de normalidad se realizará en el siguiente capítulo.

(a) ¿Difieren significativamente las medias poblacionales de los dos grupos?

La respuesta a esta cuestión es crucial para la realización del análisis discriminante. En caso de que la respuesta fuera negativa no tendría ningún sentido continuar con el análisis, ya que significaría que las variables introducidas no tienen una capacidad discriminante significativa.

La hipótesis nula y alternativa para dar respuesta a la cuestión, en el caso de dos grupos:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

El contraste de la hipótesis se puede realizar específicamente mediante el estadístico T2 de Hotelling:

$$T^2 = (\bar{y}_1 - \bar{y}_2)\bar{S}^{-1}(\bar{y}_1 - \bar{y}_2)\left(\frac{n_1 \cdot n_2}{n_1 + n_2}\right)$$

Donde $\bar{S} = \left(\frac{V_1+V_2}{n_1+n_2-2}\right)$

La matriz S es un estimador insesgado de la matriz de covarianzas poblacional Σ, obtenido bajo el supuesto de que la matriz de covarianzas poblacional es la misma en los dos grupos.

Bajo la hipótesis nula, el estadístico T2 de Hotelling se distribuye:

$$\left(\frac{n_1 + n_2 - K - 1}{K}\right) \cdot \frac{T^2}{n_1 + n_2 - 2} \approx F_{K,n_1+n_2-K-1}$$

Existen otros estadísticos para realizar el contraste, diseñados para el caso general de G grupos, como el estadístico de Rao o el estadístico V de Barlett (estos dos últimos estadísticos están construidos a partir de la Λ de Wilks).

En el caso de que se rechace la hipótesis nula H_0: $\mu_1 = \mu_2$, se puede aplicar el análisis univariante de la varianza para contrastar la hipótesis de igualdad de medias para cada una de las variables clasificadoras por separado.

Como medida de evaluación de la bondad de ajuste se utiliza el coeficiente eta cuadrado(η^2)que es el coeficiente de determinación obtenido al realizar la regresión entre la variable dicotómica, que indica la pertenencia al grupo, y las puntuaciones discriminantes. A la raíz cuadrado de este coeficiente se le denomina correlación canónica.

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

(correlación canónica)
λ ≡ ratio que se obtiene al maximizar

$$Máx\lambda = \frac{w'_1 F w_1}{w'_1 V w_1} = \frac{separaciónentregrupos}{separacióndentrogrupos}$$

(b) ¿Se cumple la hipótesis de homocedasticidad del modelo?

Para el contraste de homocedasticidad (si la matriz de covarianzas es la misma para los distintos grupos) uno de los estadísticos que se pueden utilizar es el estadístico de Barlett-Box:

$$M = \frac{\prod_{g=1}^{K} \left|S_g\right|^{(n_g-1)/2}}{\left|\bar{S}\right|^{(n-k)/2}}$$

En el numerador, los determinantes de las estimaciones de la matriz de covarianzas para cada grupo.

En el denominador, el determinante de la estimación global de la matriz de covarianzas.

Cuando el numerador sea muy superior al denominador, será indicativo de que existe heteroscedasticidad, es decir, no existe homogeneidad entre las matrices de covarianzas de cada grupo.

donde: $S_g = \frac{V_g}{n_g - 1}$

$$\bar{S} = \frac{\sum_{g=1}^{G} V_g}{n-G} = \frac{\sum_{g=1}^{G} (n_g - 1) S_g}{n-G} K \equiv variables$$

La matriz S_g es una estimación de la matriz de covarianzas correspondiente a la celda g-ésima$\sum_g$, $\bar{S}$ es una estimación de la matriz de covarianzas global$\sum$.

Otra forma y además es la elegida en este Proyecto Fin de Carrera es la Prueba de Levene:

Esta prueba se utiliza para probar la hipótesis acerca de la igualdad de varianza de una variable, siendo la hipótesis nula aquella donde la variable exhibe igual varianza dada frente a la alternativa de que la variable no exhibe igual varianza. Para su cálculo se siguen los siguientes pasos:

Se presentas las hipótesis a continuación:

$$H_0: Noexistendiferenciassignificativasentrelasvarianzas$$

$$deloscuatroratingdelasCCAA$$

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

$$H_1: ExistendiferenciassignificativasentrelasvarianzasdeloscuatroratingdelasCCAA$$

$$H_1: \sigma_i^2 \neq \sigma_j^2 paraalmenos i \neq j; i = 1,2,3; j = 2,3,4$$

dondeσ_i^2es la varianza poblacional de la prueba de salida de la i-ésima calificación de la Comunidad Autónoma (i= 1, 2, 3, 4).

1. *Calcular la diferencia (en valor absoluto) entre cada valor y la media de su grupo:*

$$D_{ij} = \left|X_{ij} - \bar{X}_j\right|$$

donde: $X_{ij}: puntuación del sujeto i perteneciente al grupo j$
$\bar{X}_j: media del grupo j$

2. *Calcular la media de las diferencias de cada grupo:*

$$\bar{D}_j = \frac{\sum D_{ij}}{n_j}$$

donde: $\sum D_{ij}: suma de las puntuaciones D en el grupo j$

$n_j: tamaño del grupo j$

3. *Calcular la media total de las diferencias:*

$$\bar{D}_t = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} D_{ij}}{N}$$

donde: $\sum_{i=1}^{n} \sum_{j=1}^{k} D_{ij}$: $es la suma de las puntuaciones D de todos los sujetos$
$N: suma de todos los sujetos$

4. *Calcular la suma de cuadrados intragrupo (SCintra):*

$$SC_{intra} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\bar{D}_{ij} - \bar{D}_j\right)^2$$

5. *Calcular la suma de cuadrados intergrupo (SCinter):*

$$SC_{inter} = \sum_{j=1}^{k} n_j \left(\overline{D}_j - \overline{D}_t\right)^2$$

6. *Calcular los grados de libertad:*

$$G.L.(inter) = k - 1; siendo k el número de grupos$$

$$G.L.(intra) = \sum_{j=1}^{k} \left(n_j - 1\right); siendo n_j el tamaño muestral del grupo j$$

7. *Calcular la media cuadrática intergrupos:*

$$MC_{inter} = \frac{SC_{inter}}{G.L._{inter}}$$

8. *Calcular la media cuadrática intragrupos:*

$$MC_{intra} = \frac{SC_{intra}}{G.L._{intra}}$$

9. *Calcular la F = MCinter / MCintra*

$$F_{Prueba} = \frac{MC_{inter}}{MC_{intra}}$$

Paso 4: El Modelo Matemático: Análisis Discriminante

Partiendo de q grupos donde se asignan a una serie de objetos y de p variables medidas sobre ellos $(x_1, x_2, \dots, x_p)$, se trata de obtener para cada objeto una serie de puntuaciones que indican el grupo al que pertenecen $(y_1, y_2, \dots, y_m)$, de modo que sean funciones lineales de $(x_1, x_2, \dots, x_p)$:

$$\begin{cases} y_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p + w_{10} \\ y_m = w_{m1}x_1 + w_{m2}x_2 + \cdots + w_{mp}x_p + w_{10} \end{cases}$$

m= mín. [q-1,p]

tales que discriminen o separen lo máximo posible a los q grupos.

Estas combinaciones lineales de las p variables deben maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos.

Descomposición de la varianza

La variabilidad total de la muestra se puede descomponer en variabilidad dentro de los grupos y entre los grupos. Para ello, se parte:

$$Cov(x_j, x_{j'}) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ij'} - \overline{x_{j'}})$$

se puede considerar la media de la variable X_j en cada uno de los grupos $(I_1, I_2, \dots, I_q)$, es decir,

$$\bar{x}_{kj} = \frac{1}{n_k}\sum_{i \epsilon I_k} x_{ij}$$

para k=1, …, q.

De esta forma, la media total de la variable X_j se puede expresar como función de las medias dentro de cada grupo: $n_k\bar{x}_{kj} = \sum_{i\epsilon I_k} x_{ij}$

con lo cual,

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \frac{1}{n}\sum_{k=1}^{q}\sum_{i\epsilon I_k} x_{ij} = \frac{1}{n}\sum_{k=1}^{q} n_k\bar{x}_{kj} = \sum_{k=1}^{q}\frac{n_k}{n}\bar{x}_{kj}$$

Poniendo en cada uno de los términos: $\begin{cases} (x_{ij} - \bar{x}_j) = (x_{ij} - \bar{x}_{kj}) - (\bar{x}_{kj} - \bar{x}_j) \\ (x_{ij'} - \bar{x}_{j'}) = (x_{ij'} - \bar{x}_{kj'}) - (\bar{x}_{kj'} - \bar{x}_{j'}) \end{cases}$

se obtiene,

$$Cov(x_j, x_{j'}) = \frac{1}{n}\sum_{k=1}^{q}\sum_{i \epsilon I_k}(x_{ij} - \bar{x}_j) - (\bar{x}_{ij'} - \bar{x}_{j'}) =$$

$$= \frac{1}{n}\sum_{k=1}^{q}\sum_{i \epsilon I_k}(x_{ij} - \bar{x}_{kj}) - (x_{ij'} - \bar{x}_{kj'}) + \sum_{k=1}^{q}\frac{n_k}{n}(\bar{x}_{kj} - \bar{x}_j)(\bar{x}_{ij'} - \bar{x}_{j'}) =$$

$$= v(x_j, x_{j'}) + f(x_j, x_{j'}) \rightarrow t(x_j, x_{j'}) = v(x_j, x_{j'}) + f(x_j, x_{j'}) \rightarrow$$

$$\rightarrow T = V + F$$

La covarianza total es igual a la covarianza dentro de los grupos más la covarianza entre grupos.

Extracción de las Funciones Discriminantes

La idea básica del Análisis Discriminante consiste en extraer a partir de $(x_1, x_2, \dots, x_p)$ variables observadas en k grupos, m funciones $(y_1, y_2, \dots, y_p)$de forma que:

$$y_i = w_{i1}x_1 + w_{i2}x_2 + \cdots + w_{ip}x_p + w_{i0} dondem = mín(q-1, p)$$

$$talesquecorre(y_i, y_j) = 0 \; \forall i \neq 0$$

Si las variables $(x_1, x_2, \dots, x_p)$están tipificadas, las funciones $(y_i = w_{i1}x_1 + + w_{i2}x_2 + \cdots + w_{ip}x_p)$ para (i=1, …, m) se denominan discriminantes canónicas.

Las funciones $(y_1, y_2, \dots, y_p)$ se extraen de modo que:

- Y_1 sea la combinación lineal de $(x_1, x_2, \dots, x_p)$que proporciona la mayor discriminación posible entre los grupos.
- Y_2sea la combinación lineal de $(x_1, x_2, \dots, x_p)$ que proporciona la mayor discriminación posible entre los grupos, trasY_1 , tal que corre $(y_1, y_2,) = 0$

En general, i y es la combinación lineal de $(x_1, x_2, \dots, x_p)$ que proporciona la mayor discriminación posible entre los grupos, después de y_{i-1} , tal que corre $(y_1, y_2,) = 0$ para J=1, …, (i-1).

Matricialmente

Se busca una función lineal de $(x_1, x_2, \dots, x_p)$: Y = w'X

Es conocido que la covarianza total es igual a la covarianza dentro de los grupos más la covarianza entre grupos:

T = **F** + **V** (matricialmente)

De modo que:

Var(y) = w'Tw = w'Fw+w'Vw

Se maximiza la variabilidad entre los grupos para discriminarlos mejor, es decir, se maximiza la varianza entre grupos en relación con el total de la varianza:

$$máx\left[\frac{w'Fw}{w'Tw}\right]$$

Considerando la función:

$$f(w)=\frac{w'Fw}{w'Tw}$$

se observa que es una función homogénea, es decir,

$$f(w)=f(\mu w)\forall\mu\in R$$

El hecho de que sea homogénea implica que calcular:

$$máx\left[\frac{w'Fw}{w'Tw}\right]$$

Equivale a calcular:

$$Máx[w'Fw]\text{tal que }[w'Fw]=1$$

Como es el esquema habitual de los multiplicadores de Lagrange, se define:

$$L=w'Fw-\lambda(\mathrm{w'T\,w}-1)\rightarrow\frac{\vartheta\mathrm{L}}{\vartheta\mathrm{w}}=2\mathrm{F\,w}-2\lambda\mathrm{T\,w}=0\rightarrow\mathrm{F\,w}=\lambda\mathrm{T\,w}\rightarrow(\mathrm{T}^{-1}\mathrm{F})\mathrm{w}=\lambda\mathrm{w}$$

En consecuencia, el autovector asociado a la primera función discriminante lo es de la matriz $(T^{-1}-F)$, que en general no es simétrica.

Como:

$$Fw=\lambda\cdot Tw,\text{ se tiene }w'Fw=\lambda w'Tw$$

Pues, tomando el vector asociado al máximo autovalor se obtendrá la función que recoge el máximo poder discriminante.

El autovalor asociado a la función discriminante indica la proporción de varianza total explicada por las m funciones discriminantes que recoge la variable y_i.

Para obtener más funciones discriminantes se siguen sacando los autovectores de la matriz $(T^{-1}-F)$ asociados a los autovalores elegidos en orden decreciente:

$$\begin{cases} w'_2\rightarrow w'_2X=Y_2\\ \dots\dots\dots\dots\dots\dots\dots\\ w'_m\rightarrow w'_mX=Y_m\end{cases}$$

Siendo $m=mín(q-1,p)$

Estos vectores son linealmente independientes y dan lugar a funciones sin correlación entre sí.

La suma de todos los autovalores:

$$\sum_{i=1}^{m}\lambda_i$$

es la proporción de varianza total que queda explicada, o se conserva, al considerar sólo los ejes o funciones discriminantes.

Como consecuencia, el porcentaje explicado por la variable i y del total de varianza explicada por las funciones $(y_1, y_2, \ldots, y_p)$ es:

$$\frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} \cdot 100\%$$

Clasificación en dos grupos

En primer lugar, se desarrolla el método para el caso de dos grupos. Posteriormente se hará lo propio con el caso de más de dos grupos, que es el aplicado en este Proyecto Fin de Carrera.

Se estudia la aplicación Método del Análisis Discriminante a la clasificación de individuos en el caso de que se puedan asignar solamente a dos grupos a partir de k variables discriminadoras.

Fisher resuelve el problema mediante su función discriminante:

$$D = w_1X_1 + w_2X_2 + \cdots + w_kX_k$$

Las puntuaciones discriminantes son los valores que se obtienen al dar valores a $(X_1, X_2, \dots, X_k)$en la ecuación anterior.

Se trata de obtener los coeficientes de ponderación w_j.

Si se consideran N observaciones → La función discriminante X para:

$$D = w_1X_{1i} + w_2X_{2i} + \cdots + w_kX_{ki}i = 1, \dots, N.$$

(D_i)es la puntuación discriminante correspondiente a la observación i-ésima.

La función discriminante en forma matricial:

$$\begin{pmatrix} D_1 \\ D_2 \\ \dots \\ D_N \end{pmatrix} = \begin{pmatrix} X_{11}X_{11} \dots X_{k1} \\ X_{12}X_{22} \dots X_{k2} \\ \dots \\ X_{1N}X_{2N} \dots X_{kN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{pmatrix}$$

Expresando el modelo en función de las desviaciones a la media, resulta:

$$\begin{pmatrix} D_1 - \overline{d_1} \\ D_2 - \overline{d_2} \\ \dots \\ D_N - \overline{d_N} \end{pmatrix} = \begin{pmatrix} X_{11}X_{11} \dots X_{k1} \\ X_{12}X_{22} \dots X_{k2} \\ \dots \\ X_{1N}X_{2N} \dots X_{kN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{pmatrix}$$

es decir, $d = Xw$ (función discriminante en diferencias)

La variabilidad de la función discriminante (suma de cuadrados de las desviaciones de las variables discriminantes con respecto a su media) se expresa:

Suma de cuadrados explicada por esta función: $dd' = w'X'Xw$

$X'X$es una matriz simétrica que expresa las desviaciones cuadráticas con respecto a la media de las variables (suma de cuadrados total).

Se puede descomponer en suma de cuadrados entre grupos F y suma de cuadrados dentro de los grupos V:

$$T = X'X$$

(matriz de suma de cuadrados y productos cruzados (varianzas-covarianzas) para el conjunto de observaciones).

$$T = X'X = F + V$$

con lo cual:

$$dd' = w'X'Xw = w'(F + V)w = w'Fw + w'Vw$$

Los ejes discriminantes vienen dados por los vectores propios asociados a los valores propios de la matriz $(V^{-1}F)$ ordenados de mayor a menor.

Las puntuaciones discriminantes se corresponden con los valores obtenidos al proyectar cada punto del espacio k-dimensional de las variables originales sobre el eje discriminante.

Los coeficientes w se obtienen de la siguiente forma:

$$Máx\,\lambda = \frac{w'F\,w}{w'V\,w} = \frac{\text{separación entre grupos}}{\text{separación dentro grupos}}$$

Como puede observarse, se trata de que el primer término (entre-grupos) sea lo mayor posible en detrimento del segundo término (intra-grupos).

La función discriminante de Fisher suele ir acompañada del calificativo lineal debido a que se obtiene como combinación lineal de las variables originales.

Los coeficientes $(w_1, w_2, \dots, w_k)$ (normalizados) que se obtienen en el proceso de maximización de λ puede contemplarse como un conjunto de cosenos que definen la situación del eje discriminante. Para esta interpretación, la normalización se refiere a que la suma de sus cuadrados sea la unidad.

Las puntuaciones discriminantes son aquellas que se obtienen al dar valores a $X_1, X_2, \dots, X_K$ en la función discriminante de Fisher, y se corresponden con los valores obtenidos al proyectar cada punto del espacio K-dimensional de las variables originales sobre el eje discriminante.

Clasificación de los casos

Se obtienen las puntuaciones discriminantes d_i para cada observación, introduciendo los correspondientes valores de las k variables en la función discriminante.

Se aplica el criterio de clasificación:

$$d_i < C\ (d_i - C < 0) \rightarrow pertenece\ al\ grupo\ 1º$$

$$d_i > C\ (d_i - C > 0) \rightarrow pertenece\ al\ grupo\ 2º$$

O bien:

Funciones Discriminantes para cada grupo → se clasifica la observación en el grupo en que la función correspondiente arroja mayor valor.

Centroides para los grupos I y II

Los centros de gravedad o centroides, es decir, el vector de medias son aquellos estadísticos básicos que resumen la información sobre los grupos de la variable dependiente. Las denominaciones que se utilizarán para designar a los centroides de los grupos I y II son las siguientes:

$$\bar{x}_I = \begin{pmatrix} \bar{X}_{1I} \\ \bar{X}_{2I} \\ \cdots \\ \bar{X}_{kI} \end{pmatrix}$$

$$\bar{x}_{II} = \begin{pmatrix} \bar{X}_{1II} \\ \bar{X}_{2II} \\ \cdots \\ \bar{X}_{kII} \end{pmatrix}$$

Los subíndices I y II indican a qué grupo pertenece la variable.

Puntuaciones para cada grupo

Si se sustituye en la función discriminante de Fisher por los elementos de los vectores mostrados anteriormente se obtiene:

$$\begin{cases} \bar{D}_I = w_1\bar{X}_{1I} + w_2\bar{X}_{2I} + \cdots + w_k\bar{X}_{kI} \\ \bar{D}_{II} = w_1\bar{X}_{1II} + w_2\bar{X}_{2II} + \cdots + w_k\bar{X}_{kII} \end{cases}$$

De esta manera se determinan las puntuaciones de las dos funciones discriminantes, utilizando los valores de w y X.

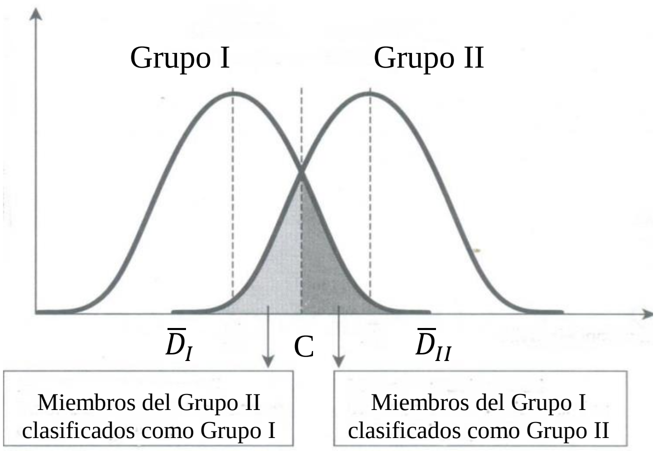
Criterio para clasificar a un individuo

Si D < C se clasifica al individuo i en el grupo I

Si D > C se clasifica al individuo i en el grupo II

C: punto de corte discriminante $C = \frac{\bar{D}_I + \bar{D}_{II}}{2}$

Figura 2:
Funciones de distribución de frecuencias hipotéticas de dos grupos



En la figura 2 se ha representado un ejemplo de funciones de frecuencia que corresponden a dos grupos, que es el caso que se está analizando. La varianza y distribución de la frecuencia es la misma en las dos funciones, siendo diferente tan sólo en sus medias. Se observa que ambas funciones no se solapan, ya que si ocurriera precisamente esto el análisis no tendría sentido realizarlo ya que no habría distinción sobre dónde pertenecería un individuo.

En general:

$$\langle D - C = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k - C \rangle$$

se clasifica dependiendo si (D−C) es positivo o negativo (grupo I o grupo II).

Utilizando la ecuación anterior, si se iguala a cero el segundo miembro, en el caso de dos variables se obtiene la ecuación de la recta:

$$w_1 X_1 + w_2 X_2 - C = 0$$

Que delimita en el plano (X_1, X_2) a los grupos I y II.

Además, como ya se puntualizó en el apartado “Clasificación de los casos”, existe una forma alternativa a la utilización de:

$$\langle D - C = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k - C \rangle$$

que consiste en construir funciones discriminantes para cada grupo, basadas también en el maximización de λ. Estas funciones tienen la siguiente estructura:

$$\begin{cases} F_I = w_1 X_{1I} + w_2 X_{2I} + \cdots + w_k X_{kI} \\ F_{II} = w_1 X_{1II} + w_2 X_{2II} + \cdots + w_k X_{kII} \end{cases}$$

Cuando se utilizan estas funciones, se clasifica a un individuo en el grupo para el que la función F_j sea mayor. Este tipo de funciones clasificadoras tiene la ventaja de que se generalizan fácilmente al caso de que existan más de 2 grupos. Además, en buena parte de los programas de análisis multivariante se suministran estas funciones. A partir de los coeficientes de las dos funciones anteriores se pueden obtener los coeficientes de la función

$$\langle D - C = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k - C \rangle$$

Clasificación con más de dos grupos

En el caso de Análisis Discriminante que existan más de dos grupos, es decir, G grupos (G > 2) es el llamado Análisis Discriminante Múltiple que será el utilizado en el presente Proyecto Fin de Carrera. En este tipo de análisis, el número máximo de ejes discriminantes que se pueden obtener viene dado por mín. (G-1, k), siendo G el número de grupos y k el número de variables explicativas. En el caso del presente Proyecto, G será igual a 4 y k han sido 20 en primer lugar, posteriormente se han irán eliminando variables por no cumplir con algunos de los tres requisitos expuestos anteriormente (normalidad, homocedasticidad o independencia). Finalmente, el análisis se hará con 8 variables. El número de ejes discriminantes que se obtendrán será de 3. Normalmente, el número de grupos suele ser menor que el de variables explicativas, como en este caso.

Cada una de las funciones discriminantes D_i se obtiene como función lineal de las k variables explicativas:

$$D_i = w_{i1}X_1 + w_{i2}X_2 + \cdots + w_{ik}X_k$$
$$para\ i = 1, \dots, G-1$$

Las funciones discriminantes de este Proyecto Fin de Carrera serían tres:

$$D_1 = w_{11} \cdot X_1 + w_{12} \cdot X_2 + w_{13} \cdot X_3$$

$$D_2 = w_{21} \cdot X_1 + w_{22} \cdot X_2 + w_{23} \cdot X_3$$

$$D_3 = w_{31} \cdot X_1 + w_{32} \cdot X_2 + w_{33} \cdot X_3$$

$$i = 1,2,3$$

Los (G−1) ejes vienen definidos respectivamente por los vectores $(w_1, w_2, \dots, w_{G-1})$, siendo éstos:

$$w_1 = \begin{pmatrix} w_{11} \\ w_{12} \\ \dots \\ w_{1k} \end{pmatrix}; w_2 = \begin{pmatrix} w_{21} \\ w_{22} \\ \dots \\ w_{2k} \end{pmatrix}$$

En general:

$$w_{G-1} = \begin{pmatrix} w_{G-1,1} \\ w_{G-1,2} \\ \dots \\ w_{G-1k} \end{pmatrix}$$

Para la obtención del primer eje discriminante se maximiza la ratio variabilidad entre grupos entre variabilidad dentro grupos, es decir: separación dentro grupos separación entre grupos.

$$Máx\lambda_1 = \frac{w_1\ F\ w_1}{w_1\ V\ w_1} = \frac{separación\ entre\ grupos}{separación\ dentro\ grupos}$$

(*criterio de obtención del primer eje discriminante*)

Derivando la ratio e igualando a cero:

$$\frac{\vartheta\lambda_1}{\vartheta w_1} = 0$$

con lo cual:

$$\frac{\vartheta\lambda_1}{\vartheta w_1} = \frac{2Fw_1(w'Vw_1) - 2Vw_1(w'Fw_1)}{(w'Vw_1)^2} = 0 \rightarrow$$

$$\rightarrow 2\ F\ w_1(w'Vw_1) - 2\ V\ w_1(w'Vw_1) = 0$$

operando con la expresión, resulta:

$$\frac{2\ F\ w_1}{2\ F\ w_1} = \frac{(w'F\ w_1)}{(w'V\ w_1)} = \lambda_1 \rightarrow F\ w_1 = V\ w_1\lambda_1$$

siendo, por tanto,

$$\lambda_1 w_1 = V^{-1}\ F\ w_1$$

La obtención del vector 1 w resulta un problema de cálculo de un vector característico asociado a la matriz no simétrica $(V^{-1}F)$. De las raíces características que se obtienen al resolver la ecuación:

$$\lambda_1 w_1 = V^{-1}\ F\ w_1$$

se retiene la mayor, ya que $\lambda_1$es la ratio que se pretende maximizar y $w_1$es el vector característico asociado a dicha raíz característica.

Como $\lambda_1$es la ratio $\left[\frac{w_1\ F\ w_1}{w_1\ V\ w_1}\right]$medirá el poder discriminante del primer eje discriminante. El resto de los ejes discriminantes son otros vectores característicos de la matriz $(V^{-1}F)$, ordenados según el orden decreciente de las raíces características.

Así, el segundo eje discriminante tendrá menor poder discriminante que el primero, pero más que cualquiera de los restantes.

Puesto que la matriz $(V^{-1}F)$ no es simétrica, en general, esto implicará que los ejes discriminantes no sean ortogonales, es decir, no sean perpendiculares entre sí.

Contrastes de significación

En el análisis discriminante múltiple se plantean contrastes específicos para determinar si cada uno de los valores λ_i es estadísticamente significativo, es decir, para determinar si cada uno de los valores $\lambda_i$contribuye o no a la discriminación entre los diferentes grupos.

Este tipo de contrastes se realiza a partir del estadístico V de Barlett. El estadístico V es una función de la Δde Wilks y se aproxima a una chi-cuadrado, tiene interés en el análisis discriminante por su descomponibilidad.

Estadístico V de Barlett:

$$V = \left[n - 1 - \frac{K+G}{2}\right] (\ln \Delta) V \approx \mathcal{X}^2_{K(G-1)} \qquad \begin{cases} K \equiv variables\ categóricas \\ G \equiv grupos \end{cases}$$

Este estadístico se utiliza en el análisis multivariante para contrastar las hipótesis

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_G \\ H_1: No\ todas\ las\ \mu_g\ son\ iguales \end{cases}$$

En el análisis multivariante de la varianza con un factor se contrasta esta hipótesis para determinar si el factor (variable categórica con G grupos) explica la variabilidad del vector de variables dependientes de forma significativa.

En el análisis discriminante múltiple la hipótesis a contrastar sigue siendo la misma, aunque los papeles se han invertido. Ahora se realiza el contraste para tratar de dar respuesta a la siguiente pregunta:

¿Las K variables clasificadoras contribuyen significativamente a discriminar entre los G grupos?

Si no se rechaza la hipótesis nula citada, no se debería continuar el análisis, puesto que las variables clasificadoras utilizadas en la investigación no tienen ningún poder discriminante significativo.

Para examinar el poder discriminante de cada uno de los ejes que se construyen en el análisis discriminante, se descompone el estadístico V en productos a partir de la descomposición de la Δde Wilks. De acuerdo con su definición, el recíproco de Λ se puede descomponer:

$$\frac{1}{\Delta} = \frac{|T|}{|V|} = |V|^{-1}|T| = |V^{-1}||T| = |V^{-1}T| = |V^{-1}(F+V)| = |I + V^{-1}F|$$

teniendo en cuenta que el determinante de una matriz es igual al producto de sus raíces características, se obtiene que:

$$\frac{1}{\Delta} = |I + V^{-1}F| = (1+\lambda_1)(1+\lambda_2) \dots (1+\lambda_{G-1})$$

sustituyendo en el estadístico V de Barlett, se obtiene la expresión alternativa del estadístico:

Estadístico V de Barlett:

$$V = \left[n - 1 - \frac{K+G}{2}\right] \sum_{g=1}^{G-1} ln\left(1 + \lambda_g\right)$$

Si se rechaza la hipótesis nula, significa que al menor uno de los ejes discriminantes es estadísticamente significativo. Esto implica a su vez que el primer eje discriminante es estadísticamente significativo, debido a que es precisamente el que tiene mayor poder discriminante.

En caso de que se acepte la hipótesis de que el primer eje discriminante es significativo, se pasa a contrastar la significación conjunta del resto de los ejes discriminantes, utilizando el estadístico:

$$V = \left[n - 1 - \frac{K+G}{2}\right] \sum_{g=2}^{G-1} ln\left(1 + \lambda_g\right)$$

De forma general, se puede establecer la expresión de contrastación secuencial mediante el estadístico:

Estadístico V de Barlett:

$$V_j = \left[n - 1 - \frac{K+G}{2}\right] \sum_{g=j+1}^{G-1} ln\left(1 + \lambda_g\right) \text{donde j=0, 1, 2, …, G-2}$$

Así, en el proceso secuencial se van eliminando del estadístico V las raíces características que van resultando significativas, deteniendo el proceso cuando se acepte la hipótesis nula de no significatividad de los ejes discriminantes que queden por contrastar.

Método Paso a Paso (Stepwise)

El método Paso a Paso (también nombrado a través de su denominación en inglés, Stepwise) es uno de los algoritmos para seleccionar las variables útiles que pueden considerarse, desde el punto de vista de la selección, hacia adelante o hacia atrás.

La aplicación de procedimientos iterativos requiere definir previamente una regla de decisión para medir la bondad del ajuste en cada paso o iteración. Una regla es la de minimización del estadístico Λ de Wilks, ya que cuanto menor sea este estadístico, mayor será el grado del ajuste. En cada paso se selecciona la variable para la que se obtenga un menor estadístico Λ. Otra regla de decisión consiste en minimizar la distancia de Mahalanobis entre los dos centroides.

En el Método de selección paso a paso hacia adelante (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, de modo que encontremos la pareja que produce la mayor discriminación. La variable que contribuye a la mejor pareja es seleccionada en segundo lugar. Con ambas variables, podrán formarse grupo de tres variables para determinar cuál de éstas resulta más discriminante. De este modo quedaría seleccionada esa la tercera variable. El proceso continuaría hasta que todas las variables hayan sido seleccionadas o las variables restantes no supongan un suficiente incremento en la capacidad de discriminación. Por último, en el Método de selección paso a paso hacia atrás (backward), todas las variables son consideradas inicialmente, y van siendo excluidas una a una en cada etapa, eliminando del modelo aquellas cuya supresión produce el menor descenso en la discriminación entre los grupos.

En el proceso Stepwise, en cada paso puede entrar, y también salir, una variable en el conjunto seleccionado, dependiendo del valor que tenga el estadístico F correspondiente a la lambda de Wilks o, en general, al estadístico que se utilice como criterio. Cuanto mayor sea el valor de la F, más significativa será la variable para la que se calcula. Antes de comenzar la aplicación del procedimiento es necesario fijar un valor F mínimo para entrar y un valor de la F máximo para salir. Este valor vendrá dado por IBM SPSS Statistics.

El valor de F mínimo debe ser mayor que el valor de F máximo. De lo contrario, una variable podría estar entrando y saliendo de forma indefinida en la selección. Cuando se está aplicando el procedimiento Paso a Paso, los niveles de significación con los que se está trabajando en cada paso no son los verdaderos. Para el cálculo de los niveles de significación verdaderos se requeriría tener en cuenta todas las contrastaciones que se han realizado en el proceso.

En la aplicación de este procedimiento, también se suele fijar un nivel de tolerancia, que es una medida del grado de asociación lineal entre las variables clasificadoras. Para la variable i la tolerancia se define igual a $1 - r_i^2$, donde r_i^2 es el coeficiente de determinación entre la variable i y el resto de variables explicativas que figuran en el modelo, es decir, que están seleccionadas en ese momento. Cuando la tolerancia de la variable i es muy pequeña supone que dicha variable está muy correlacionada con el resto de las variables explicativas, lo que puede provocar problemas en la estimación. Para solucionar esta posible dificultad, generalmente se suele fijar un nivel mínimo de tolerancia del 0,001 con lo que las variables con una tolerancia menor a ese límite son excluidas del análisis.

Al iniciar el procedimiento todas las variables explicativas están fuera del modelo. En la primera iteración se introduce en el modelo aquella variable con menor Λ de Wilks, si éste es el criterio que se está empleando, siempre que el valor correspondiente de la F sea mayor que la F mínima. El estadístico F inicial de cada variable, correspondiente a la lambda de Wilks, se obtiene aplicando el análisis de la varianza a cada una de las variables clasificadoras por separado.

Para cada una de las variables que están fuera (en la primera iteración, todas menos una) se calcula el valor de la F correspondiente al cambio que produce la introducción de la variable en la lambda de Wilks del modelo. En la primera iteración, la lambda del modelo es la lambda de la variable que ya se ha introducido. Al valor calculado de esta forma se le denomina F-para-entrar. Se introduce en el modelo aquella variable con la F-para-entrar más alta, siempre que su valor sea mayor que la F mínima, como se ha comentado anteriormente.

Si se cumple la condición anterior, el modelo estará integrado por esas dos variables. Cuando hay una exclusión de una variable, previamente se ha calculado su F que se denomina F-para-salir. Se elimina del modelo aquella variable cuya F-para-salir sea menor que la F máxima para entrar.

Este proceso continúa hasta que, entre las variables fuera del modelo, no exista ninguna variable a la que corresponda una F-para-entrar mayor que la F mínima para entrar. IBM SPSS Statistics suele fijar (según quiera el usuario) un número máximo de iteraciones deteniéndose el proceso en ese límite en el caso de que no se haya hecho antes.

Probabilidades en el Análisis Discriminante: Análisis Discriminante Bayesiano

Fisher obtuvo la función discriminante aplicando un enfoque descriptivo. Cuando en el análisis discriminante se desean abordar cuestiones de carácter inferencial y otros relativos al modelo poblacional se requiere la formulación previa de hipótesis estadísticas. Esta perspectiva es debido al Teorema de Bayes.

Las cuestiones de tipo inferencial se refieren a diversos contrastes de significación sobre el modelo, así como otros que se usan durante en el proceso de selección de variables cuando el número de variables es elevado y además, no se conoce de antemano si dichas variables vayan a ser relevantes o no en el análisis.

Por otra parte, el cálculo de probabilidad de pertenencia a un grupo requiere que previamente se haya postulado algún modelo probabilístico de la población. Las hipótesis estadísticas que se acogen son semejantes a las postuladas en el análisis multivariante de la varianza y se refieren a la población, así como al proceso de obtención de la muestra.

Las funciones discriminantes vistas anteriormente clasifican a los diferentes individuos en uno u otro grupo, pero no dan más información sobre éstos.

En ocasiones es conveniente tener información adicional a las puntuaciones discriminantes. Con estas puntuaciones se puede clasificar a cada individuo, pero sería interesante disponer además de información sobre la probabilidad de su pertenencia a cada grupo, ya que ello permitiría realizar análisis más matizados e incluir otras informaciones como información a priori o posibles costes que impliquen una clasificación errónea (en el caso que fuera necesario).

La clasificación de los individuos se realizará utilizando el teorema de Bayes. La aplicación de este teorema permite el cálculo de las probabilidades a posteriori a partir de las probabilidades a priori y de la información muestral contenida en las puntuaciones discriminantes. Si se considera 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 [Prob(g/D)] es la siguiente:

Teorema de Bayes

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

En el segundo miembro aparecen las probabilidades a priori π_g y las probabilidades condicionadas $Prob\left({}^{D}/_{g}\right)$.

La probabilidad condicionada $Prob\left({}^{D}/_{g}\right)$ se obtiene calculando la probabilidad de la puntuación observada suponiendo la pertenencia a un grupo g.

Dado que el denominador del segundo miembro es una constante, se utiliza también, de manera equivalente, la siguiente expresión:

$$Prob\left({}^{g}/_{D}\right) \propto \left(\pi_g + Prob\left({}^{D}/_{g}\right)\right)$$

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 su probabilidad a posteriori. Se va a presentar el cálculo de probabilidades de forma que sea fácilmente generalizable al caso de G grupos (se analiza el caso de dos grupos)

El cálculo de probabilidades se realizará en función del cálculo de probabilidades sin información a priori y del cálculo de probabilidades con información a priori. En el presente Proyecto Fin de Carrera se han utilizado ambas, con información y sin información previa.

En primer lugar, se afronta el cálculo de probabilidades sin información a priori.

Probabilidad sin información a priori

El cálculo de esta probabilidad de pertenecer a diferentes grupos permite introducir matices en la información acerca de cada individuo.

En el cálculo de probabilidades siguientes se considera que no existe conocimiento previo de las probabilidades de pertenencia a cada grupo. Cuando no se dispone de dicha información, se adopta el supuesto de que la probabilidad de pertenencia a los dos grupos es la misma, es decir, $\pi_I = \pi_{II}$.Ésto implica que estas probabilidades a priori no afectan a los cálculos de las probabilidades a posteriori.

La probabilidad de pertenencia a cada grupo, dada la puntuación discriminante obtenida, viene según la siguiente expresión:

$$Prob\left({}^{g}/_{D}\right) = \frac{e^{F_g}}{e^{F_I} + e^{F_{II}}}$$

Para g= I, II y F_I y F_{II} son las funciones que se habían definido anteriormente.

Un individuo se clasifica en el grupo en la que la probabilidad sea mayor. Este criterio implica que un individuo se clasificará en el grupo I si:

$$F_I > F_{II}$$

Con la ecuación anterior se consiguen los mismos resultados que aplicando la función discriminante de Fisher.

Otro método de clasificar consiste en minimizar la probabilidad de clasificación errónea. Si se denomina Prob(I/II), la probabilidad de clasificar a un individuo en el grupo I perteneciendo éste al grupo II y Prob(II/I) que es la probabilidad de clasificar a un individuo en el grupo II perteneciendo al I.

La probabilidad total de clasificación errónea es:

$$Prob(II/I) + Prob(I/II)$$

Si se minimiza esta probabilidad se obtiene también como punto de corte el valor de C.

Probabilidad con información a priori

A veces ocurre se dispone de información de la probabilidad a priori sobre pertenencia a un individuo a cada uno de los grupos. Así, en el caso de concesión de préstamos, por ejemplo, se puede tener la información de que los préstamos fallidos suponen un 10% del total de préstamos concedidos a lo largo de los 5 años. Por ello es importante conocer esta probabilidad.

Cuando se utilizan probabilidades a priori, los casos se clasifican en el grupo para el que la probabilidad a posteriori sea mayor. La probabilidad a posteriori de pertenencia a cada grupo es:

$$Prob\left({}^{g}/_{D}\right) = \frac{\pi_I e^{F_g}}{\pi_I e^{F_I} + \pi_{II} e^{F_{II}}}$$

Para g= I, II.

Con este criterio se clasifica a un individuo en el grupo I si:

$$F_I \ln(\pi_I) > F_{II} \ln(\pi_{II})$$

El punto de corte con información a priori viene definido por la siguiente expresión:

$$C_P = \frac{\overline{D}_I + \overline{D}_{II}}{2} - ln\frac{\pi_{II}}{\pi_I}$$

El ratio de probabilidades a priori debe establecerse de forma que el punto de corte se desplace hacia el grupo con menor probabilidad a priori. Al desplazar el punto de corte de esta forma, se tenderá a clasificar una proporción menor de individuos en el grupo con menor probabilidad a priori.

Cuando las dos probabilidades a priori son igual a 0.5, entonces la expresión anterior se convierte en el valor de C:

$$C = \frac{\overline{D}_I + \overline{D}_{II}}{2}$$

Si se introducen probabilidades a priori, la probabilidad total de clasificación errónea es:

$$\pi_I \cdot Prob(II/I) + \pi_{II} \cdot Prob(I/II)$$

Como se puede observar, cada probabilidad de clasificación errónea va multiplicada por la probabilidad a priori del grupo real de pertenencia.

Los pasos 5 y 6 (interpretación de los resultados y conclusiones) se realizarán en los capítulos 6 y 7 del presente Proyecto Fin de Carrera, respectivamente.

5. Análisis previo de los datos

Diagnóstico de las hipótesis básicas del modelo en el Análisis Multivariante:

Cada técnica multivariante está basada en una serie de hipótesis de naturaleza estadística. Algunas de ellas son muy exigentes, como el Análisis Discriminante Múltiple (que será el utilizado en este Proyecto Fin de Carrera) ya que asume la normalidad de los datos, la homocedasticidad y la independencia de las observaciones.

A continuación, se muestran las variables independientes, el análisis de la variable dependiente y, por último, se comprueban las tres hipótesis básicas del modelo Análisis Discriminante.

5.1 Recogida de datos

Variables

A continuación, se detalla las variables económico-financieras seleccionadas para la discriminación. Para el análisis se han recogido un total de 20 variables.

Se ha creado una base de datos con toda la información. Esta información se ha obtenido del *Ministerio de Hacienda y Administraciones Públicas*, *Ministerio de Economía y Competitividad*, *Ministerio de Política Territorial y Administraciones Públicas* y del *Diario Expansión del grupo Unidad Editorial* (líder en prensa económica en España), principalmente. En la bibliografía se encuentra de manera más detallada esta información además de alguna otra fuente que se ha utilizado, como, por ejemplo, para el desarrollo del método estadístico.

Los datos de las variables recogidos en la base de datos son de todas las CCAA de España con la excepción de Ceuta y Melilla, entre los años 2009 y 2014.

Las variables utilizadas en este proyecto son:

Variable 1: Superávit o (Déficit / Ingresos Corrientes)

Variable 2: Tasa Anual Crecimiento de Ingresos Corrientes

Variable 3: Tasa Anual Crecimiento de Gastos Corrientes

Variable 4: Deuda / PIB

Variable 5: (Deuda – Ingresos Corrientes) / PIB

Variable 6: PIB / Habitante año respecto a la media nacional

Variable 7: Deuda / Ingresos Corrientes

Variable 8: Carga Financiera / Ingresos Corrientes

Variable 9: Gastos Financieros / Ingresos Totales

Variable 10: Ingresos Corriente / Ingresos Totales

Variable 11: Ingresos por Impuestos / Ingresos Corrientes

Variable 12: Ingresos por Impuestos Totales / Ingresos Totales

Variable 13: Gasto Per Cápita

Variable 14: Ingresos Corrientes / Ingresos Totales

Variable 15: Ingresos Capítulos (6, 7 y 8) / Ingresos Totales

Variable 16: Nº Matriculaciones / Nº Habitantes

Variable 17: Importaciones % PIB

Variable 18: Exportaciones % PIB

Variable 19: Esperanza de vida

Variable 20: Balanza Comercial Exportaciones – Importaciones

Seguidamente se explican los diferentes términos que han sido utilizados en las variables:

- *Déficit:* El Déficit en Economía y Contabilidad es el saldo negativo que se produce cuando los egresos son mayores a los ingresos. En contabilidad representa el exceso de pasivo sobre activo. Cuando se refiere al déficit público se habla del exceso de gasto gubernamental sobre sus ingresos; cuando se trata de déficit comercial de la balanza de pagos se relaciona con el exceso de importaciones sobre exportaciones.

- *Ingresos Corrientes:* Constituyen los ingresos que se recaudan normalmente por impuestos, contribuciones, regalías, tasas y otros ingresos no tributarios, venta de bienes y servicios, operaciones financieras, transferencias y donaciones corrientes y otros conceptos similares.

- *Deuda:* Por deuda pública o deuda soberana se entiende al conjunto de deudas que mantiene un Estado frente a los particulares u otro país. Constituye una forma de obtener recursos financieros por el Estado o cualquier poder público y se consigue normalmente mediante emisiones de títulos de valores o bonos.

- *PIB:* es una magnitud macroeconómica que expresa el valor monetario de la producción de bienes y servicios de demanda final de un país (o una región) durante un período determinado de tiempo (normalmente un año).

  El ratio deuda/PIB indica si una economía que produce y vende bienes y servicios son suficientes para pagar sus deudas sin incurrir en más deuda. Uno de los criterios de convergencia de la Unión Europea era que la deuda pública en relación al PIB debería de estar por debajo del 60 %.

- *Carga Financiera:* el índice de carga financiera muestra el sacrificio soportado por cada ciudadano como consecuencia del recurso al endeudamiento. Mide la carga presupuestaria de la deuda, que anualmente se calcula como sumatorio de los gastos financieros y los gastos derivados de la amortización del principal de la deuda.

- *Gastos Financieros:* se podrían definir los gastos financieros como toda retribución satisfecha o gasto incurrido que sea consecuencia de la titularidad de toda clase de pasivos financieros.

- *Ingresos Totales:* son los ingresos que recibe una empresa procedentes de la venta de sus productos o servicios.

- *Ingresos por Impuestos:* Ingresos recibidos por el Estado u otro sujeto activo por la aplicación de determinados impuestos.

- *Gasto per Cápita:* son los gastos totales por habitante.

- *Nº Matriculaciones:* número de matriculaciones que se hacen en una Comunidad Autónoma.

- *Importaciones:* son el transporte legítimo de bienes y servicios del extranjero los cuales son adquiridos por un país para distribuirlos en el interior de este. Las importaciones pueden ser cualquier producto o servicio recibido dentro de la frontera de un Estado con propósitos comerciales.

- *Exportaciones:* es cualquier bien para la economía o el servicio enviado fuera del territorio nacional.

- *Esperanza de vida:* estimación del promedio de años que viviría un grupo de personas nacidas el mismo año si los movimientos en la tasa de mortalidad de la región evaluada se mantuvieran constantes.

- *Exportaciones – Importaciones (Balanza):* es la diferencia entre las exportaciones e importaciones de un país. La Balanza comercial es positiva, cuando el valor de las exportaciones es superior que el de las importaciones, entonces se dice que hay superávit comercial.

No ha sido posible recopilar toda la información económica-financiera de todos los años de las CCAA ya que no se encuentra publicada en la bibliografía utilizada en este estudio. Por este motivo, el total de casos a analizar se han visto reducidos.

Figura 3:

Captura de base de datos creada con Hoja de Cálculo.

INDICADORES ECONÓMICO-FINANCIEROS DE LAS CCAA

CCAA	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7	Variable 8	Variable 9	Variable 10	Variable 11	Variable 12	Variable 13	Variable 14
Andalucía-2014	-0,0798038	0,00363953	0,01104775	0,2053647	0,0393836	0,75230731	1,23727759	0,3420	1,05	0,335788312	0,544259484	0,432125895	3049	0,793970353
Andalucía-2013	-0,0919677	0,06484803	0,04510007	0,17299566	0,00590949	0,75958028	1,03536789	0,3630	1,04	0,335105486	0,530526197	0,407847613	3183	0,768760554
Andalucía-2012	-0,1165469	-0,0034847	-0,0242319	0,14873815	-0,0295094	0,7560285	0,83444701	0,3570	1,05	0,214330895	0,696225678	0,548870342	3290	0,788351189
Andalucía-2011	-0,3801171	0,04363498	0,01392297	0,10175052	-0,0712756	0,75901897	0,58806446	0,3100	1,13	0,334397677	0,542805645	0,430983313	3047	0,79399195
Andalucía-2010	-0,2263988	0,0638648	-0,0203876	0,08596749	-0,0940369	0,75716669	0,47758552	0,2800	1,2	0,478928279	0,354034862	0,27601852	3291	0,779636554
Andalucía-2009	-0,082427	0,02606305	-0,0426389	0,06876944	-0,1232655	0,76165856	0,35810901	0,2630	1,1	0,451935491	0,42520598	0,353846199	3552	0,832175971
Aragón-2014	-0,0726062	0,01930698	-0,0112437	0,18123153	0,06037268	1,11471508	1,49953048	0,1870	0,92	0,175554127	0,722933494	0,539210027	3717	0,745863944
Aragón-2013	-0,1764206	0,03394114	0,08880169	0,16289442	0,03890093	1,11661791	1,31373362	0,2060	0,9	0,198670468	0,716630583	0,573457199	3505	0,800213127
Aragón-2012	-0,1283563	0,03700326	-0,0149287	0,13958068	0,01140973	1,10505014	1,08901964	0,1870	0,93	0,059346295	0,892979617	0,70893688	3451	0,793900405
Aragón-2011	-0,3667228	0,02388139	0,02636727	0,09971285	-0,0290074	1,11081803	0,77464775	0,1680	0,94	0,208428143	0,701943578	0,582508486	3647	0,829850866
Aragón-2009	-0,1136247	0,01415915	-0,0877436	0,05536352	-0,0905534	1,09413825	0,37941818	0,1360	1,12	0,376007695	0,519399144	0,443201528	4034	0,853296608
Aragón-2008	-0,0678824	-0,6062043	-0,3149308	0,04259441	-0,0992799	1,09874008	0,30022629	0,0970	1,03	0,3783545	0,543305481	0,493055092	3637	0,90750988
Asturias-2014	-0,085192	0,02457509	-0,0010498	0,16225922	0,01553864	0,9110264	1,10590635	0,2080	0,88	0,167675141	0,746695062	0,631102329	3598	0,845194192
Asturias-2013	-0,0669748	0,0291564	0,02703948	0,14329984	-0,0081272	0,90708244	0,94632932	0,2230	0,86	0,2101894	0,699557079	0,593159077	3561	0,847906618
Asturias-2012	-0,0659252	9,0308E-06	9,1258E-06	0,12287552	-0,0297175	0,91350907	0,80525017	0,2380	0,89	0,190123311	0,699788306	0,533126146	3784	0,761839176
Asturias-2010	-0,233448	0,03423114	-0,0487716	0,07438017	-0,0771542	0,9177097	0,4908467	0,1660	1,05	0,348552167	0,488779608	0,368217484	3943	0,75334052
Asturias-2009	-0,09754	-0,0746305	-0,1246511	0,04752684	-0,1103798	0,90890898	0,30098061	0,1410	1,01	0,355788856	0,514207642	0,410730605	4131	0,798764101
Canarias-2014	-0,0691244	0,00998439	-0,0081125	0,14531705	0,01292449	0,87007871	1,09762244	0,3110	0,5	0,412441665	0,447415578	0,356254132	3243	0,796248834
Canarias-2013	-0,075998	-0,0193482	0,03774471	0,12970013	-0,0066748	0,86834214	0,95105555	0,3310	0,5	0,41494129	0,45268435	0,369530314	3216	0,81630901
Cantabria-2014	-0,104729	0,0245857	-0,0230281	0,19854444	0,04394518	0,93175886	1,28425218	0,1840	0,95	0,211363206	0,676751982	0,522736446	4181	0,772419528
Cantabria-2013	-0,0794532	-0,0271349	-0,0007332	0,17912657	0,0197181	0,93314496	1,12369545	0,1980	0,91	0,23971548	0,665739635	0,562693093	3902	0,84521495
Cantabria-2012	-0,1229437	0,03779074	-0,0167581	0,16432153	0,01172224	0,93906393	1,0768171	0,1920	0,98	0,073418928	0,866771276	0,670211802	4061	0,773227979
Cantabria-2011	-0,4262791	-0,0125117	0,03817311	0,1018752	-0,0526439	0,93710451	0,65930489	0,1590	1,07	0,207305899	0,70476939	0,576347038	3731	0,81778103
Cantabria-2010	-0,2653695	0,09919718	-0,0314254	0,0773429	-0,0736724	0,9371673	0,51215272	0,1470	1,22	0,361600427	0,501813121	0,395215931	3980	0,787575921
CastillaLaMancha-2014	-0,1150145	0,04178429	-0,0004904	0,33976324	0,18606234	0,81833567	2,21054812	0,2850	0,53	0,260521772	0,601056785	0,431674767	3473	0,718192985
CastillaLaMancha-2013	-0,1242113	-0,0036508	0,03587792	0,29782597	0,13844194	0,82775093	1,86860614	0,2900	0,53	0,3065725	0,586794097	0,478725505	3196	0,815832176
CastillaLaMancha-2012	-0,0815116	0,18716818	0,03738	0,26706854	0,10855146	0,81627944	1,68479346	0,3010	0,64	0,16352685	0,724521731	0,528541826	4117	0,729504448
CastillaLaMancha-2011	-0,5382397	0,04624277	0,10871029	0,17625678	-0,014204	0,81140839	0,92542281	0,2450	0,68	0,405886263	0,502253601	0,433932633	4403	0,863971173

Fuente. Elaboración propia.

5.2 Análisis de la variable dependiente

Calificaciones

La variable dependiente debe ser cualitativa (nominal u ordinal) y tener como mínimo dos grupos definidos que sean mutuamente excluyentes y exhaustivos, bien desde el punto de vista de su naturaleza categórica, bien desde su naturaleza dicotómica. A continuación, se expone la variable dependiente utilizada.

El modo que tienen de evaluar las tres agencias más importantes a nivel mundial es el siguiente:

Tabla 2:
Clasificación de las calificaciones realizadas por las principales agencias de calificación a nivel internacional.

Calidad	Moody's	S&P	Fitch
Principal	Aaa	AAA	AAA
Alto grado	Aa1	AA+	AA+
	Aa2	AA	AA
	Aa3	AA-	AA-
Grado medio superior	A1	A+	A+
	A2	A	A
	A3	A-	A-
Grado medio inferior	Baa1	BBB+	BBB+
	Baa2	BBB	BBB
	Baa3	BBB-	BBB-
Grado de no inversión	Ba1	BB+	BB+
especulativo	Ba2	BB	BB
	Ba3	BB-	BB-
Altamente especulativa	B1	B+	B+
	B2	B	B
	B3	B-	B-
Riesgo sustancial	Caa1	CCC+	CCC+
	Caa2	CCC	CCC
	Caa3	CCC-	CCC-
Extremadamente especulativa	Ca	CC	CC
			C
A falta de pocas			

Fuente. Elaboración propia.

Para el análisis, las calificaciones se han dividido en cuatro grupos (grupo 1, grupo 2, grupo 3 y grupo 4). Las equivalencias que se han efectuado con respecto a la calificación han sido las siguientes:

- Alto Grado y Principal → 1
- Grado medio superior → 2
- Grado medio inferior → 3
- Grado de no inversión especulativo → 4

En algunos casos, las agencias diferían en su calificación (aunque siempre la diferencia era tan solo en una unidad de calificación). Cuando ha ocurrido ésto, se ha tomado el dato en la que al menos dos agencias coincidían en la calificación y cuando la calificación era diferente en las tres, se consideró el promedio.

5.3 Análisis de la normalidad

Para comprobar si los datos de la muestra se distribuyen normalmente se utilizará el contraste de normalidad de Shapiro-Wilk. La lógica de la prueba se basa en las desviaciones que presentan las estadísticas de orden de la muestra respecto a los valores esperados de los estadísticos de orden de la normal estándar. El contraste de normalidad de Shapiro-Wilk se utiliza para muestras menores de 50 observaciones. La muestra del presente proyecto no supera dicha cantidad (siempre habrá menos de 50 observaciones en cada uno de los cuatro grupos).

Las hipótesis a contrastar son:

$$H_0: Los\ elementos\ de\ la\ muestra\ aleatoria\ siguen\ una\ distribución\ normal$$

$$H_1: Los\ elementos\ de\ la\ muestra\ aleatoria\ no\ siguen\ una\ distribución\ normal$$

El estadístico es el siguiente:

$$W = \left\{\sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i)\right\}^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{b^2}{S^2}$$

Siendo:

$$a_{n-i+1} = cuantiles\ esperados\ de\ x_i$$

$$y_{n-i+1} = dato\ mayor\ de\ la\ muestra\ ordenada$$

$$y_i = dato\ menor\ de\ la\ muestra\ ordenada$$

$$x_i = dato\ de\ la\ muestra\ ordenada$$

$$\bar{x} = media\ de\ la\ muestra$$

Se rechazará la hipótesis nula $(\mathrm{H_0})$ si al nivel de significación α 100%:

$$si\ W_0 \leq W_{t,\alpha} \rightarrow Rechazar\ H_0$$

El contraste se realizará seleccionando un α de 0.05 por lo que la región de rechazo serán todos los valores menores o iguales a $\mathrm{W_t}$con un α de 0.05, es decir, $\mathrm{W_{t,0.05}}$, cuantil 0,95.

En la tabla 1 se muestra los datos de la variable 1 (Superávit o Déficit / Ingresos Corrientes) para el cálculo del estadístico de Shapiro-Wilks del grupo 1, es decir, aquellas Comunidades Autónomas calificadas con calificación 1.

Tabla 3:
Análisis de Normalidad (I).

Datos Variable 1	Datos ordenados de menor a mayor	$(x_i - \bar{x})^2$	a_{n-i+1}	$y_{n-i+1} - y_i$	$a_{n-i+1}(y_{n-i+1} - y_i)$
-0,2263988	-0,426279108	0,054425593	0,4254	0,37171041	0,15812561
-0,082427	-0,385307065	0,036987325	0,2944	0,31742465	0,09344982
-0,12835634	-0,366722753	0,030184399	0,2487	0,29365191	0,07303123
-0,36672275	-0,325057562	0,017442852	0,2148	0,24263056	0,05211704
-0,1136247	-0,30431266	0,012393587	0,1870	0,20677265	0,03866649
-0,06788241	-0,281319455	0,007802769	0,1630	0,18106397	0,02951343
-0,23344796	-0,265369454	0,005239339	0,1415	0,15338160	0,02170350
-0,09754001	-0,23595929	0,001846689	0,1219	0,12233459	0,01491259
-0,12294369	-0,233447962	0,001637156	0,1036	0,11050428	0,01144824
-0,42627911	-0,226398803	0,001116404	0,0862	0,09975819	0,00859916
-0,26536945	-0,214246576	0,000452005	0,0697	0,08589024	0,00598655
-0,32505756	-0,209679975	0,000278683	0,0537	0,07540097	0,00404903
-0,20545385	-0,205453846	0,000155443	0,0381	0,04698295	0,00179005
-0,12664061	-0,200401597	5,49885E-05	0,0227	0,04161689	0,00094470
-0,20967998	-0,196143747	9,97027E-06	0,0076	0,01349065	0,00010253
-0,10025549	-0,182653101	0,000106772	0,0000		
-0,1826531	-0,158784704	0,001169741			
-0,23595929	-0,158470898	0,001191304			
-0,1584709	-0,134279006	0,003446532			
-0,07307084	-0,128356341	0,004177015			
-0,21424658	-0,126640612	0,004401734			
-0,0545687	-0,122943685	0,00490595			
-0,11198785	-0,113624703	0,006298243			
-0,30431266	-0,111987851	0,006560728			
-0,38530707	-0,10025549	0,00859898			
-0,1587847	-0,097540012	0,00910997			
-0,13427901	-0,082427	0,012223331			
-0,19614375	-0,073070839	0,014379687			
-0,2004016	-0,067882411	0,015650951			
-0,28131946	-0,054568697	0,019159398			

Siendo:

$$\bar{x} = -0{,}192986173$$

$$\sum (x_i - \bar{x})^2 = 0{,}281407538$$

$$\sum a_{n-i+1}(y_{n-i+1} - y_i) = 0{,}51443996$$

$$\left\{\sum a_{n-i+1}(y_{n-i+1} - y_i)\right\}^2 = 0{,}26464847$$

$$W_0 = \left\{\sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i)\right\}^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{b^2}{S^2} = \frac{0{,}281407538}{0{,}26464847}$$
$$= 0{,}94044557$$

Según tablas estadísticas, la distribución del estadístico de Shapiro-Wilk (W) para el contraste de normalidad es:

$$W_{t;\alpha} = W_{30;0,05} = 0{,}927$$

Se comprueba:

Se rechazará la hipótesis nula (H_0) si:

$$si\ W_0 \leq W_{t,\alpha} \rightarrow Rechazar\ H_0$$

$$W_{Tabla} = W_{t;\alpha} = W_{30;0,05} = 0{,}927$$

$$W_0 = \frac{b^2}{S^2} = \frac{0{,}281407538}{0{,}26464847} = 0{,}94044557$$

Dado que $W_0 > W_{t;\alpha}$, entonces no se rechaza la hipótesis nula. Dado que no existe evidencia muestral para rechazar la hipótesis nula por lo que se puede afirmar que la distribución de la muestra es normal.

A continuación, se presenta la tabla 2 donde se representarán los distintos valores para el cálculo del estadístico del grupo 2, es decir, aquellas Comunidades Autónomas con calificación 2:

Tabla 4:
Análisis de Normalidad (II).

Datos	Datos ordenados de menor a mayor	$(x_i - \bar{x})^2$	a_{n-i+1}	$y_{n-i+1} - y_i$	$a_{n-i+1}(y_{n-i+1} - y_i)$
-0,38011711	-0,53823967	0,10824988	0.4643	0,47231452	0,21929563
-0,53823967	-0,39128075	0,03314387	0.3185	0,32430594	0,10329144
-0,35967875	-0,38737801	0,03173808	0.2578	0,31477179	0,08114817
-0,39128075	-0,38011711	0,02920371	0.2119	0,30178435	0,0639481
-0,34527099	-0,36357772	0,0238244	0.1736	0,2841245	0,04932401
-0,38737801	-0,35967875	0,02263598	0.1339	0,27448677	0,03675378
-0,36357772	-0,34527099	0,01850819	0.1092	0,25328517	0,02765874
-0,09600104	-0,18677623	0,000504	0.0804	0,09077519	0,00729832
-0,07833276	-0,17642055	0,00107621	0.0530	0,07169158	0,00379965
-0,18677623	-0,10535762	0,01078868	0.0263		
-0,09198582	-0,10472897	0,01091967	0.0000		
-0,10535762	-0,09600104	0,01281993			
-0,07260622	-0,09198582	0,0137453			
-0,17642055	-0,08519198	0,01538448			
-0,08519198	-0,07945321	0,01684102			
-0,06697481	-0,07833276	0,01713309			
-0,06592515	-0,07260622	0,01866501			
-0,10472897	-0,06697481	0,02023545			
-0,07945321	-0,06592515	0,02053518			

Siendo:

$$\bar{x} = -0{,}20922618$$

$$\sum (x_i - \bar{x})^2 = 0{,}42595212$$

$$\sum a_{n-i+1}(y_{n-i+1} - y_i) = 0{,}59251785$$

$$\left\{\sum a_{n-i+1}(y_{n-i+1} - y_i)\right\}^2 = 0{,}35107741$$

$$W = \left\{\sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i)\right\}^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{b^2}{S^2} = \frac{0{,}35107741}{0{,}42595212} = 0{,}824218$$

Según la tabla, la distribución del estadístico de Shapiro-Wilk (W) para el contraste de normalidad es:

$$W_{t;\alpha} = W_{19;0,05} = 0{,}901$$

Se comprueba:

Se rechazará la hipótesis nula (H_0) si:

$$si\ W_0 \leq W_{t,\alpha} \rightarrow Rechazar\ H_0$$

$$W_{Tabla} = W_{t;\alpha} = W_{30;0,05} = 0{,}901$$

$$W_0 = \frac{b^2}{S^2} = \frac{0{,}35107741}{0{,}42595212} = 0{,}824218$$

Dado que $W_0 < W_{t;\alpha}$, entonces se rechaza la hipótesis nula, por lo tanto se puede afirmar que la distribución de la muestra no es normal.

Este mismo proceso se repite para los grupos 3 y 4 aunque debido a que al menos este grupo 2 no se distribuye normalmente, la variable 1 no tendrá una distribución normal para los cuatro grupos de calificaciones.

Si se utiliza el programa IBM SPSS Statistics, se aceptará o rechazará la H_0 si:

$$sip_{spss} > p \rightarrow No\ Rechazar H_0;\ siendo\ p = 0{,}05$$

$$sip_{spss} \leq p \rightarrow Rechazar H_0;\ \ siendo\ p = 0{,}05$$

En el Anexo se encuentran las ventanas de IBM SPSS para realizar el análisis de normalidad, figuras 1 y 2.

Los resultados para la Variable 1:

Tabla 5:
Pruebas de normalidad.

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Calificac.	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Variable Superávit	1	0,124	30	0,200*	0,943	30	0,112
	2	0,277	19	0,000	0,810	19	0,002
	3	0,365	23	0,000	0,646	23	0,000
	4	0,252	8	0,144	0,800	8	0,029

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

Se puede observar en la tabla que el p-valor (Sig.) es mayor a 0.05 en el grupo 1 (Calificación = 1), por lo que se acepta la hipótesis nula en ese caso particular mientras que en los demás casos (Calificación = 2, 3, 4) el p-valor es menor de 0.05 por lo que se rechaza la hipótesis nula. Todo ello se presenta a continuación:

$$sip_{spss} > p \rightarrow AceptarH_0;\ siendo\ p = 0{,}05$$
$$sip_{spss} \leq p \rightarrow RechazarH_0;\ \ siendo\ p = 0{,}05$$

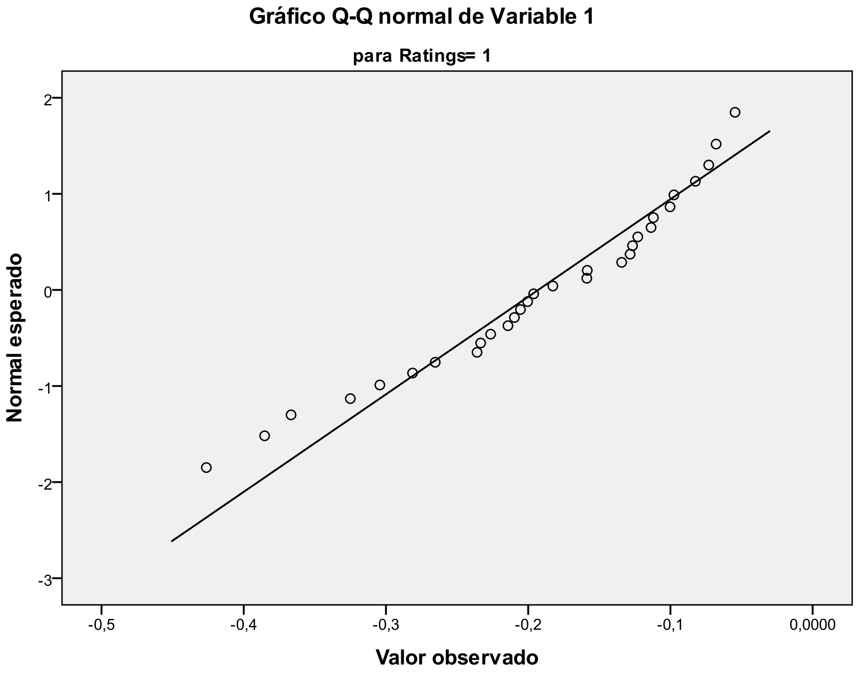
$$p_{1;spss} = 0.112 > p \rightarrow Se\ aceptaH_0;\ \ siendo\ p = 0{,}05$$

$$p_{2;spss} = 0.002 < p \rightarrow Se\ rechazaH_0;\ \ siendo\ p = 0{,}05$$
$$p_{3;spss} = 0.000 < p \rightarrow Se\ rechazaH_0;\ \ siendo\ p = 0{,}05$$
$$p_{4;spss} = 0.029 < p \rightarrow Se\ rechazaH_0;\ \ siendo\ p = 0{,}05$$

Además, se exponen las gráficas Normal Cuantil-Cuantil:

Se presenta el gráfico Q-Q (Cuantil-Cuantil) normal de la Variable Superávit para el grupo Calificación = 1:

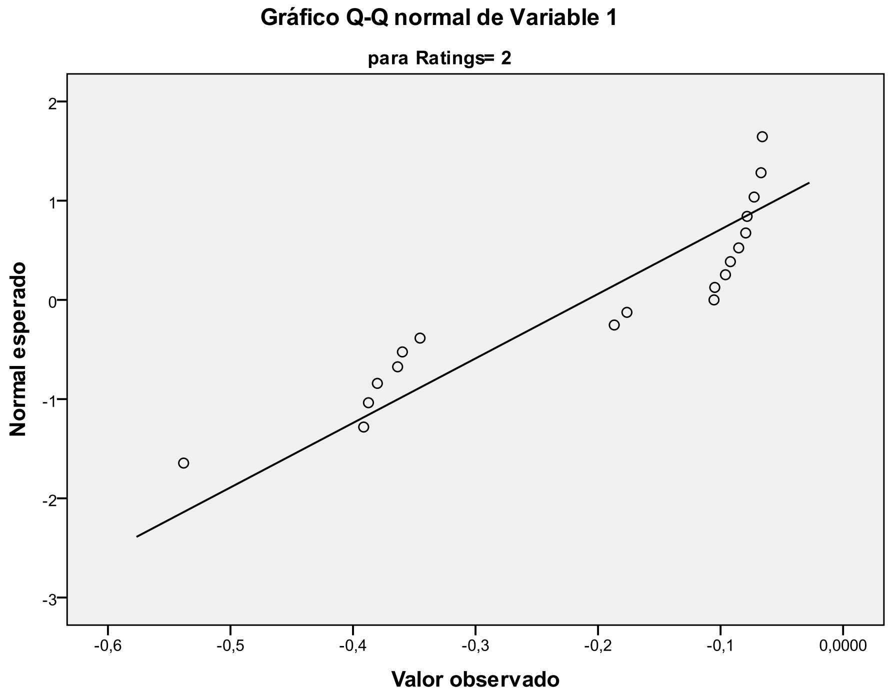
Figura 4:
Gráfico Normal Cuantil-Cuantil. Variable 1. Grupo Calificación =1.



Se observa cómo la recta se ajusta a los datos de la variable Superávit para el grupo Calificación = 1, tal y como determinaba el análisis hecho anteriormente.

A continuación, se muestra el gráfico Cuantil-Cuantil Normal de la Variable 1 para el grupo Calificación = 2:

Figura 5:
Gráfico Normal Cuantil-Cuantil. Variable 1. Grupo Calificación =2.



Se puede observar claramente que la recta normal no se ajusta a los datos de la Variable Superávit grupo Calificación = 2 que es justo lo que resultó del análisis hecho anteriormente.

En el Anexo se expone la tabla 1 sobre la prueba de normalidad para las 20 variables seleccionadas.

Aplicando el mismo análisis no se puede afirmar al 5% que los datos se rijan según un modelo Normal.

Debido a que en la mayoría de variables no se cumple la hipótesis de normalidades necesario realizar transformaciones a los datos de las variables para intentar que éstos cumplan la hipótesis de normalidad.

En el Anexo se encuentran las ventanas de IBM SPSS para realizar la transformación, figuras 3, 4, y 5.

Con la media y la desviación típica de la variable en cuestión se consiguen transformar los valores antiguos en variables con valores nuevos (Variable X Transformada) que se distribuyen normalmente (en una mayor parte de los casos).

En el Anexo se muestra la tabla 2 sobre la prueba de normalidad tras la transformación de las 20 variables.

Exceptuando las variables 9, 10 y 11, todas las demás se distribuyen normalmente para los cuatro grupos ya que su p-valor > 0.05 por lo que no existen evidencias para rechazarla hipótesis nula a un contraste de significación de 0.05:

$$\begin{cases} H_0: La\ variable\ en\ la\ población\ tiene\ distribución\ normal; p - valor > 0.05 \\ H_1: La\ variable\ en\ la\ población\ es\ distinta\ a\ la\ distribución\ normal;\ p - valor \leq 0.05 \end{cases}$$

Se muestran nuevamente las gráficas, esta vez ajustándose a la recta (variables 2 (Grupo Calificación=1) y 13 (Grupo Calificación=2):

Figura 6:
Gráfico Normal Cuantil-Cuantil Variable Transf. 2. Calificación 1.

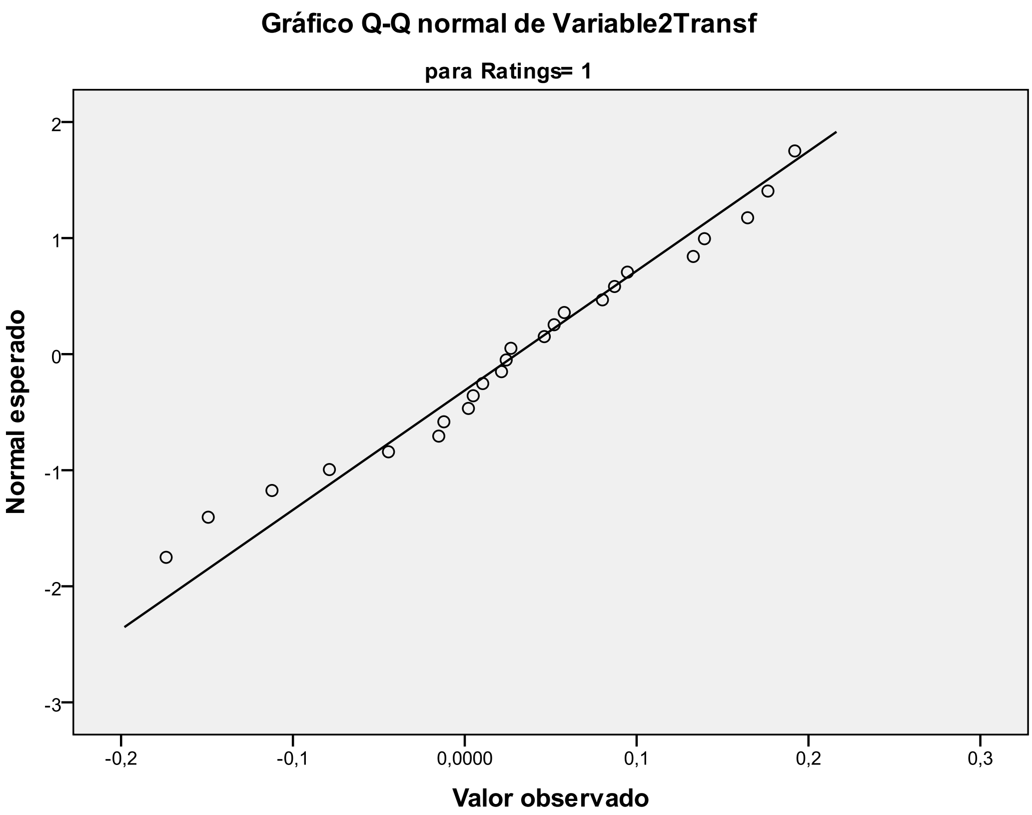
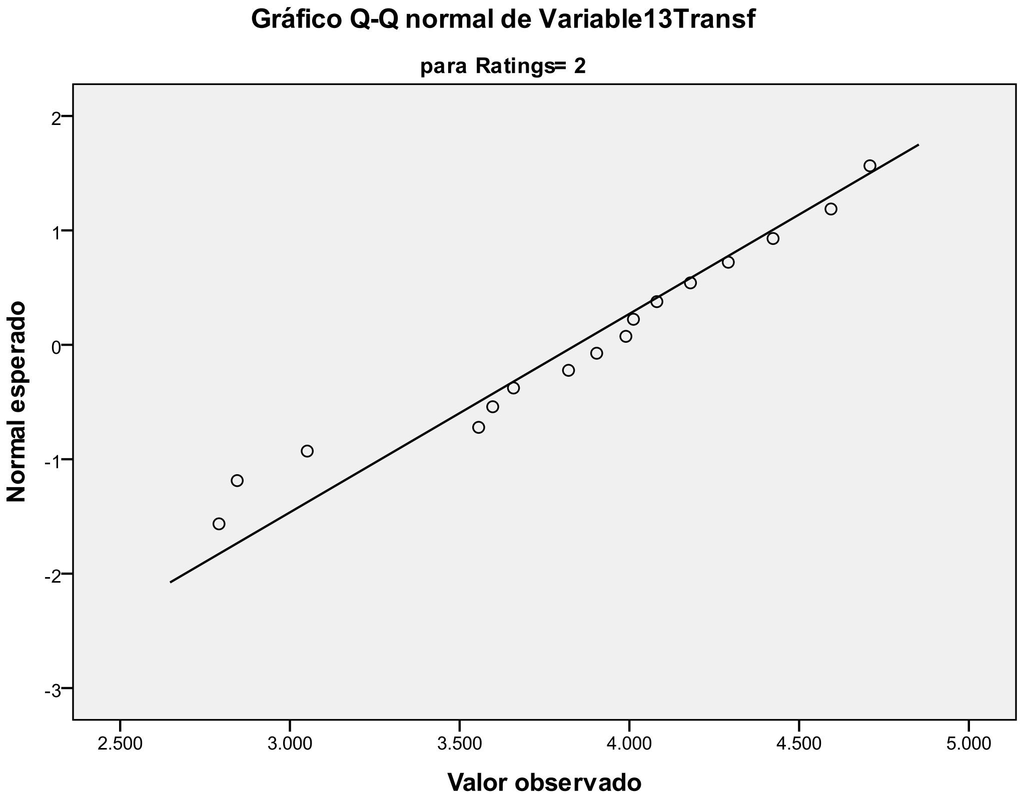


Figura 7:
Gráfico Normal Cuantil-Cuantil Variable Transf. 2. Calificación 2.



Después de este primer supuesto, las variables que continúan en el análisis son:

Tabla 6:
Estado de las variables tras la Prueba de Normalidad.

Variables que continúan el análisis	Variables excluidas
Superávit	
Tasa Anual Crecim. Ingr. Corr.	
Tasa Anual Crecim. Gast. Corr.	
Deuda/PIB	
(Deuda – Ingr. Corr.) / PIB	
PIB / Habitante año según media	
Deuda / Ingresos Corrientes	
Carga Financiera / Ingr. Corr.	
	~~Ingr. Financieros/Ingr. Totales~~
	~~Ingr. Corr./Ingr. Totales~~
	~~Ingr. Impuestos / Ingr. Corrientes~~
Ingr. Impuestos Totales / Ingr.Totales	
Gasto Per Cápita	
Ingresos Corrientes / Ingresos Totales	
Ingr. Capítulos (6, 7 y 8) / Ingr. Totales	
Nº Matriculaciones / Nº Habitantes	
Importaciones % PIB	
Exportaciones % PIB	
Esperanza de vida	
Balanza Comercial Export. –Import.	

Fuente. Elaboración propia.

Destacar que para los contrastes de significación el nivel ha sido del 5%.

5.4 Homocedasticidad de varianzas

Una hipótesis adicional para la validez del modelo de Análisis Discriminante Múltiple es la de homocedasticidad o igualdad de. Esta opción permite contrastar este supuesto mediante la prueba de Levene.

Las varianzas de la variable dependiente en los grupos que se comparan deben ser aproximadamente iguales. Por ello uno de los pasos previos a la comprobación de la existencia de diferencias entre las medias de varias muestras es determinar si las varianzas en tales muestras son iguales. Se seleccionará la prueba de Levene, esto es, aquella que emplea SPSS para comprobar que las varianzas de las variables dependientes en los grupos que se comparan sean aproximadamente iguales.

El contraste a validar es:

$$H_0: No\ existen\ diferencias\ significativas\ entre\ las\ varianzas$$

$$de\ los\ cuatro\ rating\ de\ las\ CCAA$$

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

$$H_1: Existen\ diferencias\ significativas\ entre\ las\ varianzas\ de\ los\ cuatro\ rating\ de\ las\ CCAA$$

$$H_1: \sigma_i^2 \neq \sigma_j^2\ para\ al\ menos\ i \neq j; i = 1,2,3; j = 2,3,4$$

dondeσ_i^2es la varianza poblacional de la prueba de salida del i-ésimo calificación de la Comunidad Autónoma (i= 1, 2, 3, 4).

Para su cálculo, en primer lugar, se calcula la diferencia (en valor absoluto) entre cada valor y la media de su grupo. Esta estimación se hará para la **Variable Deuda/PIB**:

$$D_{ij} = \left|X_{ij} - \bar{X}_j\right|$$

donde: $X_{ij}: puntuación\ del\ sujeto\ i\ perteneciente\ al\ grupo\ j$
$\bar{X}_j: media\ del\ grupo\ j$

Tabla 7:
Homocedasticidad de varianzas (I).

X_{ij}			
Grupo 1: X_{i1}	**Grupo 2: X_{i2}**	**Grupo 3: X_{i3}**	**Grupo 4: X_{i4}**
0,1	0,11	0,16	0,22
0,06	0,21	0,15	0,29
0,14	0,18	0,14	0,25
0,1	0,18	0,32	0,3
0,03	0,15	0,27	0,24
0	0,13	0,19	0,23
0,07	0,22	0,17	
0,02	0,2	0,15	
0,18	0,2	0,23	
0,08	0,12	0,28	
0,17	0,2	0,23	
0,12	0,22	0,21	
0,09	0,13	0,15	
0,04	0,18	0,21	
0,14	0,16	0,19	
0,17	0,12	0,16	
0,1		0,26	
0,12		0,24	
0,09		0,11	
0,08		0,14	
0,07			
0,03			
0,09			
0,08			

Las medias de cada grupo son:

Tabla 8:
Homocedasticidad de varianzas (II).

$\overline{X}_j$			
$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_4$
0,1268	0,27705882	0,33142857	0,79

En el segundo paso, se calculan la media de las diferencias de cada grupo:

$$\overline{D}_j = \frac{\sum D_{ij}}{n_j}$$

donde: $\sum D_{ij}: suma\ de\ las\ puntuaciones\ D\ en\ el\ grupo\ j$

$n_j: tamaño\ del\ grupo\ j$

Tabla 9:
Homocedasticidad de varianzas (III).

D_{ij}			
D_{i1}	D_{i2}	D_{i3}	D_{i4}
0,0268	0,16705882	0,17142857	0,57
0,0668	0,06705882	0,18142857	0,5
0,0132	0,09705882	0,19142857	0,54
0,0268	0,09705882	0,01142857	0,49
0,0968	0,12705882	0,06142857	0,55
0,1268	0,14705882	0,14142857	0,56
0,0568	0,05705882	0,16142857	
0,1068	0,07705882	0,18142857	
0,0532	0,07705882	0,10142857	
0,0468	0,15705882	0,05142857	
0,0432	0,07705882	0,10142857	
0,0068	0,05705882	0,12142857	
0,0368	0,14705882	0,18142857	
0,0868	0,09705882	0,12142857	
0,0132	0,11705882	0,14142857	
0,0432	0,15705882	0,17142857	
0,0268	0,10768382	0,07142857	
0,0068		0,09142857	
0,0368		0,22142857	
0,0468		0,19142857	
0,0568			
0,0968			
0,0368			
0,0468			

Las medias de los cuatro grupos son:

Tabla 10:
Homocedasticidad de varianzas (IV).

$\overline{D}_j$			
$\overline{D}_1$	$\overline{D}_2$	$\overline{D}_3$	$\overline{D}_4$
0,05021667	0,10768382	0,13342857	0,535

En el tercer paso, se calculan las medias totales de las diferencias:

$$\overline{D}_t = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} D_{ij}}{N}$$

donde: $\sum_{i=1}^{n} \sum_{j=1}^{k} D_{ij}$: *es la suma de las puntuaciones D de todos los sujetos*

N: *suma de todos los sujetos*

$$\overline{D}_t = 0{,}13343504$$

El cuarto paso consiste en calcular la suma de cuadrados intra-grupo:

$$SC_{intra} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\overline{D}_{ij} - \overline{D}_j\right)^2$$

Tabla 11:
Homocedasticidad de varianzas (V).

$\left(\overline{D}_{ij} - \overline{D}_j\right)^2$			
$(\overline{D}_{i1} - \overline{D}_1)^2$	$(\overline{D}_{i2} - \overline{D}_2)^2$	$(\overline{D}_{i3} - \overline{D}_3)^2$	$(\overline{D}_{i4} - \overline{D}_4)^2$
0,00054834	0,00352539	0,001444	0,001225
0,00027501	0,00165039	0,002304	0,001225
0,00137023	0,00011289	0,003364	0,000025
0,00054834	0,00011289	0,014884	0,002025
0,00217001	0,00037539	0,005184	0,000225
0,00586501	0,00155039	6,4E-05	0,000625
4,334E-05	0,00256289	0,000784	
0,00320167	0,00093789	0,002304	
8,9003E-06	0,00093789	0,001024	
1,1674E-05	0,00243789	0,006724	
4,9234E-05	0,00093789	0,001024	
0,00188501	0,00256289	0,000144	
0,00018001	0,00155039	0,002304	
0,00133834	0,00011289	0,000144	
0,00137023	8,7891E-05	6,4E-05	
4,9234E-05	0,00243789	0,001444	
0,00054834		0,003844	
0,00188501		0,001764	
0,00018001		0,007744	
1,1674E-05		0,003364	
4,334E-05			
0,00217001			
0,00018001			
1,1674E-05			

$$SC_{intra} = \sum_{i=1}^{n} \sum_{j=1}^{k} \left(\overline{D}_{ij} - \overline{D}_j\right)^2 = 0{,}11110838$$

En el quinto paso, se calcula la suma de cuadrados intergrupo:

$$SC_{inter} = \sum_{j=1}^{k} n_j \left(\overline{D}_j - \overline{D}_t\right)^2$$

Tabla 12:
Homocedasticidad de varianzas (VI).

$\sum_{j=1}^{k} n_j(\bar{D}_j - \bar{D}_t)^2$			
$n_1(\bar{D}_1 - \bar{D}_t)^2$	$n_2(\bar{D}_2 - \bar{D}_t)^2$	$n_3(\bar{D}_3 - \bar{D}_t)^2$	$n_3(\bar{D}_3 - \bar{D}_t)^2$
0,08898723	0,00018764	0,00996382	1,07810462

$$SC_{inter} = \sum_{j=1}^{k} n_j(\bar{D}_j - \bar{D}_t)^2 = 1{,}1772433$$

En este sexto paso se calculan los grados de libertad:

Grados de libertad, $G.L.(inter) = k - 1; siendo\ k\ el\ número\ de\ grupos$

$$G.L.(inter) = k - 1 = 4 - 1 = 3$$

Grados de libertad,

$$G.L.(intra) = \sum_{j=1}^{k}(n_j - 1); siendo\ n_j el\ tamaño\ muestral\ del\ grupo\ j$$

$$G.L.(intra) = \sum_{j=1}^{k}(n_j - 1) = (24-1) + (16-1) + (20-1) + (6-1) = 62$$

Por último, se calculan las medias cuadráticas intra-grupos e inter-grupos, para posteriormente determinar el valor experimental:

$$MC_{inter} = \frac{SC_{inter}}{G.L._{inter}} = \frac{0{,}11110838}{3} = 0{,}03703613$$

$$MC_{intra} = \frac{SC_{intra}}{G.L._{intra}} = \frac{1{,}1772433}{62} = 0{,}0189878$$

$$F_{Prueba} = \frac{MC_{inter}}{MC_{intra}} = \frac{0{,}03703613}{0{,}0189878} = 1{,}9505$$

Una vez obtenida F_{Prueba}, se comprobará si se rechaza o no la hipótesis nula:

$$F_{Prueba} \leq F_{Tabla} \rightarrow No\ se\ rechaza\ H_0$$

$$F_{Tabla} = 2{,}7532$$

Por lo tanto:

$$F_{Prueba} \leq F_{Tabla} \rightarrow No\ se\ rechaza\ H_0 \rightarrow 1{,}9505 \leq 2{,}7532$$

Se puede concluir que no se rechaza la hipótesis nula. Se puede afirmar que las varianzas de los cuatro grupos de la variable 4 son aproximadamente iguales.

En el Anexo se encuentran las ventanas de IBM SPSS para realizar la prueba de homocedasticidad de varianzas, figuras 6 y 7.

La prueba de Levene determina:

$$\begin{cases} H_0: No\ existen\ diferencias\ significativas\ entre\ las\ varianzas \\ H_1: Existen\ diferencias\ significativas\ entre\ las\ varianzas \end{cases}$$

Se muestra la tabla obtenida en IBM SPSS Statistics:

Tabla 13:
Prueba de homogeneidad de varianzas.

Variables	Estadístico de Levene	gl1	gl2	Sig.
Superávit	2,935	3	62	0,040
T. Anual Crec. Ingr. C.	1,391	3	62	0,254
T. Anual Crec. Gast. C.	0,160	3	62	0,923
Deuda/PIB	1,378	3	62	0,258
(Deuda – Ingr. C.) / PIB	0,857	3	62	0,468
PIB / Habit. año media	0,143	3	62	0,934
Deuda / Ingresos Corr.	1,377	3	62	0,258
Carga Financ. / Ingr. C.	1,103	3	62	0,355
Ingr. Imp.T. / Ingr. T.	0,991	3	62	0,403
Gasto Per Cápita	0,583	3	62	0,628
Ingresos C. / Ingresos T.	0,459	3	62	0,712
Ing. C. (6, 7 y 8) / Ing. T.	2,646	3	62	0,057
NºMatric. / Nº Habit.	1,301	3	62	0,282
Importaciones % PIB	2,448	3	62	0,072
Exportaciones % PIB	2,937	3	62	0,040
Esperanza de vida	1,115	3	62	0,350
Balanza C.Exp. – Imp.	0,270	3	62	0,847

Las variables 1 y 18 rechazan la hipótesis nula por lo que se eliminarán del análisis ($p - valor \leq 0.05$).

Tabla 14:
Estado de las variables tras la Prueba de Homocedasticidad.

Variables que continúan el análisis	Variables excluidas
	~~Superávit~~
Tasa Anual Crecim. Ingr. Corr.	
Tasa Anual Crecim. Gast. Corr.	
Deuda/PIB	
(Deuda – Ingr. Corr.) / PIB	
PIB / Habitante año según media	
Deuda / Ingresos Corrientes	
Carga Financiera / Ingr. Corr.	
	~~Ingr. Financieros/Ingr. Totales~~
	~~Ingr. Corr./Ingr. Totales~~
	~~Ingr. Impuestos / Ingr. Corrientes~~
Ingr. Impuestos Totales / Ingr. Totales	
Gasto Per Cápita	
Ingresos Corrientes / Ingresos Totales	
Ingr. Capítulos (6, 7 y 8) / Ingr. Totales	
Nº Matriculaciones / Nº Habitantes	
Importaciones % PIB	
	~~Exportaciones % PIB~~
Esperanza de vida	
Balanza Comercial Export. –Import.	

Fuente. Elaboración propia.

5.5 Independencia de las medias

¿Difieren significativamente las medias poblacionales de los grupos?

La respuesta que se dé a esta pregunta es crucial para la justificación de la realización del Análisis Discriminante. En el caso de que la respuesta fuese negativa carecería de interés continuar con el análisis, ya que significaría que las variables introducidas como variables clasificadoras no tienen una capacidad discriminante significativa.

La hipótesis nula y alternativa para dar respuesta a esta cuestión son las siguientes:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \mu_i \neq \mu_j \; para \; al \; menos \; i \neq j; i = 1,2,3; j = 2,3,4$$

Para su cálculo se debe seguir el siguiente procedimiento:

Se quiere calcular F_{prueba}:

$$F_{prueba} = \frac{Estimación \; intermediante \; de \; varianza}{Estimacion \; interna \; de \; varianza} \Rightarrow$$

$$\Rightarrow F_{Prueba} = \frac{s_x^2}{s_w^2} = \frac{n \cdot s_x^2}{(s_1^2 + s_2^2 + \cdots + s_k^2)/k}$$

Primeramente, se determinan los grados de libertad del numerador y denominador. Como se tienen cuatro grupos (variable dependiente Calificación) los grados de libertad serán:

$$G.L._{Numerador} = k - 1 = 4 - 1 \Rightarrow G.L._{Numerador} = 3$$

Los grados de libertad del denominador serán el número de datos restándole cuatro (número de grupos en variable dependiente Calificación):

$$G.L._{Denominador} = \sum_{i=1}^{K} n_i - k = 66 - 4 \Rightarrow G.L._{Denominador} = 62$$

Con tres grados de libertad en el numerador, 75 en el denominador y un nivel de significación de 0.05, con la lectura en tablas de F:

$$G.L._{Numerador} = 3$$

$$G.L._{Denominador} = 62$$

$$\alpha = 0.05$$

$$F_{Tabla} = 2{,}7532$$

$$F_{Prueba} \leq F_{Tabla} \rightarrow No\ se\ rechaza\ H_0$$

$$F_{Prueba} > F_{Tabla} \rightarrow Se\ rechaza\ H_0, aceptando\ H_1$$

A continuación, se calculan las medias aritméticas de los cuatro grupos:

$$\bar{x_j} = \frac{\sum x_i}{n_j}$$

$$\bar{x}_1 = \frac{\sum x_i}{n_1} = \frac{2{,}17}{24} = 0{,}0904$$
$$\bar{x}_2 = \frac{\sum x_i}{n_2} = \frac{2{,}71}{16} = 0{,}1693$$
$$\bar{x}_3 = \frac{\sum x_i}{n_3} = \frac{3{,}96}{20} = 0{,}198$$
$$\bar{x}_4 = \frac{\sum x_i}{n_4} = \frac{1{,}53}{6} = 0{,}255$$

Posteriormente se calculan las varianzas muestrales:

$$\sum (x_i - \bar{x}_1)^2 = 0{,}0527$$

$$\sum (x_i - \bar{x}_2)^2 = 0{,}0219$$

$$\sum (x_i - \bar{x}_3)^2 = 0{,}05992$$

$$\sum (x_i - \bar{x}_4)^2 = 0{,}00535$$

Y ahora se obtienen las varianzas de los cuatro grupos:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

$$s_1^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{0{,}0527}{24 - 1} = 0{,}0023$$

$$s_2^2 = \frac{\sum (x_i - \bar{x})^2}{n_2 - 1} = \frac{0{,}0219}{16 - 1} = 0{,}00146$$

$$s_3^2 = \frac{\sum (x_i - \bar{x})^2}{n_3 - 1} = \frac{0{,}05992}{20 - 1} = 0{,}00315$$

$$s_4^2 = \frac{\sum (x_i - \bar{x})^2}{n_4 - 1} = \frac{0{,}00535}{6 - 1} = 0{,}00107$$

Se calcula la estimación interna (denominador):

$$s_w^2 = \frac{\frac{\sum(x_w - \bar{x})^2}{n_w - 1}}{k} \; para \; w = 1,2,3,4$$

$$s_w^2 = \frac{\frac{\sum(x_w - \bar{x})^2}{n_w - 1}}{k} = \frac{0,0023 + 0,00146 + 0,00315 + 0,00107}{4} = \frac{0,00798}{4} == 0,001995$$

Se determina la varianza de cada muestra

$$\bar{x}_1 = 0,0904$$
$$\bar{x}_2 = 0,1693$$
$$\bar{x}_3 = 0,198$$
$$\bar{x}_4 = 0,255$$

$$\bar{\bar{x}} = \frac{\sum \bar{x}_i}{k} \; para \; i = 1,2,3,4$$

$$\bar{\bar{x}} = \frac{\sum \bar{x}_i}{k} = \frac{0,0904 + 0,1693 + 0,198 + 0,255}{4} = \frac{0,1782}{4} = 0,0445$$

Se calcula la estimación intermediante (numerador):

$$s_x^2 = n \cdot s_{\bar{x}}^2 = n \cdot 0,0445 = 0,67$$

$$F_{Prueba} = \frac{s_x^2}{s_w^2} = \frac{0,67}{0,001995} = 335,84$$

En el Anexo se muestran las ventanas de IBM SPSS para realizar la prueba de independencia de medias, figuras 8 y 9.

Se muestra la tabla obtenida:

Tabla 15:
Prueba de independencia de medias.

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
T. Anual Crec. Ingr. C.	Inter-grupos	0,018	3	0,006	0,984	0,406
	Intra-grupos	0,380	62	0,006		
	Total	0,398	65			
T. Anual Crec. Gast. C.	Inter-grupos	0,065	3	0,022	11,376	0,000
	Intra-grupos	0,118	62	0,002		
	Total	0,182	65			
Deuda/PIB	Inter-grupos	0,198	3	0,066	29,732	0,000
	Intra-grupos	0,137	62	0,002		
	Total	0,335	65			

(Deuda – Ingr. C.) / PIB	Inter-grupos	0,249	3	0,083	21,324	0,000
	Intra-grupos	0,241	62	0,004		
	Total	0,490	65			
PIB / Habit. año media	Inter-grupos	0,051	3	0,017	0,609	0,612
	Intra-grupos	1,728	62	0,028		
	Total	1,779	65			
Deuda / Ingresos Corr.	Inter-grupos	13,519	3	4,506	23,136	0,000
	Intra-grupos	12,076	62	0,195		
	Total	25,595	65			
Carga Financ. / Ingr. C.	Inter-grupos	0,079	3	0,026	11,610	0,000
	Intra-grupos	0,140	62	0,002		
	Total	0,219	65			
Ingr. Imp. T. / Ingr. T.	Inter-grupos	0,217	3	0,072	2,254	0,091
	Intra-grupos	1,986	62	0,032		
	Total	2,203	65			
Gasto Per Cápita	Inter-grupos	2732094,388	3	910698,129	2,600	0,060
	Intra-grupos	21714834,540	62	350239,267		
	Total	24446928,928	65			
Ingresos C. / Ingresos T.	Inter-grupos	0,020	3	0,007	2,901	0,042
	Intra-grupos	0,146	62	0,002		
	Total	0,166	65			
Ing. C. (6, 7 y 8) / Ing. T.	Inter-grupos	0,001	3	0,000	0,559	0,644
	Intra-grupos	0,052	62	0,001		
	Total	0,053	65			
Nº Matric. / Nº Habit.	Inter-grupos	0,001	3	0,000	3,728	0,016
	Intra-grupos	0,003	62	0,000		
	Total	0,004	65			
Importaciones % PIB	Inter-grupos	882,017	3	294,006	4,172	0,009
	Intra-grupos	4368,853	62	70,465		
	Total	5250,871	65			
Esperanza de vida	Inter-grupos	4,491	3	1,497	2,654	0,056
	Intra-grupos	34,970	62	0,564		
	Total	39,461	65			
Balanza C. Exp. – Imp.	Inter-grupos	201,745	3	67,248	2,468	0,070
	Intra-grupos	1689,372	62	27,248		
	Total	1891,117	65			

Las variables que rechazan la hipótesis nula son: 2, 6, 12, 13, 15, 19 y 20 quedando para el Análisis Discriminante Múltiple: 3, 4, 5, 7, 8, 14, 16 y 17.

Tabla 16:
Estado de las variables tras la Prueba de Independencia de Medias.

Variables que continúan el análisis	Variables excluidas
	~~Superávit~~
	~~Tasa Anual Crecim. Ingr. Corr.~~
Tasa Anual Crecim. Gast. Corr.	
Deuda/PIB	
(Deuda – Ingr. Corr.) / PIB	
	~~PIB / Habitante año según media~~
Deuda / Ingresos Corrientes	
Carga Financiera / Ingr. Corr.	
	~~Ingr. Financieros/Ingr. Totales~~
	~~Ingr. Corr./Ingr. Totales~~
	~~Ingr. Impuestos / Ingr. Corrientes~~
	~~Ingr. Impuestos Totales / Ingr. Totales~~
	~~Gasto Per Cápita~~
Ingresos Corrientes / Ingresos Totales	
	~~Ingr. Capítulos (6, 7 y 8) / Ingr. Totales~~
Nº Matriculaciones / Nº Habitantes	
Importaciones % PIB	
	~~Exportaciones % PIB~~
	~~Esperanza de vida~~
	~~Balanza Comercial Export. – Import.~~

Fuente. Elaboración propia.

6 Resultados del Análisis Discriminante Múltiple

En este sexto capítulo se exponen las diferentes tablas y gráficas obtenidas con IBM SPSS Statistics con la diferente información que ellas ofrecen. Además, en el Anexo se muestra cómo hacer dicho análisis con la herramienta software.

En el proceso de Análisis Discriminante se buscan funciones discriminantes a partir de las variables independientes para clasificar a los individuos según los valores de la variable dependiente (Variable Calificación). Por ello, inicialmente se seleccionan las variables independientes que más discriminen (que proporcionen los centros de los grupos muy distintos entre sí y muy homogéneos dentro de sí).

En el Anexo se encuentras las ventanas de IBM SPSS para aplicar el Método Análisis Discriminante a las variables seleccionadas, figuras 10 y 11.

Seguidamente se muestran los resultados obtenidos en el visor de resultados que proporciona IBM SPSS Statistics:

<u>Resumen inicial</u>

La siguiente tabla 17 muestra el número de casos válidos en el análisis y los que se han excluidos, que en este caso es cero. En todos los casos analizados, sus variables independientes están completas, es decir, no hay ningún valor perdido. Es por ello por lo que el 100% de los casos son válidos y analizables. En aquellos casos en los que faltaba alguna variable, se excluyeron previamente.

Tabla 17:
Resumen del procesamiento para el análisis de casos.

Casos no ponderados		N	Porcentaje
Válidos		66	100,0
Excluidos	Códigos de grupo para perdidos o fuera de rango	0	Excluidos
	Perdida al menos una variable discriminante	0	
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables discriminantes.	0	
	Total excluidos	0	
Casos Totales		66	100,0

Estadísticos de grupo

A continuación, se muestra el conjunto de medias y desviaciones típicas de las variables introducidas según las cuatro calificaciones y el total. Ya en esta primera tabla se puede obtener algunas conclusiones:

Se observa, por ejemplo, que las variables 'Importaciones %PIB' con calificación igual a 4 tiene una media bastante superior que al resto de los grupos cuyas medias están más igualadas. También la variable 'Tasa Anual Crecimiento de Gastos Corrientes' es negativa para el grupo 1 y positiva para los restantes. Ocurre lo mismo para la variable '(Deuda – Ingresos Corrientes) / PIB' que es también negativa en el grupo 1 y positiva en los restantes.

Tabla 18:
Estadísticos de grupo.

Calificaciones		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
1	T. Anual Crec. Gast. C.	-0,0465	0,04725	24	24,000
	Deuda/PIB	0,0908	0,04736	24	24,000
	(Deuda – Ingr. C.) / PIB	-0,0619	0,05581	24	24,000
	Deuda / Ingresos C.	0,6191	0,40046	24	24,000
	Carga Financ. / Ingr. C.	0,1795	0,05284	24	24,000
	Ingr.C. / Ingr.T.	0,8237	0,05016	24	24,000
	Nº Matricul. / Nº Habit.	0,0239	0,00738	24	24,000
	Importaciones % PIB	19,7850	7,35149	24	24,000
2	T. Anual Crec. Gast. C.	0,0207	0,04990	16	16,000
	Deuda/PIB	0,1682	0,03785	16	16,000
	(Deuda – Ingr. C.) / PIB	0,0167	0,06111	16	16,000
	Deuda / Ingresos C.	1,2280	0,39056	16	16,000
	Carga Financ. / Ingr. C.	0,2174	0,03841	16	16,000
	Ingr.C. / Ingr.T.	0,8042	0,04154	16	16,000
	Nº Matricul. / Nº Habit.	0,0178	0,00592	16	16,000
	Importaciones % PIB	18,5989	6,42317	16	16,000
3	T. Anual Crec. Gast. C.	0,0121	0,03390	20	20,000
	Deuda/PIB	0,1984	0,05573	20	20,000
	(Deuda – Ingr. C.) / PIB	0,0507	0,07174	20	20,000
	Deuda / Ingresos C.	1,4921	0,53437	20	20,000
	Carga Financ. / Ingr. C.	0,2561	0,04817	20	20,000
	Ingr.C. / Ingr.T.	0,7855	0,05091	20	20,000
	Nº Matricul. / Nº Habit.	0,0175	0,00830	20	20,000

	Importaciones % PIB	19,3060	10,76348	20	20,000
4	T. Anual Crec. Gast. C.	0,0293	0,03740	6	6,000
	Deuda/PIB	0,2537	0,03234	6	6,000
	(Deuda – Ingr. C.) / PIB	0,1330	0,05578	6	6,000
	Deuda / Ingresos C.	1,9985	0,36715	6	6,000
	Carga Financ. / Ingr. C.	0,2668	0,04380	6	6,000
	Ingr.C. / Ingr.T.	0,7771	0,05039	6	6,000
	Nº Matricul. / Nº Habit.	0,0217	0,00468	6	6,000
	Importaciones % PIB	31,9275	7,82021	6	6,000
Total	T. Anual Crec. Gast. C.	-0,0056	0,05295	66	66,000
	Deuda/PIB	0,1570	0,07178	66	66,000
	(Deuda – Ingr. C.) / PIB	0,0090	0,08679	66	66,000
	Deuda / Ingresos C.	1,1566	0,62751	66	66,000
	Carga Financ. / Ingr. C.	0,2198	0,05800	66	66,000
	Ingr.C. / Ingr.T.	0,8031	0,05056	66	66,000
	Nº Matricul. / Nº Habit.	0,0203	0,00762	66	66,000
	Importaciones % PIB	20,4562	8,98792	66	66,000

Pruebas de igualdad de las medias

En la tabla 19 de Pruebas de igualdad de las medias de los grupos muestra información sobre F, es decir, muestra el poder que tiene esa variable en discriminar j.

Como variable discriminatoria mayor es claramente la variable Deuda/PIB, con $F = 29{,}732$. Les siguen las variables Deuda / Ingresos C. y (Deuda-Ingresos C.) / PIB, con una $F = 23{,}136$ y $F = 21{,}324$ respectivamente. Y, asimismo, continúan las variables Carga Financ. / Ingr. C. y T. Anual Crec. Gast. C en este orden. Cabe señalar, que las variables Nº Matricul. / Nº Habit., Importaciones % PIB e Ingr. C. / Ingr. T. tienen un poder discriminante bastante menor que las anteriores, precisamente aquellas con sigma cercano a 0,05. En estas variables no debería haber mucha diferencia, en principio, para clasificar un registro en las cuatro calificaciones, es decir, no serían buenas variables para clasificar comunidades autónomas en una calificación u otro. La información de esta tabla se suele utilizar como prueba preliminar para detectar si los grupos difieren en las variables clasificadoras seleccionadas. A pesar de ello, este análisis se está realizando a nivel univariante y podría cambiar cuando se realice a nivel multivariante, esto es, cuando se utilicen varias variables a la vez.

Se observa que sigma es menor a 0,05 para todas las variables por lo que no existen evidencias de que las medias de los distintos grupos sean iguales.

$$\sigma < 0{,}05 \rightarrow Se\ rechaza\ la\ H_0$$

Tabla 19:
Pruebas de igualdad de las medias de los grupos.

	Lambda de Wilks	F	gl1	gl2	Sig.
T. Anual Crec. Gast. C.	0,645	11,376	3	62	0,000
Deuda/PIB	0,410	29,732	3	62	0,000
(Deuda – Ingr. C.) / PIB	0,492	21,324	3	62	0,000
Deuda / Ingresos C.	0,472	23,136	3	62	0,000
Carga Financ. / Ingr. C.	0,640	11,610	3	62	0,000
Ingr.C. / Ingr.T.	0,877	2,901	3	62	0,042
Nº Matricul. / Nº Habit.	0,847	3,728	3	62	0,016
Importaciones % PIB	0,832	4,172	3	62	0,009

Variables introducidas/excluidas

En este caso se muestra la información de las distintas variables introducidas en el modelo. Se observa que el estadístico lambda de Wilks va disminuyendo a medida que se van introduciendo variables en el modelo, lo que significa que los grupos van estando cada vez menos solapados, es decir, su distinción en la nube de puntos es mayor.

En la tabla 20 de variables introducidas/excluidas se muestra un resumen de todos los pasos llevados a cabo para construir la función discriminante (o funciones discriminantes). A su vez, también recuerda los criterios utilizados para seleccionar las variables. En cada paso se muestra la variable introducida en el modelo y (si la hubiere), la variable excluida. En el caso que se está analizando solamente se muestran las variables introducidas ya que no hay ninguna variable que se excluya en esta ocasión. Si tal fuera el caso, se mostraría en otra columna con el encabezado “Excluidas”. En este análisis se incorpora primeramente la variable Deuda/PIB, posteriormente la variable Importaciones % PIB y por último la variable Carga Financ. / Ingr. C.

La leyenda que se encuentra debajo de la tabla indica información sobre el número de pasos máximo (16), valor de la F de entrada y salida.

Como ya se ha explicado anteriormente, en el proceso de Análisis Discriminante se buscan funciones discriminantes a partir de las variables independientes para clasificar a los individuos según los valores de la variable dependiente. Por ello, inicialmente se seleccionan las variables independientes que más discriminen (que proporcionen los centros de los grupos muy distintos entre sí y muy homogéneos dentro de sí). Las variables introducidas para discriminar en el modelo son la variable Deuda/PIB, Importaciones % PIB y Carga Financiera / Ingresos Corrientes. Estas son las variables con un poder de discriminación mayor y serán las utilizadas para la predicción de las calificaciones.

A continuación, se muestran estas tres variables que tienen un poder discriminador mayor:

Tabla 20:
Variables introducidas/excluidas[a,b,c,d] (I)

		Lambda de Wilks			
Paso	**Introducidas**	**Estadístico**	**gl1**	**gl2**	**gl3**
1	Deuda/PIB	0,410	1	3	62,000
2	Importaciones % PIB	0,334	2	3	62,000
3	Carga Financ. / Ing. C.	0,253	3	3	62,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 16.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

En el paso 1 se seleccionó la variable Deuda/PIB, en el paso 2 entró la variable Importaciones % PIB y en el paso 3 fue la variable Carga Financ. / Ing. C. Los valores de la Lambda de Wilks de la tabla anterior no son suficientemente próximos a cero por lo que es posible que los grupos no estén separados de manera clara. Los p-valores de Lambda de Wilks y los estadísticos F exacta (tabla siguiente) certifican la significatividad de dos ejes discriminantes, con lo que su capacidad explicativa será buena y el modelo formado por las tres variables es significativo (p-valores nulos).

En esta tabla 21 se muestra el valor transformado de lambda de Wilks y su significación. Los diferentes valores del estadístico se refieren al estadístico global.

Tabla 21:
Variables introducidas/excluidas[a,b,c,d] (II)

	Lambda de Wilks			
	F exacta			
Paso	Estadístico	gl1	gl2	Sig.
1	29,732	3	62,000	0,000
2	14,841	6	122,000	0,000
3				

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 16.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Variables en el análisis

Cuanto mayor sea la tolerancia de una variable, más información independiente del resto de variables recogerá. En la tabla 22 se muestra la tolerancia de las variables en los distintos pasos del método. Siempre la variable (Deuda/PIB) será aquella que recoja más información.

De este modo, si en una iteración dada del procedimiento por pasos la variable seleccionada verifica que su tolerancia con respecto a las variables ya incluidas en la función discriminante es muy pequeña entonces la variable no se incluye en dicha etapa. Así, se evita la redundancia de información.

Tabla 22:
Variables en el análisis.

	Paso	Tolerancia	F para salir	Lambda de Wilks
1	Deuda/PIB	1,000	29,732	
2	Deuda/PIB	0,976	30,295	0,832
	Importaciones % PIB	0,976	4,618	0,410
3	Deuda/PIB	0,976	17,643	0,476
	Importaciones % PIB	0,848	6,709	0,338
	Carga Financ. / Ing. C.	0,867	6,409	0,334

Autovalores

Esta siguiente tabla 23 muestra los autovalores y algunos estadísticos descriptivos multivariantes. Es posible comparar, de esta forma, las distintas capacidades discriminantes de cada función, en este caso tres funciones.

El autovalor es el cociente entre la variación debida a las diferencias entre los grupos (medida mediante la suma de cuadrados inter-grupos) y la variación que se da dentro de cada grupo (suma de cuadrados intra-grupos). Este estadístico se diferencia de la F del análisis de varianza multivariante en que no intervienen los grados de libertad. El autovalor sirve para comparar cómo se distribuye la dispersión inter-grupos cuando existe más de una función. Debido a que no tiene un valor máximo, se utiliza a la vez que el estadístico lambda de Wilks (el autovalor puede tomar un valor mínimo de cero y no tiene un valor máximo definido).

El autovalor de una función se interpreta como la parte de variabilidad total de la nube de puntos proyectada sobre el conjunto de todas las funciones atribuible a la función. Si su valor es alto, la función tendrá una mayor discriminación.

Tabla 23:
Autovalores.

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2,397[a]	93,7	93,7	0,840
2	0,140[a]	5,5	99,2	0,351
3	0,020[a]	0,8	100,0	0,140

a. Se han empleado las 3 primeras funciones discriminantes canónicas en el análisis.

Los autovalores de las tres funciones del modelo son bastante desiguales (sobre todo el de la función primera con respecto a los otros dos). La primera función explica el 93,7% de la variabilidad disponible en los datos, mientras que las dos siguientes explican tan sólo el 5,5% y 0,8% respectivamente.

En el AD se han utilizado cuatro grupos para clasificación, por lo tanto, se obtendrán tres funciones discriminantes:

$$g = 2; p = 9$$

$$q = \min(k, g-1) = min(9, 4-1) \rightarrow q = 3$$

De manera parecida al análisis de los autovalores, la correlación canónica de la primera función es alta (0,840) mientras van disminuyendo las siguientes correlaciones (0,351 y 0,140).

El cálculo analítico se muestra a continuación:

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

$$\eta_1 = \sqrt{\frac{\lambda_1}{\lambda_1 + 1}} = \sqrt{\frac{2{,}397}{2{,}397 + 1}} = 0{,}840$$

$$\eta_2 = \sqrt{\frac{\lambda_2}{\lambda_2 + 1}} = \sqrt{\frac{0{,}140}{0{,}140 + 1}} = 0{,}351$$

$$\eta_3 = \sqrt{\frac{\lambda_3}{\lambda_3 + 1}} = \sqrt{\frac{0{,}020}{0{,}020 + 1}} = 0{,}140$$

Se puede observar que el primer autovalorλ_1 es el que tiene una mayor discriminación (correlación canónica 0,840).

Las correlaciones canónicas miden las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones totales sin distinguir grupos. Si su valor es grande (próximo a 1) la dispersión será debida a las diferencias entre grupos, y, en consecuencia, la función discriminará mucho.

Contraste de las funciones. Lambda de Wilks

En la tabla 24 se muestra el contraste de la significación de las tres funciones obtenidas. En la primera línea de la tabla se contrasta la hipótesis nula de que el modelo completo (las tres funciones discriminantes) no permite distinguir las medias de los grupos. El valor de lambda de Wilks está más próxima a cero que a uno y, además, su nivel de significación es menor a 0.05, p-valor < 0,05. De esta manera, se puede afirmar que el modelo permite distinguir significativamente entre los grupos.

Para las dos funciones siguientes, lambda de Wilks está bastante próxima a uno y además sus p-valor son mayores a 0,05. No existirían evidencias para rechazar la hipótesis nula para los dos casos restantes, esto es, no se podría concluir que las funciones segunda y tercera permitan discriminar entre grupos. Con este análisis se podría prescindir de las funciones 2 y 3, no afectando demasiado al resultado.

Tabla 24:
Lambda de Wilks (I).

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 3	0,253	84,507	9	0,000
2 a la 3	0,860	9,305	4	0,054
3	0,980	1,226	1	0,268

El estadístico del contraste de significación global Lambda de Wilks:

$$\Lambda_1 = \frac{1}{(1+\lambda_1)\cdot(1+\lambda_2)\cdot(1+\lambda_3)} = \frac{1}{(1+2{,}397)\cdot(1+0{,}140)\cdot(1+0{,}020)} \Rightarrow$$

$$\Rightarrow \Lambda_1 = 0{,}2531$$

que conduce a rechazar la hipótesis nula de igualdad de medias, p-valor = 0,000 < 0,05, lo que indica que la función discriminante 1 es significativa. Las demás funciones discriminantes, 2, y 3, no serán significativas ya que sus p-valor > 0,05.

Igualmente, en el Análisis Discriminante Múltiple, se utiliza el estadístico V de Barlett para comprobar si los valores λ_i son estadísticamente significativos.

$$H_0: \lambda_1 = \lambda_4 = \lambda_3$$
$$H_1: No\ todos\ los\ \lambda_G\ \ son\ iguales$$

$$V = \left[N - 1 - \frac{K+G}{2}\right]\sum_{g=1}^{G-1} ln\left(1+\lambda_g\right) = \left[66 - 1 - \frac{3+4}{2}\right]\sum_{g=1}^{G-1} ln\left(1+\lambda_g\right) =$$

$$\left[66 - 1 - \frac{3+4}{2}\right]\left(ln(1+2{,}397) + ln(1+0{,}140) + ln(1+0{,}020)\right) = 84{,}484$$

$$\begin{cases} N: nº\ de\ casos \\ K: nº\ de\ variables\ discriminantes \\ G: nº\ de\ grupos \end{cases}$$

Por lo que se rechaza la hipótesis nula, pudiéndose afirmar que son estadísticamente significativos con un nivel de significación de 0,05; $\chi^2_{0,05;9} = 3{,}325$

El valor de lambda de Wilks de la tabla anterior es próximo a cero por lo que indica que no existe solapamiento entre los grupos. Además, el valor transformado de lambda es moderadamente alto y su nivel crítico es 0.0000 lo que reafirma que no consta que haya solapamiento entre los grupos.

Funciones en los centroides de los grupos

En esta siguiente tabla se muestra la ubicación de los centroides de los cuatro grupos para cada función discriminante.

Se observa que la primera función discrimina claramente al grupo primero (su valor es negativo mientras que los demás grupos son positivos) del resto de grupos. La segunda función hace lo propio con el grupo cuarto ya que discrimina con respecto al 2 y 3 (cuyos valores son negativos).

Por último, para la función tercera se distinguiría entre el grupo 2 del resto.

Según la tabla 25, se podría deducir que los grupos en los que va a haber un error mayor serán en aquellos casos con una calificación 2 o 3. Habrá casos en los que se clasifiquen en el grupo 2 y pertenezcan al 3 y viceversa, pertenezcan al 2 y se clasifiquen en el 3. En los grupos 1 y 4 el error será bastante menor.

Tabla 25:
Funciones en los centroides de los grupos.

Calificaciones	Función		
	1	2	3
1	-1,709	0,207	0,049
2	0,049	-0,240	-0,225
3	1,115	-0,321	0,135
4	2,989	0,884	-0,046

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Estos centroides se mostrarán en una gráfica más adelante donde se recogen todos los casos analizados y servirá para analizar visualmente las diferentes regiones y en cuáles de estas regiones habrá más dificultad para asignar casos al grupo correcto.

Lambda de Wilks

Posteriormente se muestran dos tablas (26 y 27) que recogen información el número de variables dentro del modelo en cada paso, su lambda y el estadístico de lambda.

Aquí se observa cómo a medida que se van incorporando más variables al modelo en cada paso, los valores de lambda de Wilks y del estadístico F asociado a ella van disminuyendo.

Tabla 26:
Lambda de Wilks (II).

Paso	Número de variables	Lambda	gl1	gl2	gl3
1	1	0,410	1	3	62
2	2	0,334	2	3	62
3	3	0,253	3	3	62

Tabla 27:
Lambda de Wilks (III).

	F exacta			
Paso	Estadístico	gl1	gl2	Sig.
1	29,732	3	62,000	0,000
2	14,841	6	122,000	0,000
3				

Coeficientes estandarizados de las funciones discriminantes canónicas

Para describir las funciones discriminantes canónicas se usan los coeficientes estandarizados. Para el presente Proyecto Fin de Carrera, éstas podrían ser sus funciones discriminantes canónicas:

$$D_1 = w_{11} \cdot X_1 + w_{12} \cdot X_2 + w_{13} \cdot X_3$$
$$D_2 = w_{21} \cdot X_1 + w_{22} \cdot X_2 + w_{23} \cdot X_3$$
$$D_3 = w_{31} \cdot X_1 + w_{32} \cdot X_2 + w_{33} \cdot X_3$$
$$i = 1,2,3$$

Estos coeficientes aparecen cuando se tipifican o estandarizan cada una de las variables clasificadoras para que tengan media 0 y desviación típica 1. De esta forma se evitan los problemas de escala que pudieran existir entre las variables y, además, la magnitud de los coeficientes estandarizados son un indicador de la importancia que tiene cada variable en el cálculo de la función discriminante.

La tabla 28 que se presenta a continuación contiene tres columnas, una para cada función discriminante. Las funciones se muestran ordenadas según sus autovalores, siendo siempre la primera función aquella con un poder discriminatorio mayor.

Tabla 28:
Coeficientes estandarizados de las funciones discriminantes canónicas.

	Función		
	1	**2**	**3**
Deuda/PIB	0,817	-0,156	-0,577
Carga Financ. / Ing. C.	0,612	-0,073	0,880
Importaciones % PIB	0,516	0,935	0,198

$$D_1 = 0{,}817 \cdot X_4 + 0{,}612 \cdot X_8 + 0{,}516 \cdot X_{17}$$
$$D_2 = -0{,}156 \cdot X_4 - 0{,}073 \cdot X_8 + 0{,}935 \cdot X_{17}$$
$$D_3 = -0{,}577 \cdot X_4 + 0{,}880 \cdot X_8 + 0{,}198 \cdot X_{17}$$

Para
$$i = 1{,}2{,}3$$

Contiene tres columnas, una por cada función discriminante. Las funciones se presentan siempre según su poder discriminante (de mayor a menor). Se puede observar que la segunda función discrimina entre las variables Deuda/PIB y Carga Financ. / Ing. C. por un lado, y la variable Importaciones % PIB por otro. Según el valor que tienen los coeficientes en la segunda función se puede afirmar que la variable Importaciones % PIB tiene una mayor importancia a la hora de predecir el grupo de pertenencia de una CCAA.

<u>Coeficientes de las funciones canónicas discriminantes</u>

A continuación, se presentan los coeficientes de las funciones canónicas discriminantes. Estos coeficientes dependen de la métrica original de las variables discriminantes, por lo que estos coeficientes son menos "preferibles" que los estandarizados ya que pueden conducir a error en ciertas ocasiones.

Tabla 29:
Coeficientes de las funciones canónicas discriminantes.

	Función		
	1	2	3
Deuda/PIB	17,357	-3,316	-12,268
Carga Financ. / Ing. C.	12,871	-1,545	18,512
Importaciones % PIB	0,061	0,111	0,024
(Constante)	-6,811	-1,417	-2,626

Coeficientes no tipificados

Quedarían las tres funciones de la siguiente manera:

$$D_1 = 17,357 \cdot X_4 + 12,871 \cdot X_8 + 0,061 \cdot X_{17} - 6,811$$

$$D_2 = -3,316 \cdot X_4 - 1,545 \cdot X_8 + 0,111 \cdot X_{17} - 1,417$$

$$D_3 = -12,268 \cdot X_4 + 18,512 \cdot X_8 + 0,024 \cdot X_{17} - 2,626$$

Matriz de Estructura

Por otro lado, es conveniente conocer cuáles son las variables que tienen mayor poder discriminante para así clasificar a un individuo en uno de los grupos (Grupo 1, 2, 3 y 4). Una forma de medir ese poder discriminante es calculando el coeficiente de correlación entre cada una de las variables y la función discriminante.

Esta es precisamente la información que se da en la tabla 30, Matriz de estructura. En este caso, la correlación de la función discriminante con la variable Deuda/PIB (0,770) es mayor en valor absoluto a las demás variables.

Las comparaciones deben hacerse siempre en valor absoluto. Las variables aparecen ordenadas de acuerdo con el valor absoluto de los coeficientes de correlación.

La matriz de estructura aporta los coeficientes de correlación entre las variables independientes y las puntuaciones discriminantes de cada función. Tal como se muestra al final de la tabla 30, los coeficientes seguidos de asterisco indican que esa variable es la que más correlaciona a esa función (que no quiere decir que sea la función que más discrimina a esa variable).

Si la colinealidad es alta (existe alta correlación entre las variables independientes), los coeficientes de la matriz pueden ser bastante distintos a los coeficientes estandarizados.

La primera función correlaciona con la variable Deuda/PIB, la variable T. Anual Crec. Gast. C. y la variable Ingr. C. / Ingr. T., la segunda función con la variable Importaciones % PIB mientras que la función tercera lo hace con las variables Carga Financ. / Ing. C., (Deuda – Ingr. C.) / PIB, Importaciones % PIB y Nº Matricul. / Nº Habit..

Tabla 30:
Matriz de estructura.

	Función		
	1	**2**	**3**
T. Anual Crec. Gast. C.	0,770*	-0,306	-0,561
Deuda/PIB[a]	0,230*	0,045	0,078
(Deuda – Ingr. C.) / PIB[a]	-0,173*	-0,031	0,167
Deuda / Ingresos C.	0,165	0,986*	-0,033
Carga Financ. / Ingr. C.	0,468	-0,423	0,776*
Ingr.C. / Ingr.T.[a]	0,544	0,008	-0,683*
Nº Matricul. / Nº Habit.[a]	0,592	-0,097	-0,665*
Importaciones % PIB[a]	0,014	0,022	-0,282*

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.

*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

a. Esta variable no se emplea en el análisis.

Coeficientes de la función de clasificación

Seguidamente se exponen en la tabla 31 los coeficientes de las funciones de clasificación de Fisher. Estos coeficientes se utilizan solamente para la clasificación. Se obtienen cuatro funciones de clasificación (una por cada grupo de Calificación).

Tabla 31:
Coeficientes de la función de clasificación.

	Calificaciones			
	1	**2**	**3**	**4**
Deuda/PIB	50,349	85,722	100,059	130,812
Carga Financ. / Ing. C.	112,682	130,922	151,433	170,348
Importaciones % PIB	0,557	0,609	0,674	0,919
(Constante)	-19,300	-28,494	-37,208	-55,384

Funciones discriminantes lineales de Fisher

Las cuatro funciones discriminantes lineales de Fisher son:

$$\begin{cases} F_I = 50{,}349 \cdot V_4 + 112{,}682 \cdot V_8 + 0{,}557 \cdot V_{17} - 19{,}300 \\ F_{II} = 85{,}722 \cdot V_4 + 130{,}922 \cdot V_8 + 0{,}609 \cdot V_{17} - 28{,}494 \\ F_{III} = 100{,}059 \cdot V_4 + 151{,}433 \cdot V_8 + 0{,}674 \cdot V_{17} - 37{,}208 \\ F_{IV} = 130{,}812 \cdot V_4 + 170{,}348 \cdot V_8 + 0{,}919 \cdot V_{17} - 55{,}384 \end{cases}$$

Analíticamente, se calcula el valor de las cuatro funciones para un sujeto determinado. Dicho sujeto se clasificará en aquel grupo donde la puntuación sea mayor.

Aquí se muestra un ejemplo. Se escoge el caso de la CCAA de Andalucía en el año 2014, la cual está clasificada en el grupo 4:

Andalucía 2014 – Calificación 4

- Variable Deuda/PIB = 0,2223
- Variable Carga Financ. / Ing. C. = 0,339
- Variable Importaciones % PIB = 22,88

$$\begin{cases} F_I = 50{,}349 \cdot 0{,}2223 + 112{,}682 \cdot 0{,}339 + 0{,}557 \cdot 22{,}88 - 19{,}300 \\ F_{II} = 85{,}722 \cdot 0{,}2223 + 130{,}922 \cdot 0{,}339 + 0{,}609 \cdot 22{,}88 - 28{,}494 \\ F_{III} = 100{,}059 \cdot 0{,}2223 + 151{,}433 \cdot 0{,}339 + 0{,}674 \cdot 22{,}88 - 37{,}208 \\ F_{IV} = 130{,}812 \cdot 0{,}2223 + 170{,}348 \cdot 0{,}339 + 0{,}919 \cdot 22{,}88 - 55{,}384 \end{cases}$$

$$\begin{cases} F_I = 954{,}255 \\ F_{II} = 1148{,}720 \\ F_{III} = 1325{,}300 \\ F_{IV} = 1512{,}466 \end{cases}$$

Según el resultado obtenido para las cuatro funciones, se clasifica en el grupo 4 ya que $F_{IV} = 1512{,}466$ es la puntuación más alta conseguida.

En la práctica, con el programa IBM SPSS Statistics no utiliza estos coeficientes para el cálculo de la clasificación.

Clasificación de los casos (CCAA/Año)

Seguidamente se recoge en la tabla 32 la información de todos los casos a los que se les ha realizado el Análisis Discriminante Múltiple. Se muestran el número del caso, el grupo al que pertenece, el grupo pronosticado, probabilidades, puntuaciones de las funciones discriminantes, etc.

Se han señalado un caso en verde y otro en rojo un ejemplo donde la predicción fue satisfactoria y otro en la que fue errónea. En el caso 2, el grupo real y pronosticado es 3. En el caso 42, sin embargo, el grupo real de pertenencia es 2 y fue clasificado en el grupo 3.

Tabla 32:
Puntuaciones discriminantes.

							Puntuaciones discriminantes		
Número de caso	Grupo real	Grupo pronosticado	P(D>d | G=g) p	P(D>d | G=g) gl	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1	Función 2	Función 3
1	4	4	0,343	3	0,652	3,330	2,816	-0,131	1,461
2	3	3	0,071	3	0,635	7,017	1,990	0,155	2,589
3	2	3**	0,181	3	0,536	4,875	0,273	0,140	2,123
4	1	2**	0,281	3	0,410	3,824	-0,518	-0,162	1,644
5	1	1	0,299	3	0,812	3,670	-1,687	-0,831	1,659
6	2	3**	0,532	3	0,443	2,202	1,028	0,704	-0,934
7	2	2	0,964	3	0,539	0,278	0,392	0,144	-0,340
8	1	2**	0,920	3	0,540	0,493	-0,476	0,221	-0,291
9	1	1	0,906	3	0,736	0,557	-1,283	0,787	-0,149
10	1	1	0,436	3	0,990	2,727	-3,341	0,358	-0,156
11	1	1	0,046	3	0,998	8,002	-4,286	1,137	-0,654
12	2	2	0,994	3	0,596	0,083	0,048	-0,488	-0,371
13	2	2	0,945	3	0,551	0,378	-0,157	-0,431	0,322
14	2	2	0,775	3	0,504	1,107	-0,251	-0,164	0,780
15	1	1	0,905	3	0,944	0,563	-2,436	0,033	-0,010
16	1	1	0,112	3	0,996	5,983	-4,070	-0,429	-0,019
17	3	3	0,341	3	0,558	3,348	0,326	-1,372	1,408
18	3	3	0,225	3	0,585	4,358	0,318	-1,288	1,805
19	2	2	0,635	3	0,626	1,709	0,413	-0,679	-1,402
20	2	2	0,773	3	0,642	1,118	0,141	-0,916	-1,033
21	1	2**	0,841	3	0,654	0,835	-0,294	-0,792	-0,868
22	1	1	0,711	3	0,947	1,378	-2,613	-0,223	-0,564
23	3	4**	0,146	3	0,674	5,383	3,321	-1,193	-1,024
24	3	3	0,481	3	0,648	2,469	2,414	-1,098	-0,288
25	2	3**	0,825	3	0,517	0,900	0,657	-1,117	-0,104
26	1	2**	0,908	3	0,605	0,549	-0,243	-0,918	-0,164
27	1	1	0,676	3	0,642	1,528	-1,482	-0,987	-0,175
28	3	2**	0,860	3	0,493	0,758	0,690	0,231	-0,579
29	3	2**	0,968	3	0,501	0,254	0,409	-0,054	0,075
30	3	2**	0,990	3	0,569	0,118	-0,129	-0,003	-0,053
31	2	1**	0,957	3	0,673	0,314	-1,323	0,049	-0,325

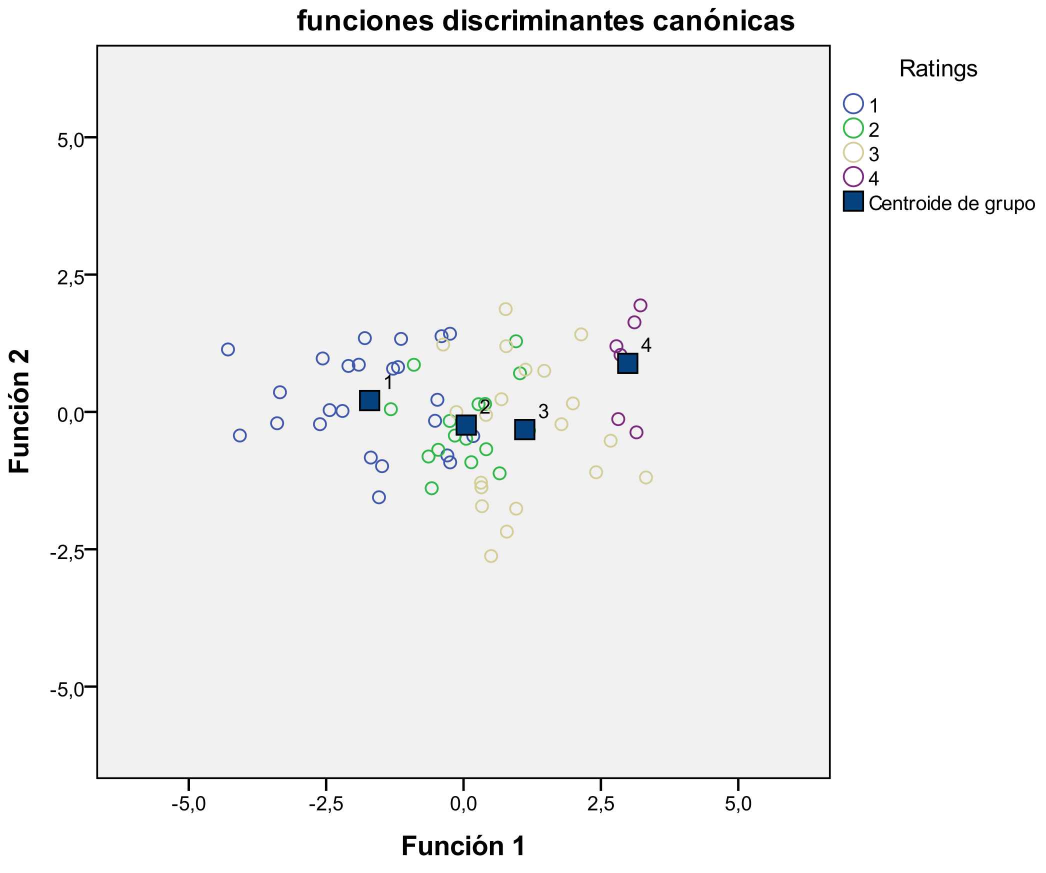
32	1	1	0,923	3	0,909	0,481	-2,203	0,017	-0,400
33	1	1	0,382	3	0,988	3,061	-3,392	-0,206	-0,191
34	4	4	0,510	3	0,928	2,312	2,858	1,038	-1,553
35	4	4	0,976	3	0,919	0,209	2,782	1,196	-0,307
36	3	4**	0,682	3	0,806	1,503	2,142	1,411	-0,758
37	2	2	0,308	3	0,397	3,603	0,955	1,286	-0,897
38	1	2**	0,412	3	0,507	2,871	-0,246	1,423	-0,358
39	4	4	0,506	3	0,795	2,333	3,148	-0,372	-0,902
40	3	4**	0,504	3	0,548	2,343	2,678	-0,525	-0,556
41	3	3	0,904	3	0,609	0,567	1,782	-0,224	-0,200
42	2	3**	0,964	3	0,587	0,278	1,192	-0,340	-0,386
43	1	2**	0,987	3	0,544	0,137	0,177	-0,440	0,060
44	3	3	0,535	3	0,657	2,185	0,958	-1,761	0,434
45	3	3	0,152	3	0,602	5,291	0,336	-1,716	1,790
46	2	2	0,327	3	0,510	3,452	-0,579	-1,389	1,093
47	1	1	0,280	3	0,663	3,836	-1,538	-1,554	0,889
48	3	3	0,650	3	0,465	1,641	1,469	0,748	-0,475
49	3	3	0,753	3	0,516	1,201	1,128	0,770	0,036
50	3	3	0,487	3	0,460	2,435	0,778	1,197	0,265
51	1	1	0,766	3	0,678	1,144	-1,192	0,816	-0,662
52	1	1	0,674	3	0,911	1,536	-2,094	0,837	-0,946
53	3	2**	0,096	3	0,546	6,353	0,788	-2,178	-1,658
54	3	2**	0,080	3	0,566	6,751	0,502	-2,623	-1,157
55	2	2	0,681	3	0,666	1,506	-0,459	-0,690	-1,248
56	2	2	0,843	3	0,585	0,828	-0,637	-0,811	-0,400
57	1	1	0,621	3	0,758	1,771	-1,137	1,329	0,478
58	4	4	0,877	3	0,971	0,682	3,110	1,631	0,283
59	4	4	0,617	3	0,982	1,790	3,220	1,939	0,743
60	3	3	0,065	3	0,474	7,226	0,768	1,871	1,651
61	1	1	0,090	3	0,480	6,483	-0,400	1,378	1,892
62	1	1	0,291	3	0,937	3,741	-1,905	0,860	1,859
63	3	2**	0,481	3	0,514	2,470	-0,371	1,226	-0,605
64	2	1**	0,716	3	0,565	1,357	-0,903	0,858	-0,484
65	1	1	0,605	3	0,890	1,846	-1,796	1,343	-0,691
66	1	1	0,489	3	0,961	2,427	-2,563	0,973	-1,005
1	4	4	0,343	3	0,652	3,330	2,816	-0,131	1,461
2	3	3	0,071	3	0,635	7,017	1,990	0,155	2,589
3	2	3**	0,181	3	0,536	4,875	0,273	0,140	2,123
4	1	2**	0,281	3	0,410	3,824	-0,518	-0,162	1,644
5	1	1	0,299	3	0,812	3,670	-1,687	-0,831	1,659

6	2	3**	0,532	3	0,443	2,202	1,028	0,704	-0,934
7	2	2	0,964	3	0,539	0,278	0,392	0,144	-0,340
8	1	2**	0,920	3	0,540	0,493	-0,476	0,221	-0,291
9	1	1	0,906	3	0,736	0,557	-1,283	0,787	-0,149
10	1	1	0,436	3	0,990	2,727	-3,341	0,358	-0,156
11	1	1	0,046	3	0,998	8,002	-4,286	1,137	-0,654

Centroides

A continuación, en la figura 8 se muestra de los diferentes casos con los centroides de los cuatro grupos:

Figura 8:
Centroides de grupo de las funciones discriminantes canónicas.



Se puede observar que los grupos 1 y 4 se dividen claramente del resto (sus centroides están también apartados de los demás), mientras que los grupos 2 y 3 están más próximos, lo que repercutirá en la capacidad del poder predictivo en esa zona sea peor (hallándose casos mal clasificados en los que se debía haber asignado al grupo 2 y se clasifica en el 3 y viceversa, debería haberse clasificado en el grupo 3 y se le asigna al 2).

Clasificación

En la tabla 33 se obtienen los resultados de la clasificación. Como se apuntaba anteriormente, los casos pertenecientes a los grupos 1 y 4 tienen una clasificación bastante mejor que los casos pertenecientes a los grupos 2 y 3.

Los casos pertenecientes al grupo 1 se clasifican bien en el 75% y el grupo 4 el 100%. Sin embargo, se consiguen un tanto por ciento menor para los grupos 2 (62,5%) y sobre todo para el grupo 3 (55%).

Tabla 33:
Resultados de la clasificación[a] (I).

		Calificaciones	Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	18	6	0	0	24
		2	2	10	4	0	16
		3	0	6	11	3	20
		4	0	0	0	6	6
	%	1	75,0	25,0	0,0	0,0	100,0
		2	12,5	62,5	25,0	0,0	100,0
		3	0,0	30,0	55,0	15,0	100,0
		4	0,0	0,0	0,0	100,0	100,0

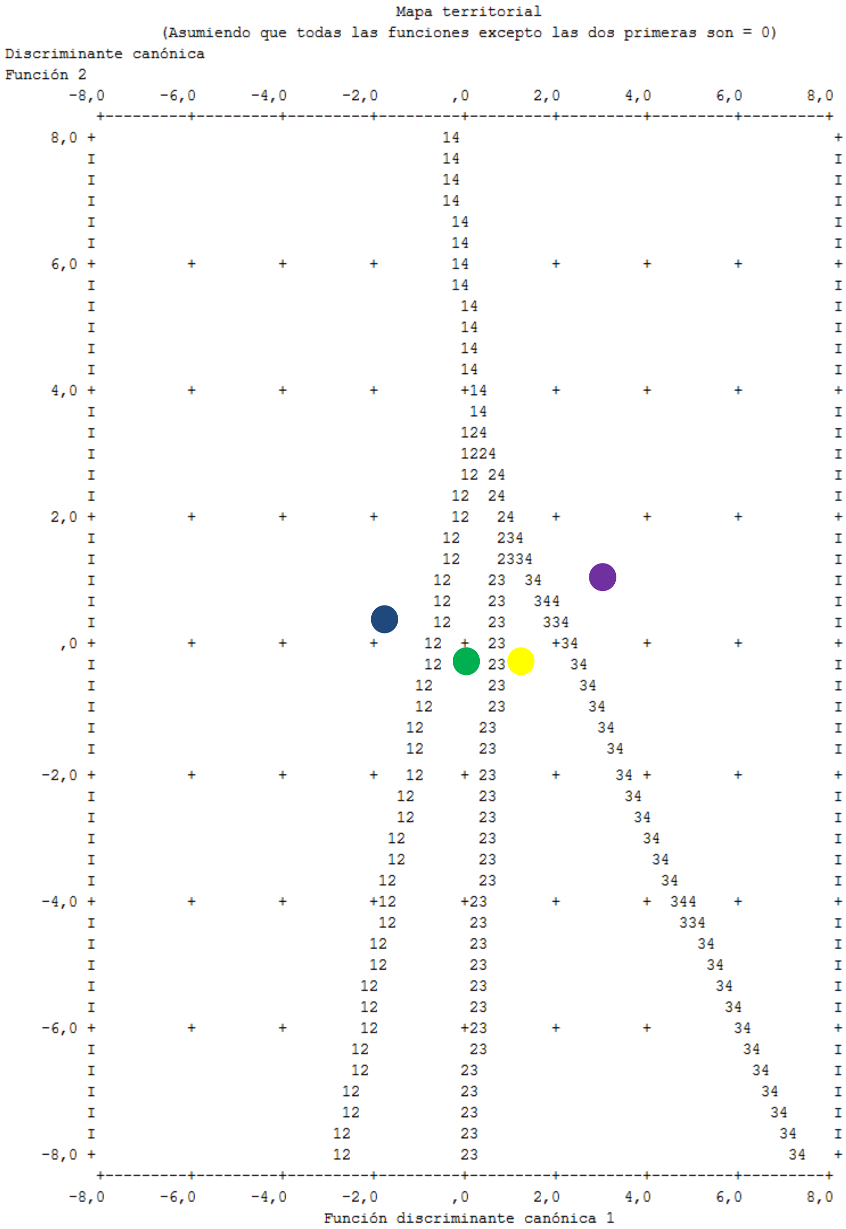
a. Clasificados correctamente el 68,2% de los casos agrupados originales.

Con el método Análisis Discriminante Múltiple se consigue clasificar correctamente el 68,2% de las CCAA/Año.

Mapa Territorial

Se presenta el Mapa Territorial en la Figura 9. Cada eje está representado por las funciones discriminantes 1 y 2 y donde los números 1, 2, 3 y 4 delimitan el espacio de cada grupo. En el mapa se observa sin ninguna dificultad que los grupos 2 y 3 son, como se había confirmado anteriormente, los grupos en los que la clasificación correcta es más dificultosa. Los centroides de estos grupos (2 y 3) están prácticamente alineados por lo que se puede decir que la función 2 no hace una buena discriminación entre estos grupos (debido también a que anteriormente se había detectado que no era significativa esta segunda función). También destacar que las regiones para los grupos 2 y 3 son especialmente más reducidas que las regiones de los grupos 1 y 4.

Figura 9:
Mapa territorial.



Centroides del mapa territorial:

- Centroide Grupo 1
- Centroide Grupo 2
- Centroide Grupo 3
- Centroide Grupo 4

Clasificación según el tamaño de los grupos

Al hacer el análisis con IBM SPSS Statistics, se eligió la opción de que todos los grupos eran iguales. Sin embargo, es posible pensar que dado el caso que se está analizando en este proyecto (las calificaciones asignadas por las agencias no cambian bruscamente y mucho menos en periodos cortos de tiempo, puesto que la economía se irá o bien recuperando o bien emporando progresivamente), es posible que el número de casos asignados del grupo 1 seis meses después sea prácticamente el mismo, al igual que para los restantes grupos.

Por ello, se ha elegido la opción del cálculo de la clasificación según el tamaño de los grupos. De esta forma, se ha pasado de un 74% de casos correctamente clasificados a un 6% más, es decir, el 74,2% de los casos estarán correctamente clasificados.

Se muestra la tabla proporcionada por la herramienta que recoge esta información. Para el grupo 1 habrá un acierto del 79,2%, para el grupo 2 un 56,3%, para el grupo 3 el 80 y, por último, para el grupo 4 el 83,3%.
Se observa que dos de los grupos han aumentado (1 y 3) mientras que los grupos 2 y 4 ha disminuido el porcentaje, y que han bajado en su acierto.

Tabla 34:
Resultados de la clasificación[a] (II).

		Calificaciones	Grupo de pertenencia pronosticado				Total
			1	2	3	4	
Original	Recuento	1	19	5	0	0	24
		2	2	9	5	0	16
		3	1	2	16	1	20
		4	0	0	1	5	6
	%	1	79,2	20,8	0,0	0,0	100,0
		2	12,5	56,3	31,3	0,0	100,0
		3	5,0	10,0	80,0	5,0	100,0
		4	0,0	0,0	16,7	83,3	100,0

a. Clasificados correctamente el 74,2% de los casos agrupados originales.

7 Conclusiones

El presente Proyecto Fin de Carrera ha perseguido fundamentalmente dos objetivos. Por un lado, conocer la influencia de una serie de variables económico-financieras independientes sobre una variable dependiente, llamada variable categórica y, por otro lado, predecir la clasificación de nuevos individuos en los distintos grupos posibles de la variable categórica. Para lograrlo, se utilizó la técnica estadística de clasificación supervisada, Análisis Discriminante. Dicha técnica se aplicó a las calificaciones de deuda realizadas a las diferentes Comunidades Autónomas españolas por las tres más importantes agencias a nivel internacional, Standard & Poor´s, Moody´s y Fitch. De esta manera, se hallaron aquellas variables independientes con un poder de discriminación mayor, esto es, variables que proporcionaron los centros de los grupos muy distintos entre sí y muy homogéneos dentro de sí.

La variable dependiente que se incluyó en el análisis contenía cuatro categorías (Grupo 1, 2, 3 y 4). Además, una serie de variables independientes relacionadas con la economía de las CCAA, exactamente un total de ocho variables (tras la eliminación de doce de ellas en los análisis previos de los datos, debido a que no cumplían con las hipótesis de normalidad, homocedasticidad o independencia). Se analizaron un total de 66 casos de los 87 inicialmente propuestos. Ésto se debió a que no se encontró información de las variables para ciertos años de varias CCAA.

Referente a conocer la influencia que tienen las ocho variables utilizadas en el análisis, se puede concluir diciendo que las variables "Importaciones % PIB", "Carga Financiera / Ingresos Corrientes" y "Deuda/PIB" en su conjunto son aquellas que poseen un poder de discriminación mayor. Estas tres variables van a ser las incluidas en las funciones discriminantes obtenidas del análisis.

Relativo a la función predictiva conseguida a través del Análisis Discriminante, se puede afirmar que es muy satisfactoria para CCAA con una clasificación de grupo 1 y 4 con un porcentaje acertado muy elevado. De esta manera, permiten prever sus calificaciones de deuda pública, así como aproximarlas cuando éstas no existan. No obstante, las CCAA con una calificación asignada al grupo 2 o 3, el error es mayor, provocando de esta forma que el resultado de la predicción sea más incierto. La función de discriminación primera es la que obtiene unos mejores resultados de discriminación, con un mayor porcentaje de acierto que las funciones segunda y tercera.

Con todo ello, los resultados obtenidos deben tomarse con prudencia ya que debido a la dificultad en la recogida de datos no se pudieron obtener toda la información que se hubiese deseado, es decir, información de todos los años para cada Comunidad Autónoma, que hubiera contribuido a la obtención de unos mejores resultados o conclusiones. Del mismo modo, no se ha podido recoger información sobre años posteriores a los analizados (2015 o 2016), por lo que no ha sido posible realizar una estimación sobre las calificaciones en dichos años.

Este trabajo puede servir de base para otras investigaciones que persigan como objetivo la calificación de su deuda ampliada a grupos de países y empresas privadas de diferente índole.

8 Bibliografía

- Anuario Económico de España. (2011). Selección de indicadores. Área de Estudios y Análisis Económico de "la Caixa". Descargado el 18 de diciembre de 2015 en http://www.joanbcasas.cat/wp-content/uploads/2012/07/AnuarioEconomicoEspa%C3%B1a2011_lacaixa.pdf

- BROTO, C. & MOLINA, L. (2014). Sovereign Ratings and their Asymmetric Response to Fundamentals. Consultado el 3 de octubre de 2016 en la World Wide Web: http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/14/Fich/dt1428e.pdf

- Diario Oficial de la Unión Europea (2016, 7 de octubre). Reglamento (UE) n 575/2013 del Parlamento Europeo y del Consejo, 136 (1-3). Consultado el 5 de octubre de 2016, en https://www.boe.es/doue/2016/275/L00003-00018.pdf.

- Indicadores Económicos y Socio-Demográficos. Diario Expansión, Datos Macro. Consultado en diciembre de 2015 y octubre de 2016, en http://www.datosmacro.com/espana-comunidades-autonomas

- LÓPEZ PASCUAL, J. (1996). El Rating y las Agencias de Calificación, Madrid: DYKINSON. Descargado el 2 de octubre de 2016 de http://www.joaquinlopezpascual.com/libro-el-rating-y-las-agencias-de-calificacion.

- LOSADA LÓPEZ, R. (2009).). Agencias de rating: Hacia una nueva regulación. Comisión Nacional del Mercado de Valores. Consultado el 3 de octubre de 2016 de la World Wide Web: http://www.cnmv.es/docportal/publicaciones/monografias/mon2009_34.pdf.

- Manual del Usuario del sistema básico de IBM SPSS Statistics 20. (2011). Consultado en la World Wide Web: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/es/client/Manuals/IBM_SPSS_Statistics_Core_System_Users_Guide.pdf.

- Ministerio de Economía y Competitividad del Gobierno de España. (2014). Indicadores e Informes Macroeconómicos. Descargado el 15 de diciembre de 2015, en http://www.mineco.gob.es/portal/site/mineco/menuitem.b6c80362d9873d0a91b0240e026041a0/?vgnextoid=bbab7ed41594c310VgnVCM1000001d04140aRCRD.

- Ministerio de Hacienda y Administraciones Públicas del Gobierno de España. (2014). Las Haciendas Autonómicas en cifras. Descargado el 15 de diciembre de 2015, en http://www.minhafp.gob.es/es-ES/CDI/Paginas/SistemasFinanciacionDeuda/InformacionCCAAs/haciendas%202005.aspx.

- Ministerio de Hacienda y Administraciones Públicas del Gobierno de España. (2013). Presupuestos Generales de las Comunidades Autónomas. Descargado el 15 de diciembre de 2015, en https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&ved=0ahUKEwix3YrKucfRAhWBVhQKHUvfD40QFgg7MAQ&url=https%3A%2F%2Fserviciostelematicosext.minhap.gob.es%2FSGCAL%2Fpublicacionpresupuestos%2Faspx%2FMenuRep_Portal.aspx%3FejerRE%3D2013&usg=AFQjCNGaccj40_vv5BFUdJ1k3lr8UTwkRA&sig2=RpN-EurbdVXWBgqyepTuyQ.

- PÉREZ, C. (2013). Análisis Multivariante de Datos. Aplicaciones con IBM SPSS, SAS Y STATGRAPHICS, Madrid: Garceta.

- Secretaría de Estado de Investigación, Desarrollo e Innovación del Gobierno de España. (2014). Indicadores del Sistema Español de Ciencia, Tecnología e Innovación. Descargado el 15 de diciembre de 2015, en http://www.idi.mineco.gob.es/portal/site/MICINN/menuitem.8ce192e94ba842bea3bc811001432ea0/?vgnextoid=403814f9bc6a9510VgnVCM1000001d04140aRCRD&vgnextchannel=c930444c25c73410VgnVCM1000001d04140aRCRD.

- URIEL, E. y ALDÁS, J. (2005). Análisis Multivariante Aplicado. Aplicaciones al Marketing, Investigación de Mercados, Economía, Dirección de Empresas y Turismo, Madrid: Thomson.

- TEMPLETON, G.F. (2011). "A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research," Communications of the AIS, 28 (4). Consultado el 10 de enero de 2016, en https://www.researchgate.net/publication/289265853_A_two-step_approach_for_transforming_continuous_variables_to_normal_Implications_and_recommendations_for_IS_research.

ANEXO - TABLAS

Tabla correspondiente a la prueba de normalidad realizada a las 20 variables seleccionadas para el Análisis Discriminante:

Tabla 1-Anexo:
Pruebas de normalidad. 20 variables.

	Calificaciones	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Superávit	1	0,124	30	0,200*	0,943	30	0,112
	2	0,277	19	0,000	0,810	19	0,002
	3	0,365	23	0,000	0,646	23	0,000
	4	0,252	8	0,144	0,800	8	0,029
T. Anual Crec. Ingr. C.	1	0,270	30	0,000	0,597	30	0,000
	2	0,186	19	0,084	0,930	19	0,175
	3	0,193	23	0,026	0,798	23	0,000
	4	0,164	8	0,200*	0,962	8	0,833
T. Anual Crec. Gast. C.	1	0,192	30	0,006	0,793	30	0,000
	2	0,161	19	0,200*	0,917	19	0,102
	3	0,200	23	0,018	0,942	23	0,197
	4	0,193	8	0,200*	0,941	8	0,623
Deuda/PIB	1	0,153	30	0,070	0,924	30	0,035
	2	0,150	19	0,200*	0,938	19	0,239
	3	0,231	23	0,003	0,884	23	0,012
	4	0,173	8	0,200*	0,948	8	0,694
(Deuda – Ingr. C.) / PIB	1	0,097	30	0,200*	0,975	30	0,679
	2	0,114	19	0,200*	0,953	19	0,450
	3	0,173	23	0,074	0,930	23	0,110
	4	0,166	8	0,200*	0,932	8	0,539
PIB / Habit. año media	1	0,221	30	0,001	0,920	30	0,026
	2	0,192	19	0,063	0,949	19	0,377
	3	0,187	23	0,037	0,915	23	0,051
	4	0,319	8	0,016	0,786	8	0,020
Deuda / Ingresos Corr.	1	0,147	30	0,098	0,896	30	0,007
	2	0,128	19	0,200*	0,968	19	0,732
	3	0,204	23	0,014	0,894	23	0,019
	4	0,172	8	0,200*	0,931	8	0,526
Carga Financ. / Ingr. C.	1	0,111	30	0,200*	0,971	30	0,577
	2	0,174	19	0,132	0,858	19	0,009

	3	0,172	23	0,075	0,934	23	0,131
	4	0,169	8	0,200*	0,973	8	0,923
Ingr. Financieros/Ingr. Totales	1	0,262	30	0,000	0,815	30	0,000
	2	0,281	19	0,000	0,753	19	0,000
	3	0,188	23	0,035	0,878	23	0,009
	4	0,287	8	0,051	0,789	8	0,022
Ingr. Corr./Ingr. Totales	1	0,158	30	0,055	0,953	30	0,202
	2	0,195	19	0,055	0,870	19	0,014
	3	0,135	23	0,200*	0,915	23	0,052
	4	0,182	8	0,200*	0,955	8	0,764
Ingr. Impuestos / Ingr. Corrientes	1	0,158	30	0,054	0,944	30	0,115
	2	0,198	19	0,048	0,880	19	0,021
	3	0,134	23	0,200*	0,931	23	0,117
	4	0,226	8	0,200*	0,938	8	0,595
Ingr. Imp.T. / Ingr. T.	1	0,126	30	0,200*	0,947	30	0,138
	2	0,175	19	0,128	0,890	19	0,032
	3	0,138	23	0,200*	0,931	23	0,117
	4	0,236	8	0,200*	0,955	8	0,760
Gasto Per Cápita	1	0,179	30	0,016	0,803	30	0,000
	2	0,133	19	0,200*	0,917	19	0,099
	3	0,146	23	0,200*	0,946	23	0,243
	4	0,182	8	0,200*	0,950	8	0,712
Ingresos C. / Ingresos T.	1	0,115	30	0,200*	0,943	30	0,112
	2	0,153	19	0,200*	0,940	19	0,259
	3	0,136	23	0,200*	0,955	23	0,375
	4	0,180	8	0,200*	0,936	8	0,575
Ing. C. (6, 7 y 8) / Ing. T.	1	0,127	30	0,200*	0,947	30	0,139
	2	0,205	19	0,035	0,897	19	0,042
	3	0,145	23	0,200*	0,902	23	0,028
	4	0,184	8	0,200*	0,973	8	0,923
NºMatric. / Nº Habit.	1	0,216	30	0,001	0,838	30	0,000
	2	0,288	19	0,000	0,600	19	0,000
	3	0,149	23	0,200*	0,893	23	0,018
	4	0,189	8	0,200*	0,930	8	0,512
Importaciones % PIB	1	0,136	30	0,165	0,958	30	0,282
	2	0,128	19	0,200*	0,959	19	0,551
	3	0,123	23	0,200*	0,935	23	0,143
	4	0,239	8	0,200*	0,839	8	0,073
Exportaciones % PIB	1	0,080	30	0,200*	0,961	30	0,335
	2	0,169	19	0,154	0,956	19	0,492

	3	0,104	23	0,200*	0,951	23	0,313
	4	0,165	8	0,200*	0,970	8	0,897
Esperanza de vida	1	0,082	30	0,200*	0,981	30	0,853
	2	0,139	19	0,200*	0,985	19	0,986
	3	0,102	23	0,200*	0,973	23	0,766
	4	0,239	8	0,200	0,888	8	0,226
Balanza C.Exp. – Imp.	1	0,168	30	0,031	0,951	30	0,179
	2	0,155	19	0,200*	0,939	19	0,252
	3	0,135	23	0,200*	0,904	23	0,031
	4	0,184	8	0,200*	0,940	8	0,615

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

Prueba de normalidad tras la transformación de las variables.

Tabla correspondiente a la prueba de normalidad realizada a las 20 variables seleccionadas para el Análisis Discriminante después de su transformación:

Tabla 2-Anexo:
Pruebas de normalidad. Variables transformadas.

	Calificaciones	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Superávit	1	0,125	24	0,200*	0,924	24	0,073
	2	0,175	16	0,200*	0,931	16	0,257
	3	0,193	20	0,049	0,924	20	0,120
	4	0,302	6	0,093	0,813	6	0,077
T. Anual Crec. Ingr. C.	1	0,112	24	0,200*	0,971	24	0,694
	2	0,122	16	0,200*	0,980	16	0,962
	3	0,131	20	0,200*	0,942	20	0,263
	4	0,183	6	0,200*	0,945	6	0,703
T. Anual Crec. Gast. C.	1	0,130	24	0,200*	0,949	24	0,259
	2	0,179	16	0,182	0,938	16	0,329
	3	0,143	20	0,200*	0,962	20	0,576
	4	0,187	6	0,200*	0,963	6	0,840
Deuda/PIB	1	0,106	24	0,200*	0,976	24	0,821
	2	0,158	16	0,200*	0,916	16	0,143
	3	0,150	20	0,200*	0,950	20	0,368
	4	0,229	6	0,200*	0,881	6	0,272
(Deuda – Ingr. C.) / PIB	1	0,078	24	0,200*	0,982	24	0,931
	2	0,173	16	0,200*	0,916	16	0,145
	3	0,122	20	0,200*	0,959	20	0,523
	4	0,199	6	0,200*	0,953	6	0,763
PIB / Habit. año media	1	0,074	24	0,200*	0,980	24	0,887
	2	0,166	16	0,200*	0,957	16	0,610
	3	0,117	20	0,200*	0,977	20	0,890
	4	0,209	6	0,200*	0,908	6	0,422
Deuda / Ingresos Corr.	1	0,094	24	0,200*	0,967	24	0,600
	2	0,160	16	0,200*	0,926	16	0,209
	3	0,143	20	0,200*	0,958	20	0,508
	4	0,145	6	0,200*	0,976	6	0,929
Carga Financ. / Ingr. C.	1	0,106	24	0,200*	0,970	24	0,665
	2	0,129	16	0,200*	0,950	16	0,488
	3	0,136	20	0,200*	0,968	20	0,704
	4	0,217	6	0,200*	0,953	6	0,765

Ingr. Financieros/Ingr. Totales	1	0,091	24	0,200*	0,977	24	0,838
	2	0,180	16	0,177	0,877	16	0,035
	3	0,118	20	0,200*	0,965	20	0,644
	4	0,217	6	0,200*	0,858	6	0,184
Ingr. Corr./Ingr. Totales	1	0,179	24	0,045	0,942	24	0,181
	2	0,192	16	0,118	0,862	16	0,021
	3	0,095	20	0,200*	0,970	20	0,760
	4	0,253	6	0,200*	0,938	6	0,642
Ingr. Impuestos / Ingr. Corrientes	1	0,166	24	0,086	0,918	24	0,052
	2	0,158	16	0,200*	0,879	16	0,038
	3	0,130	20	0,200*	0,950	20	0,374
	4	0,251	6	0,200*	0,910	6	0,436
Ingr. Imp.T. / Ingr. T.	1	0,145	24	0,200*	0,931	24	0,103
	2	0,124	16	0,200*	0,906	16	0,099
	3	0,103	20	0,200*	0,959	20	0,533
	4	0,225	6	0,200*	0,971	6	0,896
Gasto Per Cápita	1	0,105	24	0,200*	0,974	24	0,773
	2	0,121	16	0,200*	0,948	16	0,459
	3	0,129	20	0,200*	0,955	20	0,454
	4	0,290	6	0,125	0,908	6	0,425
Ingresos C. / Ingresos T.	1	0,143	24	0,200*	0,940	24	0,165
	2	0,184	16	0,149	0,939	16	0,336
	3	0,139	20	0,200*	0,962	20	0,582
	4	0,198	6	0,200*	0,899	6	0,371
Ing. C. (6, 7 y 8) / Ing. T.	1	0,115	24	0,200*	0,980	24	0,892
	2	0,156	16	0,200*	0,913	16	0,129
	3	0,122	20	0,200*	0,944	20	0,291
	4	0,209	6	0,200*	0,953	6	0,768
NºMatric. / Nº Habit.	1	0,178	24	0,049	0,918	24	0,053
	2	0,143	16	0,200*	0,916	16	0,147
	3	0,097	20	0,200*	0,977	20	0,895
	4	0,237	6	0,200*	0,907	6	0,417
Importaciones % PIB	1	0,138	24	0,200*	0,964	24	0,524
	2	0,138	16	0,200*	0,965	16	0,745
	3	0,125	20	0,200*	0,965	20	0,641
	4	0,243	6	0,200*	0,884	6	0,290
Exportaciones % PIB	1	0,100	24	0,200*	0,968	24	0,618
	2	0,133	16	0,200*	0,982	16	0,979
	3	0,110	20	0,200*	0,965	20	0,655
	4	0,163	6	0,200*	0,967	6	0,868

Esperanza de vida	1	0,078	24	0,200*	0,984	24	0,962
	2	0,189	16	0,130	0,968	16	0,801
	3	0,102	20	0,200*	0,970	20	0,746
	4	0,260	6	0,200*	0,936	6	0,628
Balanza C.Exp. – Imp.	1	0,071	24	0,200*	0,994	24	1,000
	2	0,122	16	0,200*	0,983	16	0,982
	3	0,110	20	0,200*	0,965	20	0,640
	4	0,260	6	0,200*	0,869	6	0,222

a. Corrección de la significación de Lilliefors

*. Este es un límite inferior de la significación verdadera.

ANEXO – FIGURAS

Prueba de Normalidad

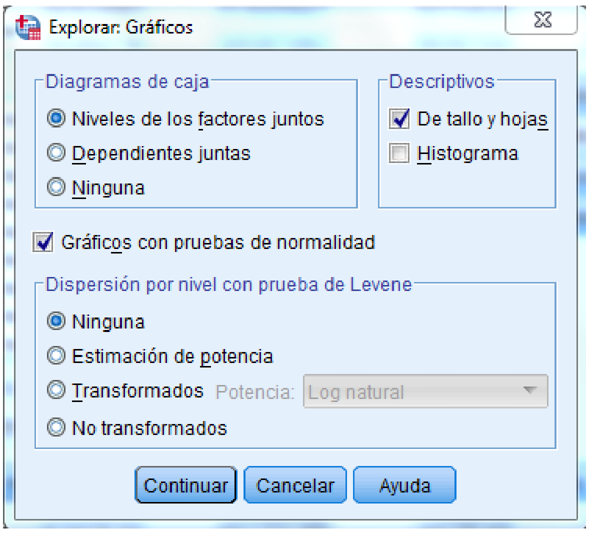
Para realizar la Prueba de Normalidad con IBM SPSS:

<u>Análisis – Explorar</u>

Introduciendo las variables dependientes, así como la variable independiente y seleccionando la prueba de normalidad se consigue verificar la normalidad de las variables.

Figura1:
SPSS Statistics. Normalidad (I).

Figura2:
SPSS Statistics. Normalidad (II).



Transformación de las variables

Se transforman los datos de todas las variables de la siguiente forma:

En el menú de IBM SPSS, Transformar – Asignar rangos a casos.

Figura 3:
Transformación (I).

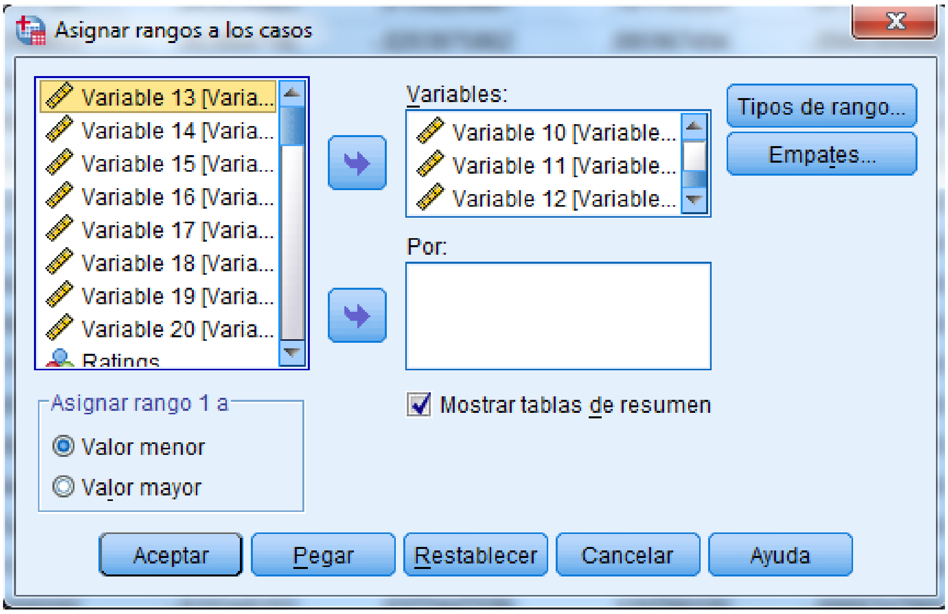
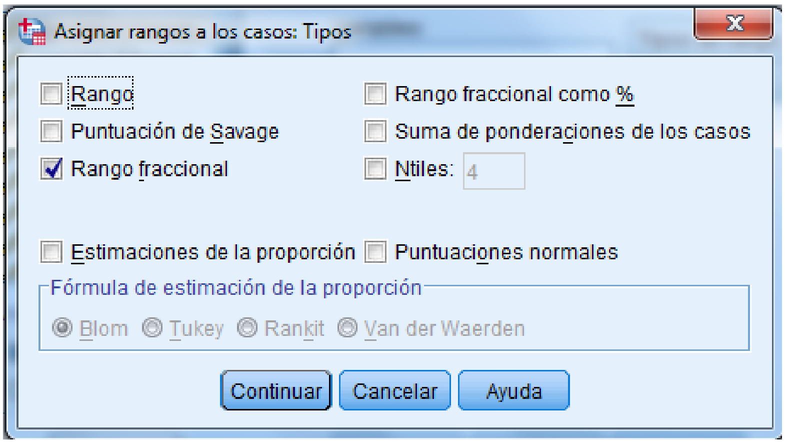
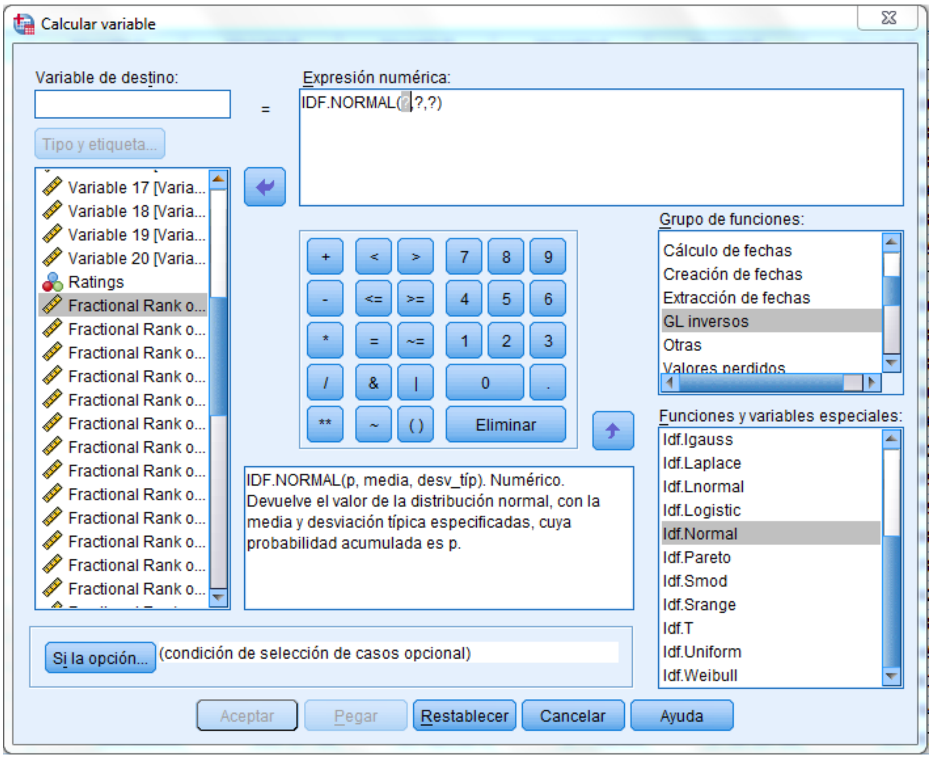


Figura4:
Transformación (II).



En el menú de IBM SPSS, Transformar – Calcular variable.

Figura5:
Transformación (III).



Prueba de Homocedasticidad

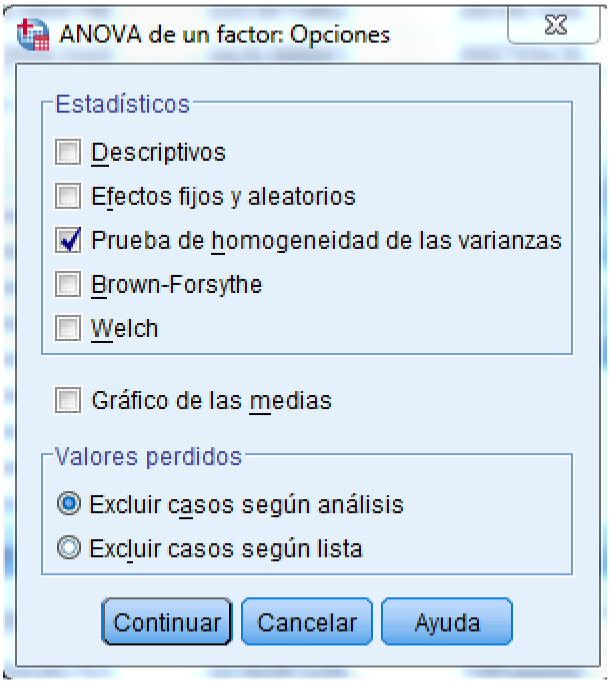
Para realizar la prueba de homocedasticidad de varianzas:

Analizar – Comparar medias – Anova de un factor

Se seleccionan las todas las variables excepto las excluidas por el requisito y se elige el factor.

Figura 6:
SPSS Statistics.Homocedasticidad (I).

Figura 7:
SPSS Statistics. Homocedasticidad (II).



Prueba de Independencia de Medias

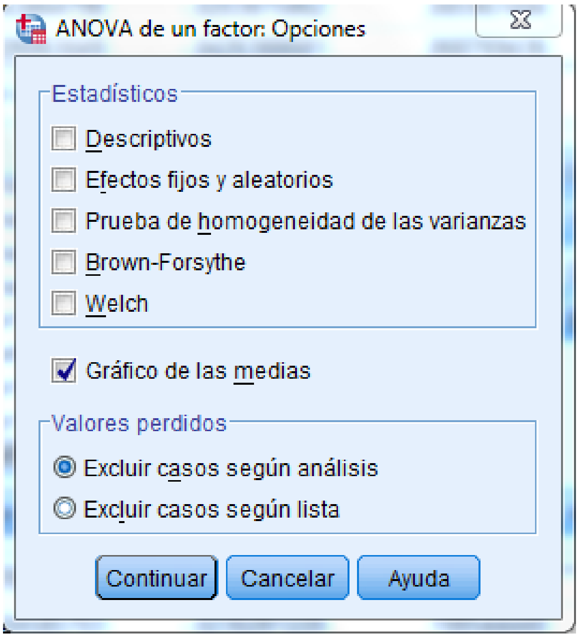
Se realiza el análisis en este caso de igual manera que para la prueba de homocedasticidad de varianzas:

Analizar – Comparar medias – Anova de un factor

Figura 8:
SPSS Statistics. Independencia (I).

En Opciones se elige Gráfico de las Medias:

Figura 9:
SPSS Statistics. Independencia (II).



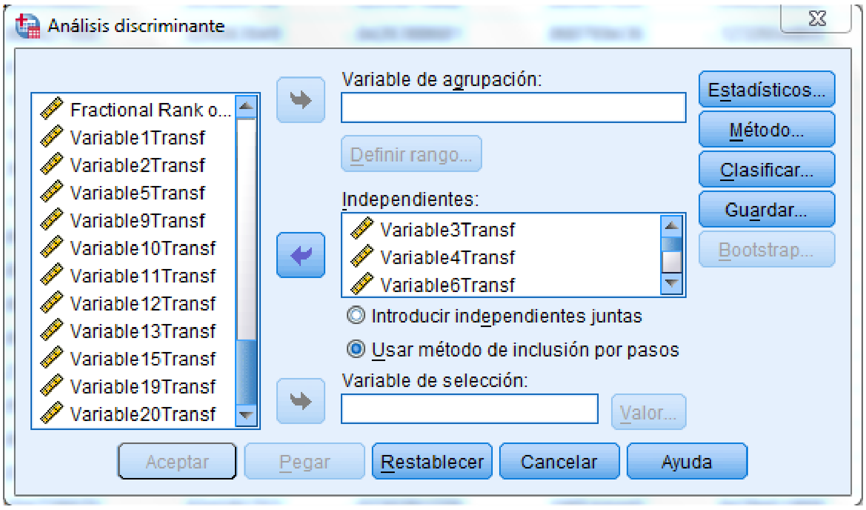
Análisis Discriminante

SPSS incorpora el procedimiento Análisis Discriminante que permite realizar Análisis Discriminante Múltiple de una manera bastante completa. Para realizar un análisis discriminante, se elige en el menú:

Analizar → Clasificar → Discriminante.

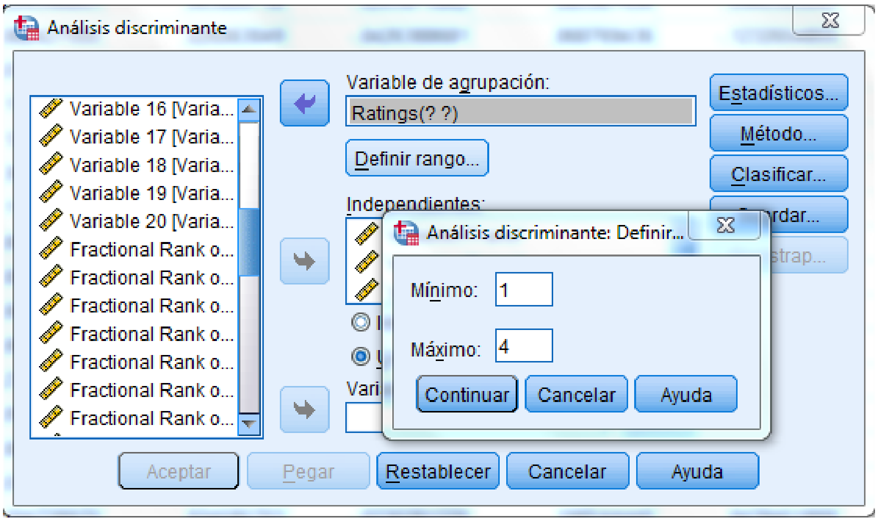
Seguidamente, se indican las variables que van a participar en el análisis. En este caso, las variables independientes elegidas para el análisis son: 3, 4, 5, 7, 8, 14, 16, y 17.

Figura 10:
SPSS Statistics. Aplicación Análisis Discriminante (I).



Se selecciona la variable dependiente (Calificación) y el rango, que en este caso es 1 (menor) y 4 (mayor).

Figura 11:
SPSS Statistics. Aplicación Análisis Discriminante (II).

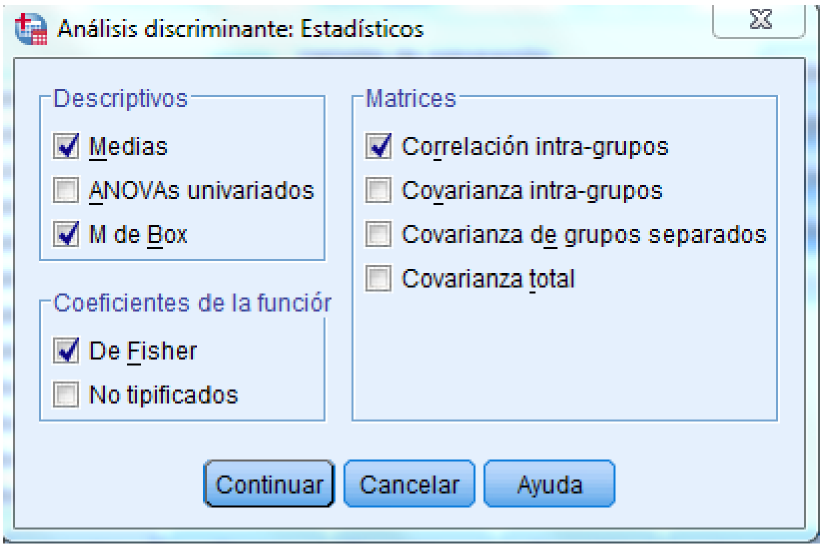


Las ventanas de *Estadísticos*, *Clasificar* y *Método* se rellenan de la siguiente forma:

Antes de nada, se marca la opción usar método de inclusión por pasos, que será el método elegido para el análisis, al ser la combinación del método hacia delante y hacia atrás.

Estadístico:

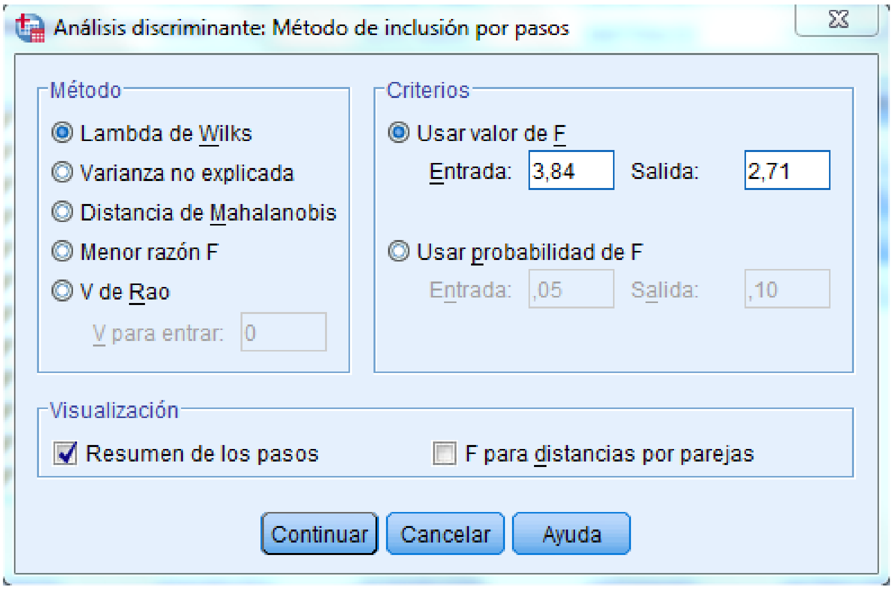
Figura 12:
SPSS Statistics. Aplicación Análisis Discriminante (III).



Método:

Los valores de F para las entradas y salidas de las variables de los que se hacían mención en el desarrollo del método se dejan con los valores por defecto que proporciona la herramienta.

Figura 13:
SPSS Statistics. Aplicación Análisis Discriminante (IV).



Clasificar:

Figura 14:
SPSS Statistics. Aplicación Análisis Discriminante (V).