



<Artículo>

Análisis Discriminante mediante SPSS

Mercedes Torrado-Fonseca, Vanesa Berlanga-Silvente

Fecha de presentación: 23/02/2013

Fecha de aceptación: 20/03/2013

Fecha de publicación: 01/07/2013

//Resumen

El análisis discriminante es un método estadístico a través del cual se busca conocer qué variables, medidas en objetos o individuos, explican mejor la atribución de la diferencia de los grupos a los cuales pertenecen dichos objetos o individuos. Es una técnica que nos permite comprobar hasta qué punto las variables independientes consideradas en la investigación clasifican correctamente a los sujetos u objetos.

Se muestran y explican los principales elementos que se relacionan con el procedimiento para llevar a cabo el análisis discriminante y su aplicación utilizando el paquete estadístico SPSS, versión 18, para el desarrollo del modelo estadístico, las condiciones para la aplicación del análisis, la estimación e interpretación de las funciones discriminantes, los métodos de clasificación y la validación de los resultados.

//Palabras clave

Análisis discriminante, clasificación de datos y sujetos, predicción, poder de discriminación.

// Referencia recomendada

Torrado-Fonseca, M. y Berlanga-Silvente, V. (2013). Análisis Discriminante mediante SPSS. [En línea] *REIRE, Revista d'Innovació i Recerca en Educació*, 6 (2), 150-166. Accesible en: <http://www.ub.edu/ice/reire.htm>

// Datos de los autores

Mercedes Torrado-Fonseca. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). mercedestorrado@ub.edu

Vanesa Berlanga-Silvente Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). berlanga.silvente@ub.edu



1. Introducción

En muchas de las investigaciones, independientemente del área de conocimiento, es habitual tener la necesidad de identificar cuáles son las características que diferencian unos grupos de sujetos u objetos respecto de otros, para así poder realizar predicciones futuras. Tanto el análisis de conglomerados como el *análisis discriminante* son técnicas que nos permiten clasificar sujetos u objetos a partir de características similares. La diferencia fundamental entre ambas pruebas es el momento del establecimiento de los grupos. En el *análisis discriminante* (AD) el investigador conoce a priori a qué grupo pertenece cada sujeto u objeto; en cambio, en el análisis de conglomerados los grupos o clúster se determinan y configuran a posteriori, es decir, una vez estudiadas y analizadas las agrupaciones.

El *análisis discriminante* es la prueba estadística apropiada para seleccionar qué variables independientes o predictivas permiten diferenciar grupos y cuántas de estas variables son necesarias para alcanzar la mejor clasificación posible. Además permite cuantificar su poder de discriminación en la relación de pertenencia de un sujeto u objeto a un grupo u otro. Por ello esta técnica es considerada, además de una prueba de clasificación, una prueba de dependencia. De hecho, su propósito es similar al análisis de regresión logística; la diferencia radica en que solo admite variables cuantitativas.

Un ejemplo de ello en el ámbito educativo podría ser el estudio del rendimiento estudiantil y la identificación de las variables que mejor lo predicen, a partir de las cualificaciones de una materia, como por ejemplo las matemáticas, que clasifica al alumnado en dos grupos: los de rendimiento alto y los de rendimiento bajo. Mediante un *análisis discriminante* se puede establecer el poder explicativo y discriminatorio de las características que diferencian a los alumnos según su rendimiento. Se procede al estudio de una serie de variables independientes como, por ejemplo, variables de carácter socioeconómico, variables académicas referentes a la preparación en bachillerato y variables actitudinales en relación con la variable dependiente que clasifica a los sujetos según el rendimiento obtenido en matemáticas. Según las características analizadas, a través de la descripción del grado de relación existente entre el conjunto de variables, se puede encontrar la frontera que separa el alto y el bajo rendimiento. Como resultado obtendremos una regla de clasificación que podrá ser utilizada en el pronóstico de adscripción al grupo de rendimiento establecido para nuevos estudiantes.

Definición de la técnica

El *análisis discriminante* tiene sus orígenes en las formulaciones del cálculo de distancias entre grupos, primero por Karl Pearson (1920), quien propuso el término de "*coeficiente de parecido racial*" y más tarde por Mahalanobis (1930), cuyo nombre ha mantenido la medida de distancia entre grupos. A partir de estos trabajos, R.A. Fisher (1936) introduce el término discriminación y da forma a la idea de combinación lineal de variables independientes para la discriminación de grupos. Desde ese momento, esta técnica se aplica a múltiples ámbitos disciplinares.



Es una técnica multivariante orientada fundamentalmente a lograr dos objetivos básicos:

- a) explicar la influencia de un conjunto de variables cuantitativas sobre una variable cualitativa.
- b) predecir la pertenencia de los sujetos a una de las categorías de la variable criterio a partir de los valores registrados en las variables predictivas.

La variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Con estos atributos medidos se forma una combinación lineal de variables predictivas para que maximice las diferencias entre los grupos y pueda construir así un modelo predictivo para clasificar nuevos casos al grupo de pertenencia.

Por tanto, el *análisis discriminante* será usado para:

- Determinar si existen diferencias significativas entre los "perfiles" de un conjunto de variables de dos o más grupos definidos a priori.
- Seleccionar las variables de predicción que contribuyen en mayor medida a explicar las diferencias entre los grupos.
- Establecer un procedimiento para clasificar a un individuo a partir de los valores de un conjunto de variables independientes.

Tras su aplicación podremos:

- Evaluar la exactitud de la clasificación en una tabla cruzada en la que se compara la pertenencia real de los individuos a un grupo con el pronóstico de pertenencia.
- Identificar qué variables independientes tienen mayor poder de discriminación y de predicción en la clasificación de sujetos.
- Obtener una ecuación denominada 'función discriminante' que exprese la combinación lineal de las variables predictivas (denominadas variables canónicas) y permita construir una regla de decisión que asigne un objeto nuevo a uno de los grupos prefijados con un cierto grado de riesgo.

2. Supuestos de aplicación de la prueba

Antes de la utilización de cualquier prueba estadística se debe comprobar el cumplimiento de los supuestos básicos de aplicación. En el caso que nos ocupa se pueden resumir en dos: (i) las variables independientes o predictivas deben seguir una distribución normal multivariante y (ii) las matrices de covarianzas deben ser iguales en todos los grupos.

Aunque el *análisis discriminante* es considerado una técnica robusta que no se ve gravemente afectada si alguno de los supuestos anteriores no se cumple, es recomendable aplicar el test de M. de Box para comprobar el segundo supuesto. El test de M. de Box parte del supuesto de que las matrices de covarianzas son iguales y se basa en el cálculo de los determinantes de covarianza de cada grupo, el valor obtenido se aproxima a la F de Snedecor.

En referencia al tipo de variables que deben utilizarse, 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. No es imprescindible que el volumen de casos sea similar en cada grupo.

Las variables independientes utilizadas para el análisis deben ser de intervalo o de razón, y seguir una distribución normal. En algunos casos, también se acepta trabajar con variables dummy resultado de la transformación de una variable cualitativa (nominal u ordinal) a una cuantitativa con valores 0 y 1. En ningún caso las variables independientes pueden ser combinaciones lineales de otras variables discriminantes.

3. Procedimiento de la técnica

El *análisis discriminante* consiste en partir de n individuos con información de p variables y agrupados en una variable y con dos categorías o más para que cada sujeto obtenga una serie de puntuaciones que indiquen a qué grupo pertenecen. Se sigue un método similar al análisis factorial, se busca una función lineal de modo que se maximice la variabilidad entre los grupos para discriminarlos mejor.

La aplicación del *análisis discriminante* obtiene como resultado una ecuación denominada función discriminante que expresa la combinación lineal de las variables predictivas (denominadas variables canónicas). El máximo número de funciones discriminantes que se obtiene es igual al mínimo entre el número de variables y el número de grupos menos 1 [con q grupos, $(q-1)$].

Al igual que en el análisis de regresión múltiple, la combinación lineal de las variables predictivas da como resultado una variable no observada (y), pero esta variable expresada cuantitativamente no puede ser considerada como la parte explicada de una variable criterio. Los coeficientes de discriminación a_j son coeficientes de ponderación de las variables predictivas y expresan la contribución de dichas variables en la función discriminante. Asimismo, en el resultado de los



coeficientes discriminantes tiene gran influencia el grado de covariación que presentan las variables predictivas.

$$Y = a_0 + a_1 * X_1 + a_2 * X_2 + \dots + a_p * X_p$$

X = las variables independientes

a_0 = la constante

a_p = los coeficientes de discriminación

La ecuación discriminante resultante será óptima en la medida en que proporcione una regla de clasificación que disminuya los errores de clasificación y en la medida en que explique una mayor parte de la variabilidad intragrupos. Para esto las variables cuantitativas deberían cumplir los supuestos paramétricos.

La aplicación de la técnica sigue tres pasos:

Selección de las variables discriminantes

El proceso de selección de las variables discriminantes suele combinar, por un lado, la identificación de variables importantes de modelos teóricos validados en investigaciones previas y, por otro, la aplicación de pruebas estadísticas. Ambas vías son complementarias y no excluyentes. La selección estadística consiste en la combinación de un proceso de análisis descriptivo (cálculo de las medias y desviaciones típicas), correlación entre las variables y diferencias significativas entre los grupos. La correlación se comprobará dentro del grupo y entre grupos. Se utilizará λ de Wilks y la correlación canónica.

Selección de la función discriminante

La función discriminante no es otra cosa que un factor o una nueva variable, combinación lineal de las anteriores. Podremos obtener más de una función discriminante pero solo una permitirá hacer predicciones de las mejores clasificaciones futuras. Se utilizará el coeficiente de "autovalor" y los coeficientes estandarizados de las funciones discriminantes.

Clasificación de los sujetos u objetos

Para poder asignar un sujeto a un grupo u otro tenemos que diseñar una regla de decisión. Una vez aplicada se hará una valoración en porcentaje de aciertos en el pronóstico de clasificación de los sujetos y se obtendrá una tabla resumen de clasificación.

4. Procedimiento del análisis discriminante en SPSS

Proponemos el siguiente caso práctico:

Se ha realizado una encuesta a 633 estudiantes del Grado de Enfermería de la Universidad de Barcelona con el objetivo de identificar qué variables independientes tienen mayor poder de discriminación y de predicción en la clasificación de sujetos que persisten y abandonan en el primer año de carrera universitaria.

En el caso que nos ocupa, se desea poder predecir el riesgo de abandono en el primer año de universidad del Grado de Enfermería. Para ello explotaremos el fichero de datos de la cohorte 2011-12 de estudiantes de nuevo ingreso en dicha titulación y se estudiarán variables cuantitativas potencialmente explicativas: nota de admisión, créditos presentados, créditos aprobados y satisfacción académica inicial (variable generada a partir del método de regresión del factor resultante de la aplicación del análisis factorial de la escala "Satisfacción Académica").

Para ejecutar el *análisis discriminante* en la matriz de datos se seleccionan los menús: Analizar – Clasificar - *Análisis discriminante*.

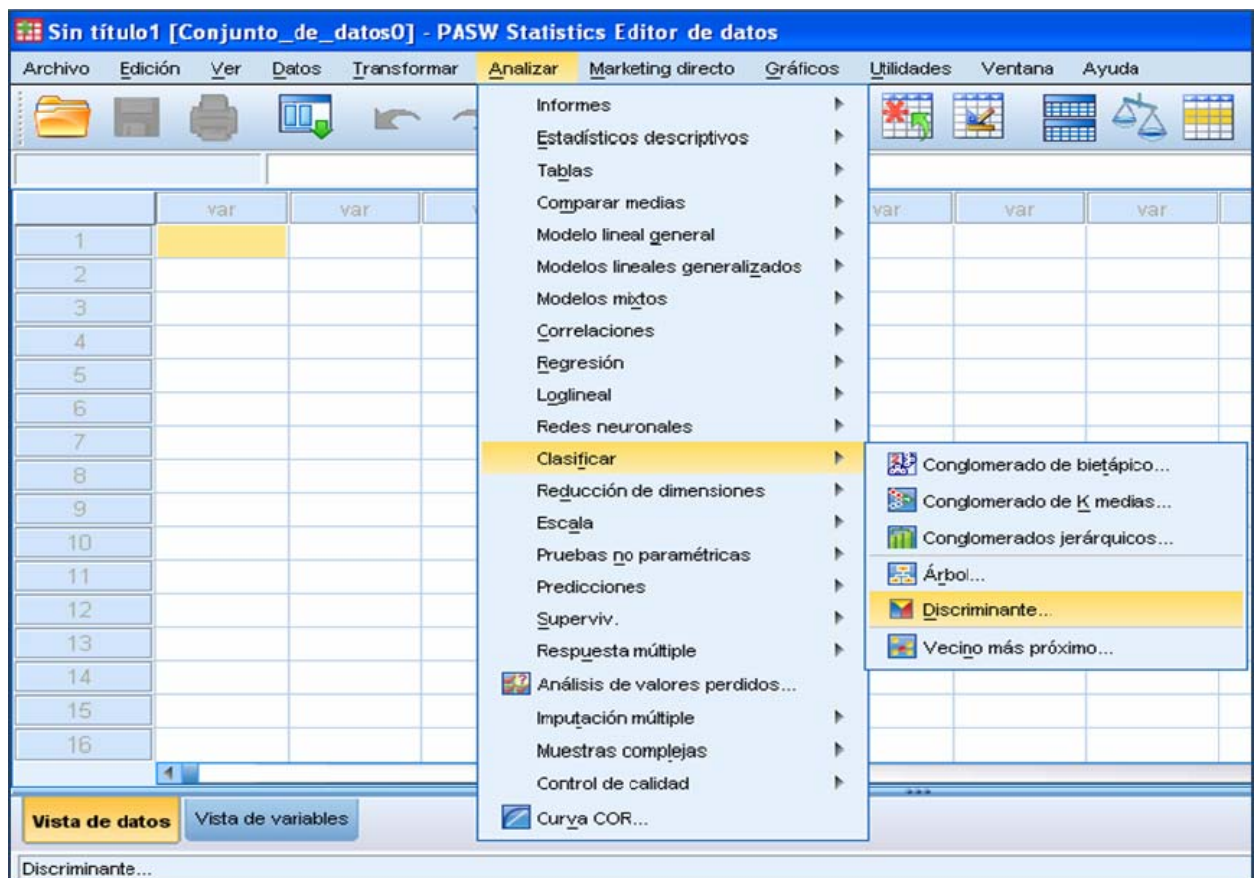


Figura 1. Cuadro de diálogo análisis discriminante en SPSS



Introduciremos como variable de agrupación la variable dependiente. Esta variable solo podría tomar valores enteros, de tal manera que indicaremos el valor máximo y el valor mínimo. En nuestro caso introduciremos la variable denominada "persistencia" y definiremos los rangos como mínimo, el 0 (abandona), y como máximo, el 1 (persiste).



Figura 2. Ventana del análisis

Como variables independientes introducimos todas las variables que queremos utilizar para formular el modelo: créditos aprobados, créditos presentados, nota de admisión y satisfacción académica.

Podemos utilizar dos formas para obtener las variables que configuran la función discriminante resultante: "introducir independientes juntas" o "usar método de inclusión por pasos". Es interesante señalar que si elegimos "usar método de inclusión por pasos" activamos automáticamente otro botón "Métodos" situado en la parte derecha-superior. Finalmente, y de manera opcional, podemos, mediante la opción de "variable de selección", aplicar el análisis en una selección de individuos de la matriz.

Por defecto obtenemos para cada función discriminante una tabla donde aparece el valor de "autovector", el estadístico con mayor poder discriminante que indica la proporción de la varianza total explicada por las m funciones discriminantes. La suma de los autovalores es la proporción de varianza total que queda explicada.



a) Botón Estadísticos

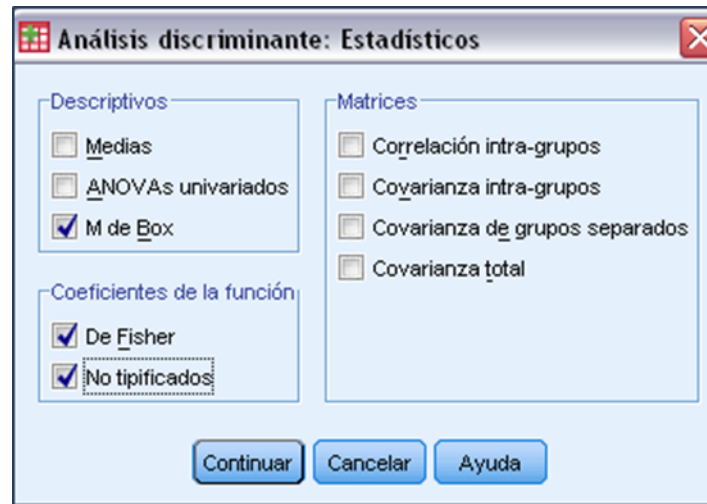


Figura 3. Botón Estadísticos

Descriptivos

Los descriptivos permitirán hacer un primer nivel de selección de las variables independientes y valorar los supuestos básicos de aplicabilidad de la prueba.

- Medias: muestra la media y desviación típica totales y las medias y desviaciones típicas de cada grupo para las variables independientes. Las medias deben ser diferentes y los valores de las desviaciones típicas pequeños.
- ANOVAs univariados: realiza un análisis de varianza de un factor sobre la igualdad de las medias de grupo para cada variable independiente. Con este análisis se puede comprobar si las varianzas para cada grupo de cada variable son iguales. Esta opción incluye además el estadístico λ de Wilks univariante, que es igual al cociente entre la suma de cuadrados dentro de los grupos y la suma de cuadros total (sin distinguir grupos). Si su valor es pequeño la variable discrimina mucho y es consecuencia de las diferencias entre los grupos y no dentro de ellos.
- M de Box: contrasta la igualdad de las matrices de covarianza de los grupos. Este test parte del supuesto de que las matrices de varianza-covarianza poblacionales correspondientes a cada grupo son iguales entre sí. El *output* ofrece su transformación en un estadístico F. La significación del test confirma que uno de los grupos es más variable que otro.



Coeficientes de la función

Los coeficientes de la función permitirán identificar la función discriminante con mayor poder explicativo.

- Coeficientes de clasificación de Fisher: muestra los coeficientes de la función de clasificación de Fisher que pueden utilizarse directamente para la clasificación. Se obtiene una función de clasificación para cada grupo, y se asigna un caso al grupo para el que tiene una mayor puntuación discriminante.
- Coeficientes no tipificados: son los coeficientes utilizados para calcular las puntuaciones discriminantes y la ubicación de los centroides de los grupos. No es habitual utilizar esta opción dado que el programa calcula automáticamente estas puntuaciones.

Para nuestro caso, en la ventana *Estadísticos* marcaremos M. de Box y los Coeficientes de la Función (de Fisher y No Tipificados).

b) Botón Clasificación

Esta opción de la técnica permite matizar situaciones de partida previas a su aplicación (desigualdad de los grupos resultantes), obtener una tabla resumen de la clasificación, gráficos de las funciones discriminantes y matrices de covarianzas.

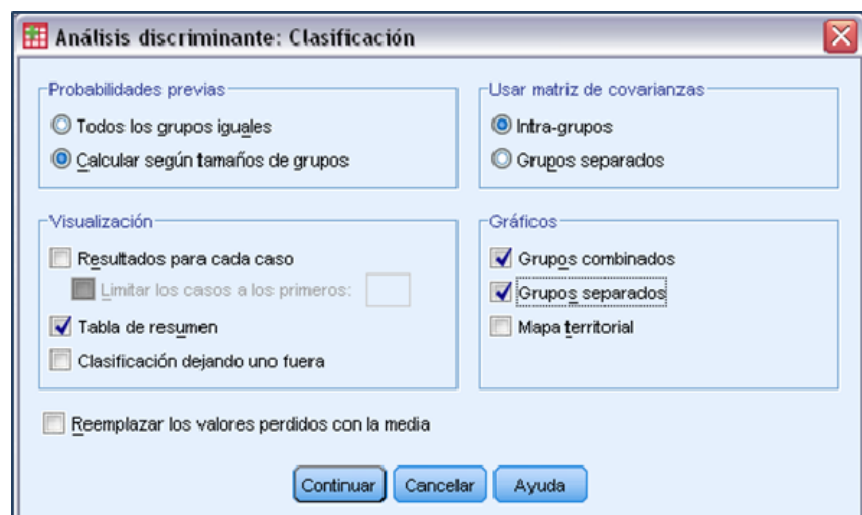


Figura 4. Botón Clasificación



Probabilidades previas

- Todos los grupos iguales: las probabilidades previas serán iguales para todos los grupos.
- Calcular según tamaños de grupos: los tamaños de grupo observados en la muestra determinan las probabilidades de la pertenencia al grupo.

Visualización

- Resultados para cada caso: muestran para cada caso los códigos del grupo real de pertenencia, el grupo pronosticado, las probabilidades posteriores y las puntuaciones discriminantes.
- Tabla de resumen: número de casos correcta e incorrectamente asignados a cada uno de los grupos, basándose en el *análisis discriminante*. Suele recibir el nombre de tabla de clasificación.
- Clasificación dejando uno fuera: se clasifica cada caso del análisis mediante la función derivada a partir de todos los casos, excepto el propio caso.

Usar matriz de covarianza

- Intra-grupos: se utiliza la matriz de covarianza intra-grupos combinada para clasificar los casos.
- Grupos separados: para la clasificación se utilizan las matrices de covarianza de los grupos separados.

Gráficos

- Grupos combinados: crea un diagrama de dispersión de los valores en las dos primeras funciones discriminantes. Si solo hay una función obtendremos un histograma.
- Grupos separados: crea diagramas de dispersión de los grupos por separado, para los valores en las dos primeras funciones discriminantes y un histograma en caso de una función.
- Mapa territorial: gráfico de las fronteras utilizadas para clasificar los casos en grupos a partir de los valores en las funciones. Los números corresponden a los grupos en los que se clasifican los casos. La media de cada grupo se indica mediante un asterisco situado dentro de sus fronteras. No se mostrará el mapa si solo hay una función discriminante.

Para nuestro análisis dentro de la ventana de *Clasificación* vamos a señalar las siguientes opciones: calcular según tamaños de grupos debido a la desigualdad manifiesta entre el volumen de estudiantes que persisten y los que abandonan, intra-grupos en la matriz de



covarianzas, marcaremos los gráficos grupos combinados y separados, y que muestre la tabla de resumen.

c) Botón Método

En los análisis de discriminante se trabaja con muchas variables, evidentemente algunas serán más influyentes que otras a la hora de discriminar a un individuo en un grupo u otro. Lo que intentaremos con esta opción será utilizar solo aquellas variables más influyentes con lo que simplificaremos el modelo. Para seleccionar las variables tendremos que usar, en la ventana del análisis, *método de selección de variables por pasos*. Al hacer esto, el botón de *Método* se activa, pudiendo seleccionar en esta ventana el método por el cual se elegirán las variables que deben utilizarse. Obtendremos información sobre la significación individual de cada variable en la incorporación de la función discriminante resultante. La tabla resultante presenta para cada variable, el nivel de tolerancia (porcentaje de variabilidad de cada variable), el valor de F (permite valorar si la variable debe o no ser rechazada en la función) y λ de Wilks global que obtendríamos si se eliminara la variable del modelo. El método más utilizado es la λ de Wilks.

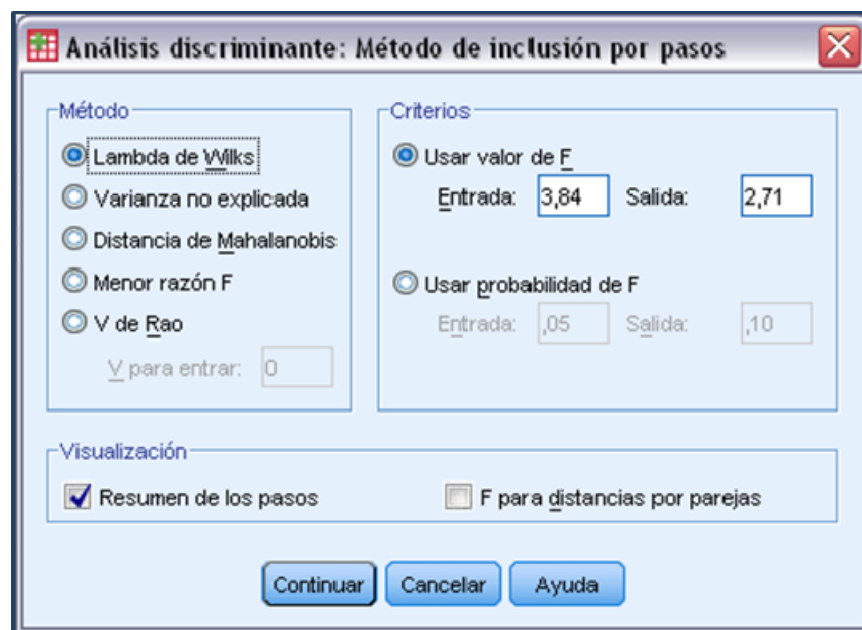


Figura 5. Botón Método

d) Guardar

Con esta opción podemos guardar en el fichero de datos para su utilización y estudio las siguientes variables:

- Grupo de pertenencia pronosticado, es decir, el grupo asignado por el análisis de discriminante. Se guarda como Dis 1.
- Puntuación discriminante: valor de las dos funciones discriminantes. Dis1 1, Dis1 2.



- Probabilidad de pertenencia: probabilidad de que ese individuo pertenezca a cada uno de los grupos. Dis2 1, Dis2 2 y Dis2 3.

5. Interpretación de Outputs

Recordemos que nuestro caso parte de la situación de persistencia y abandono de alumnos de primero de carrera. La aplicación del *análisis discriminante* permitirá responder a la siguiente pregunta: *¿Qué poder de discriminación tienen las variables de nota de admisión, créditos presentados, créditos aprobados y la satisfacción académica manifestada al inicio de curso en la predicción de la persistencia universitaria?*

Si seleccionamos todo lo que hemos indicado anteriormente, obtendremos una serie de salidas. Los supuestos paramétricos mediante el test M. de Box muestran, por un lado, el valor del test y su transformación en un estadístico F, así como su significación. El resultado obtenido de la prueba confirma que las matrices de varianzas-covarianzas son diferentes. Hay que tener en cuenta que el no cumplimiento de este supuesto paramétrico es especialmente sensible en muestras grandes y en desviaciones de la normalidad multivariante de alguna variable.

Test Results

Box's M		78,316
F	Approx.	7,638
	df1	10
	df2	45869,024
	Sig.	,000

Tests null hypothesis of equal population covariance matrices.

Figura 6. Test M. de Box

Posteriormente aparecen dos tablas que nos indican, por un lado, la estimación de la variabilidad intergrupo explicada en la función discriminante ("autovalor") y, por otro, la porción de varianza total de las puntuaciones discriminantes que no ha sido explicada por la diferencia entre los grupos (λ de Wilks).

Claves de la interpretación

Autovalor: Cuanto más alto es su valor, más eficaz será el análisis para clasificar a los sujetos. El valor mínimo es cero y no tiene un valor máximo.



Correlación canónica: Recoge la pertenencia de los sujetos a los grupos mediante un coeficiente que oscila entre 0 y 1. Interesa que presente un valor lo más próximo a 1.

λ de Wilks: Representa la diferencia entre los grupos. Los valores próximos a 1 indicarán similitudes entre los grupos. Si su valor es cercano a 0 deberá interpretarse que los grupos son diferentes y, por lo tanto, la función discriminante puede pronosticar adecuadamente a los sujetos. Su significación se realiza mediante la transformación al valor de *Chi-square*.

Tras los primeros valores obtenidos del caso (autovalor de 0.437, correlación canónica de 0.552) extraemos una primera conclusión: existe una única función discriminante que permite de forma significativa (sig. 000) clasificar a los sujetos en los dos grupos de persistencia y abandono. El valor λ de Wilks (0.696) arroja una segunda conclusión: aunque la función discriminante servirá para pronosticar la pertenencia a los grupos, seguramente no todas las variables sean discriminantes. Su valor denota ciertas similitudes entre los grupos y, por lo tanto, deberíamos estudiar la influencia de cada una de las variables en la función discriminante obtenida. En consecuencia, el proceso aconsejado es realizar el test con todas las variables mediante la opción "usar método de inclusión por pasos" para, posteriormente, eliminar las no significativas en la función.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,437 ^a	100,0	100,0	,552

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,696	225,930	2	,000

Figura 7. Lambda de Wilks y autovalor

A continuación aparecen tres tablas que nos indican todos los pasos seguidos para la construcción de la función discriminante y consecuentemente qué variables independientes de las consideradas inicialmente son significativas para el modelo. En nuestro caso solo se incluyen dos variables: créditos aprobados y créditos presentados. Las notas a pie de página de la primera tabla nos indican que se ha utilizado el valor de λ de Wilks global, el estadístico F para incorporar variables (criterio de entrada) y como estadístico para excluir variables (criterio de salida), y que el nivel de F ha sido insuficiente para continuar los cálculos; en otras palabras, no se han incluido todas las variables definidas para el análisis.

El contraste de la lambda de Wilks es un test para el contraste de las medias de todas las funciones discriminantes en todos los grupos. De tal manera que si el p-valor es inferior a 0.05 nos llevará a aceptar que existen diferencias de comportamiento entre las medias de los grupos.



Por lo tanto, el proceso realiza el test con todas las funciones para, a continuación, ir distribuyendo en dos tablas las variables seleccionadas de las que no lo son.

Variables Entered/Removed^{a, b, c, d}

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	Creditos_ aprobados	,715	1	1	624,000	249,258	1	624,000	,000
2	Creditos_ presentados	,696	2	1	624,000	136,166	2	623,000	,000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- a. Maximum number of steps is 8.
- b. Minimum partial F to enter is 3.84.
- c. Maximum partial F to remove is 2.71.
- d. F level, tolerance, or VIN insufficient for further variables in the Analysis

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	Creditos_ aprobados	1,000	249,258	
2	Creditos_ aprobados	,663	78,796	,784
	Creditos_ presentados	,663	16,774	,715

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	Creditos_ presentados	1,000	1,000	172,083	,784
	Creditos_ aprobados	1,000	1,000	249,258	,715
	Factor: Satisfacción académica inicial	1,000	1,000	19,118	,970
	Nota_ Admissió	1,000	1,000	16,172	,975
1	Creditos_ presentados	,663	,663	16,774	,696
	Factor: Satisfacción académica inicial	,970	,970	1,974	,712
	Nota_ Admissió	,917	,917	,211	,714
2	Factor: Satisfacción académica inicial	,968	,645	2,395	,693
	Nota_ Admissió	,904	,602	,001	,696

Figura 8. Selección de variables



Con la matriz de coeficientes estandarizados y la matriz de estructuras, que representa las correlaciones entre las funciones discriminantes y las variables, podremos estudiar qué variables son más influyentes en las funciones discriminantes. Se puede observar cómo las variables eliminadas de la función presentan valores en sus coeficientes cercanos a 0, frente a los valores obtenidos en las variables de créditos aprobados y presentados (0.956 y 0.794 respectivamente).

	Function
	1
Creditos_presentados	,361
Creditos_aprobados	,746

	Function
	1
Creditos_aprobados	,956
Creditos_presentados	,794
Nota_Admissió ^a	,241
Factor: Satisfacción académica inicial ^a	,154

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Figura 9. Matrices de coeficientes estandarizados y de estructuras

Con la matriz de centroides obtenemos las medias de cada grupo para las dos funciones. Si las medias de cada grupo fueran parecidas, ello implicaría que la función no discrimina los grupos y viceversa si las medias fueran diferentes. En nuestro caso son diferentes, tal y como observamos en la figura 9. Los valores negativos de los estudiantes clasificados en "abandona 1er año" ilustran la influencia negativa de las variables seleccionadas y, por lo tanto, muestran que a mayor valor de las mismas, más influencia para que los sujetos se clasifiquen en el grupo de "persiste".

Abandona-persiste	Function
	1
ABANDONA 1er AÑO	-2,027
PERSISTE 1er AÑO	,215

Unstandardized canonical discriminant functions evaluated at group means

Figura 10. Matriz de centroides



Centrando nuestra atención en la función discriminante, la tabla de coeficientes estandarizados de las funciones discriminantes permite identificar aquellas variables con mayor peso en el modelo predictivo y posibilita identificar la función discriminante resultante. Siendo en nuestro caso:

$$D1 = -4,240 + 0,037 \text{ créditos presentados} + 0,049 \text{ créditos aprobados}$$

Canonical Discriminant Function Coefficients

	Function
	1
Creditos_presentados	,037
Creditos_aprobados	,049
(Constant)	-4,240

Unstandardized coefficients

Figura 11. Coeficientes de la función discriminante

El último paso del análisis es el cuadro resumen de la clasificación de los sujetos a partir de la aplicación de la función discriminante obtenida. En nuestro caso, vemos que se ha clasificado correctamente el 91,4% a partir de los créditos presentados y aprobados. Los estudiantes que se presentan a menos créditos de los matriculados y tienen menos éxito académico son los que tras el primer año no persisten y deciden abandonar.

Classification Results^a

Abandona-persiste			Predicted Group Membership		Total
			ABANDONA 1er AÑO	PERSISTE 1er AÑO	
Original	Count	ABANDONA 1er AÑO	155	50	205
		PERSISTE 1er AÑO	57	987	1044
	%	ABANDONA 1er AÑO	75,6	24,4	100,0
		PERSISTE 1er AÑO	5,5	94,5	100,0

a. 91,4% of original grouped cases correctly classified.

Figura 12. Resultados finales



<Referencias bibliográficas>

- Bisquerra, R. (1989). Análisis discriminante. En Rafael Bisquerra. *Introducción conceptual al Análisis Multivariable. Un enfoque informático con los paquetes Spss, Bmdp, Lisrel y Spad* (pp. 243-280). Barcelona: PPU.
- Ferrán, A., M. (2001). *Spss para windows – Análisis estadístico*. España: McGraw Hill/Interamericana.
- Gil, J.; García Jiménez, E. y Rodríguez Gómez, G. (2001). *Análisis discriminante. Cuadernos de Estadística n° 12*. Madrid: La Muralla.
- Gondar, N., J. E. (2001). *Análisis discriminante* [En línea] Disponible en: <http://www.estadistico.com/arts.html?20011112>
- Hernandez, J.; Ramírez, M.J. y Ferri, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson educación.
- Johnson, D., E. (2000). *Métodos multivariados aplicados al análisis de datos*. D.F. México: C. V., International Thomson Editores.
- Lind, D.A., Marchal, W.G., Wathen, S.A. (2012). *Estadística aplicada a los negocios y la economía*. México D.F.: McGraw-Hill.
- Pérez, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson educación.
- Pérez, C. y Santín, D. (2007). *Minería de Datos: Técnicas y Herramientas*. Madrid: Ediciones Paraninfo.
- Silberschatz, A. (2007). *Fundamentos de diseño de bases de datos* (5ª ed.). Madrid: McGraw-Hill / Interamericana de España.
- Valderrey, P. (2010). *Técnicas de segmentación de mercados*. Madrid: Starbook editorial.

Copyright © 2013. Esta obra está sujeta a una licencia de Creative Commons mediante la cual, cualquier explotación de ésta, deberá reconocer a sus autores, citados en la referencia recomendada que aparece al inicio de este documento.

