ANÁLISIS DISCRIMINANTE

El Análisis Discriminante es un procedimiento estadístico que construye un modelo de pronóstico que permite clasificar casos en distintos grupos basándose en las características observadas para cada caso. A partir de unos grupos previamente definidos, el procedimiento genera una función de clasificación (o, para más de dos grupos, un conjunto de funciones de clasificación), basándose en las combinaciones lineales de las variables predictoras que proporcionan la mejor discriminación entre los casos pertenecientes a dichos grupos. Estas funciones se generan a partir de una muestra de casos para los que se conoce su pertenencia a uno de los distintos grupos; después, se pueden emplear dichas funciones para clasificar nuevos casos, de los que conocemos su valor para las variables predictoras, pero de los que no conocemos el grupo al que pertenecen, para tratar de prever a qué grupo es más probable que pertenezcan.

Para ilustrar el procedimiento de análisis discriminante que utiliza SPSS, es necesario tener un conjunto de datos en el **Editor de datos**. Para ello, nos situamos en la ventana del **Editor de datos**, y en el menú **Archivo** elegimos **Abrir**, y seleccionamos el fichero **Compra_coche.sav**, que contiene información relativa a un conjunto de 20 personas que visitaron un concesionario de vehículos de una determinada marca, de tal manera que cuando hicieron la primera visita se les recogió información sobre su salario mensual y su edad. Finalmente, hubo 10 de esas personas que adquirieron un vehículo, y otras 10 personas que no lo compraron. Esta información se encuentra en las siguientes variables:

- **compra:** Resultado de la decisión de compra de la persona, una vez que visitó el concesionario. Esta variable es categórica (con una escala de medida nominal), y presenta los siguientes valores:
    - **0:** No compra.
    - **1:** Sí Compra.
- **salario:** Salario mensual.
- **edad:** Edad (en años).

Dentro de SPSS, el procedimiento que realiza el **Análisis discriminante** se encuentra en el submenú **Clasificar** del menú **Analizar**, como se muestra en la Figura 1:.

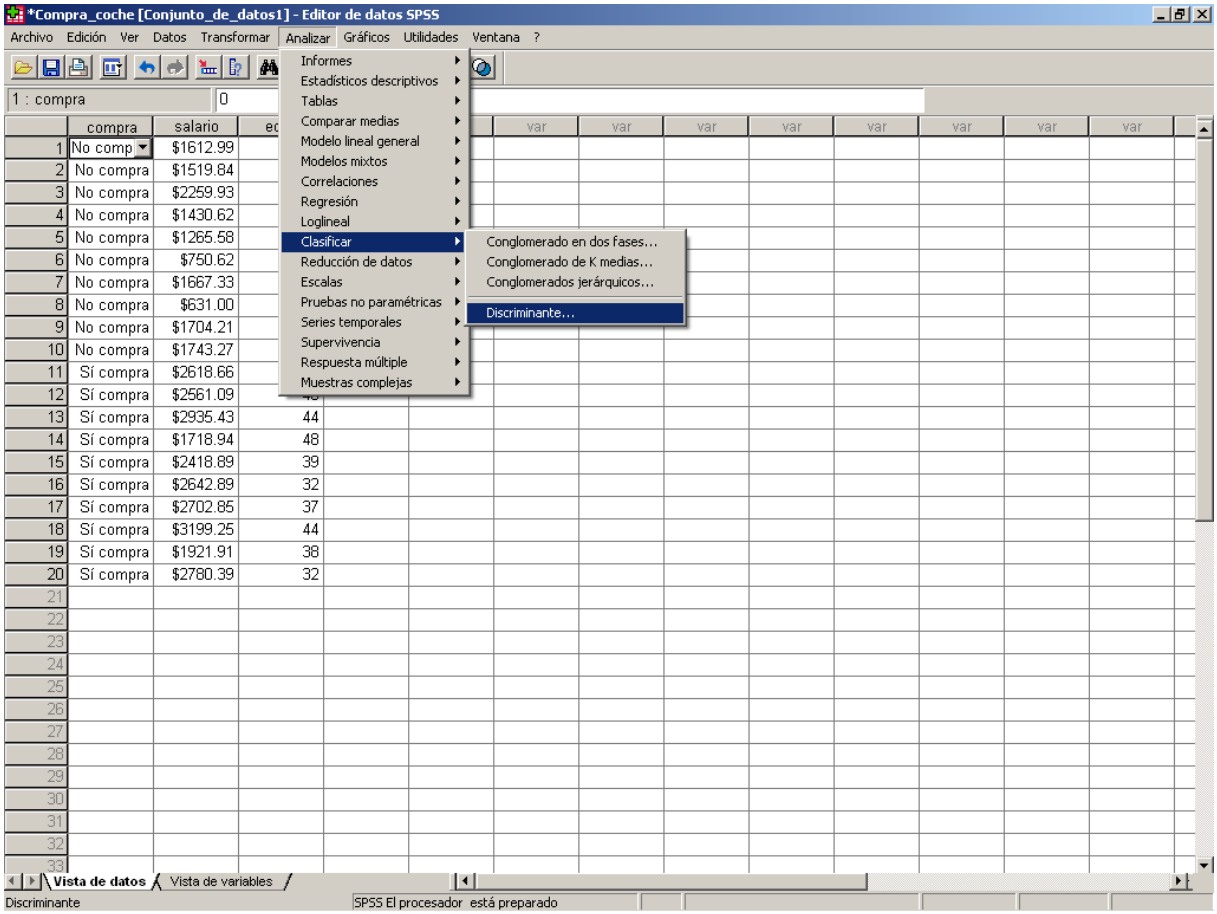


Figura 1: Selección del procedimiento **Análisis discriminante**.

Al pulsar en dicha opción, el cuadro de diálogo que aparece tiene el aspecto de la Figura 2, en el que aparecen todas las opciones propias del análisis discriminante, y que son las que emplea SPSS para la realización de este procedimiento.

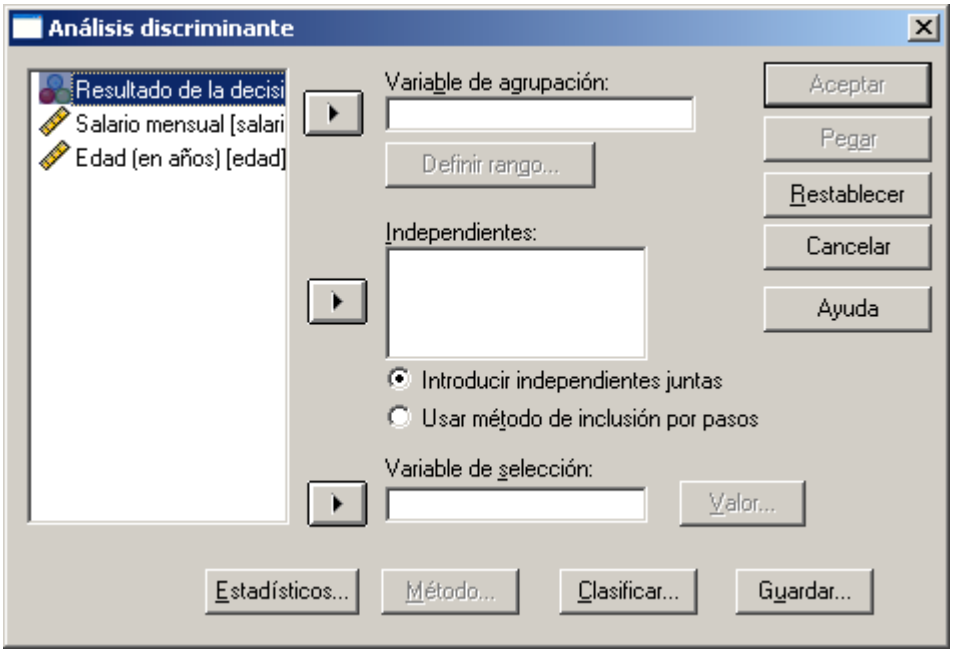


Figura 2: Cuadro de diálogo del procedimiento **Análisis discriminante**.

Para empezar, debemos seleccionar las variables que van a ser usadas como clasificadoras, y la variable de agrupación, para la cual debemos además especificar cuál es el rango de posibles valores que puede tomar (dichos valores deben ser números enteros que permitan identificar los distintos grupos de clasificación, por ejemplo de 1 a 2, si hubiera dos grupos, o de 1 a 4, si el número de grupos fuera cuatro). Con los datos del fichero **Compra_coche.sav**, vamos a tratar de explicar la decisión de compra (variable **compra**, cuyos valores son 0-1), en función del salario mensual (variable **salario**).

Si observamos la casilla de **Variable de selección**, con esta opción se permite limitar el análisis a un subconjunto de casos que tengan un valor particular en una variable. Después de seleccionar esta opción se debe elegir una variable de selección e introducir un valor para la variable de selección de casos (véase la Figura 3):

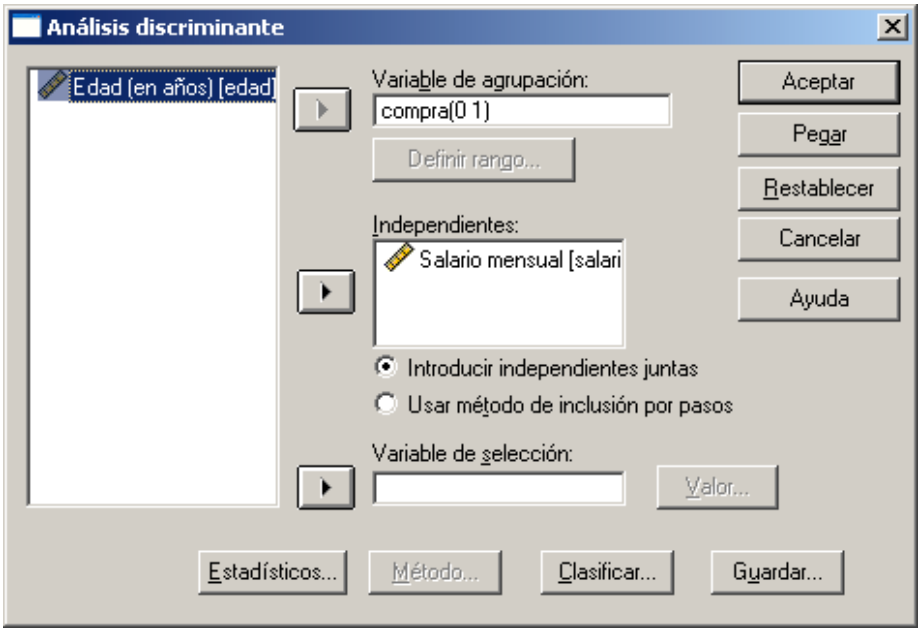


Figura 3: Cuadro de diálogo de **Análisis discriminante** con las variables introducidas.

Si pulsamos sobre el botón de **Estadísticos...**, obtenemos el cuadro de diálogo de la Figura 4, en el que podemos solicitar diferentes medidas numéricas sobre el análisis que vamos a realizar, como las medias por grupo, el contraste de la M de Box, los coeficientes de la función discriminante o las matrices de varianzas y covarianzas dentro, entre grupos y total:

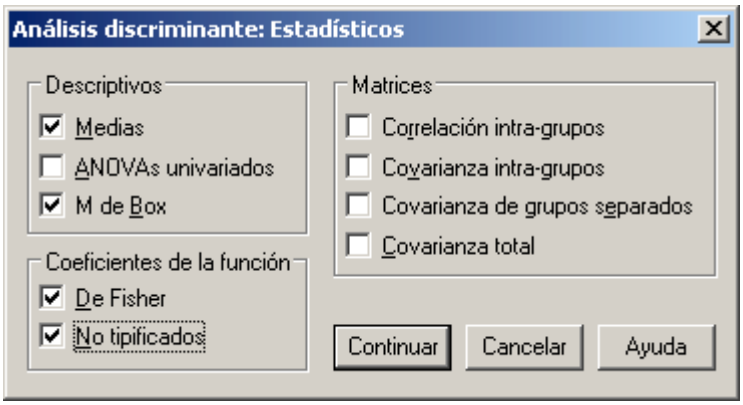


Figura 4: Opciones de **Estadísticos** de **Análisis discriminante**.

Dentro del método que se puede elegir, existen distintas posibilidades siempre que vayamos introduciendo variables explicativas paso a paso, como se muestra en la Figura 5 (la inclusión se hace de forma similar a como se hace en los modelos de regresión, aunque nosotros no analizamos este caso durante el presente curso, en el que nos limitaremos a introducir simultáneamente todas las variables independientes).

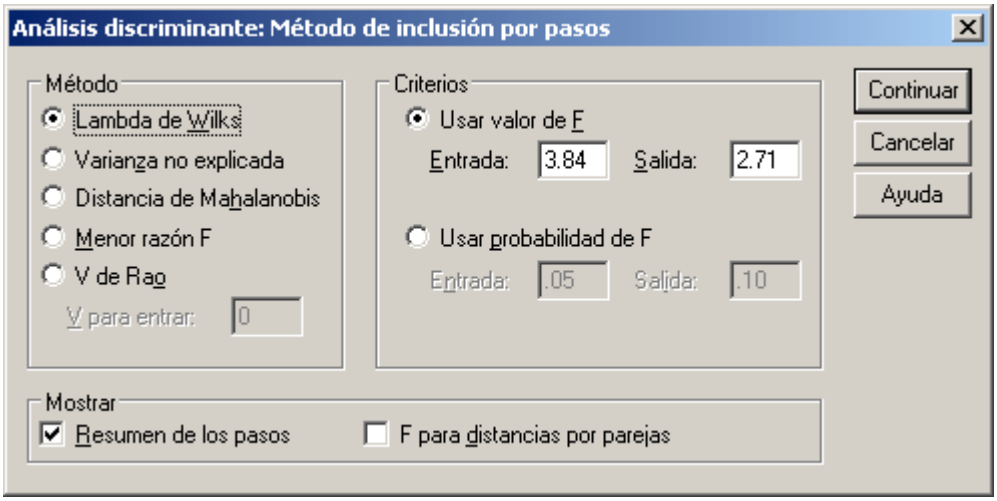


Figura 5: Opciones del **Método de inclusión por pasos** de **Análisis discriminante**.

Las opciones de clasificación se muestran en la Figura 6, en la cual se aprecian las diversas posibilidades en cuanto a los resultados que se pueden generar. En esta versión no se presentan los gráficos de los grupos combinados cuando hay dos grupos, razón por la cual no marcamos esta opción:

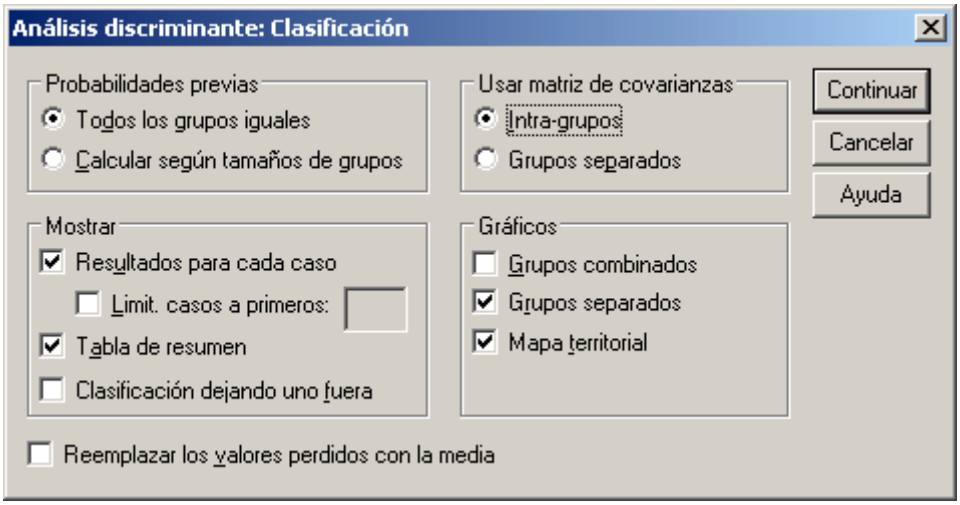


Figura 6: Opciones de **Clasificación** de **Análisis discriminante**.

Además, SPSS ofrece la posibilidad de guardar como una variable nueva el grupo al que pertenece cada caso, así como las puntuaciones discriminantes o las probabilidades de pertenencia a cada uno de los distintos grupos (Figura 7):

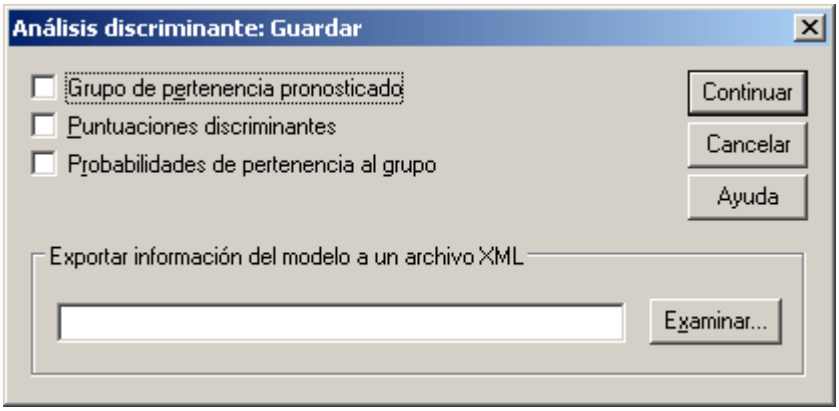


Figura 7: Opciones de **Guardar** de **Análisis discriminante**.

Al aceptar la ejecución del procedimiento, los resultados obtenidos son los siguientes:

Discriminante

Estadísticos de grupo

compra		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
No compra	Salario mensual	$1,458.54	$480.61	10	10.000
Sí compra	Salario mensual	$2,550.03	$442.51	10	10.000
Total	Salario mensual	$2,004.28	$718.11	20	20.000

En esta primera tabla se muestran las características descriptivas muestrales de las variables independientes (en este caso sólo una, X_1=salario). Debe observarse que el punto de corte de los dos grupos para esta variable se sitúa en el valor 2004.28 ((1458.54+2550.03)/2=2004.28).

Análisis 1
Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

compra	Rango	Logaritmo del determinante
No compra	1	12.350
Sí compra	1	12.185
Intra-grupos combinada	1	12.271

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

Resultados de la prueba

M de Box		.061
F	Aprox.	.058
	gl1	1
	gl2	972.000
	Sig.	.810

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Se comprueba que las matrices de varianzas-covarianzas poblacionales son iguales, ya que el p-valor asociado al estadístico de prueba en el contraste de la M de Box es 0.810 (superior al 5%).

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1.551[a]	100.0	100.0	.780

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	.392	16.387	1	.000

El autovalor de la única función discriminante es 1.551, y mide las desviaciones de las puntuaciones discriminantes entre grupos respecto a las desviaciones dentro de los grupos. El autovalor de una función discriminante se interpreta como la parte de variabilidad total de la nube de puntos proyectada sobre el conjunto de todas las expresiones que son atribuibles a la función. Si su valor es grande, la función discriminará mucho. En cuanto a 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, por tanto, la función discriminará mucho. El p-valor del cuadro de la Lambda de Wilks certifica la significatividad del eje discriminante, con lo que su capacidad explicativa será buena (separará bien los grupos).

En cuanto a los coeficientes que nos permiten encontrar los diferentes métodos posibles para la clasificación, nos encontramos con la siguiente información:

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función
	1
Salario mensual	1.000

Los denominados coeficientes estandarizados de las funciones discriminantes canónicas son los coeficientes de la función lineal discriminante calculados sobre las variables tipificadas. En este caso, la función discriminante con coeficientes estandarizados es

$$\mathbf{D_{CE} = Z_{salario} = (salario - 2004.28) / 718.11},$$

con la variable salario tipificada. El punto de corte de la función discriminante canónica con coeficientes estandarizados es el cero. Así, una puntuación discriminante estandarizada superior a 0 para un caso concreto, clasificaría dicho caso en un grupo, y por debajo de 0 le clasificaría en el otro grupo.

Con la función estandarizada se puede apreciar la importancia que tiene cada variable en la función discriminante (en este caso, al haber un única función discriminante, y una única

variable clasificatoria, ésta es la única variable influyente, lo que se aprecia en la Matriz de estructura):

Matriz de estructura

	Función
	1
Salario mensual	1.000

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.

Los coeficientes de las funciones canónicas discriminantes se muestran en la siguiente tabla:

Coeficientes de las funciones canónicas discriminantes

	Función
	1
Salario mensual	.002
(Constante)	-4.339

Coeficientes no tipificados

Por tanto, la función discriminante es la siguiente:

$$\mathbf{D = 0.002\ salario - 4.339.}$$

Funciones en los centroides de los grupos

	Función
compra	1
No compra	-1.181
Sí compra	1.181

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

El cuadro de las Funciones en los centroides de los grupos nos da la idea de cómo discrimina los grupos la función discriminante. En este caso, la función discriminante en el centroide del grupo **No compra** toma el valor **–1.181**, mientras que en el centroide del grupo **Sí compra** toma el valor **1.181**. Esto quiere decir que, como el valor de corte en la función discriminante es el punto medio de las funciones en los centroides de los grupos, se tiene:

$$\mathbf{C = (D_1 + D_2)/2 = 0},$$

y si un caso tiene una puntuación discriminante negativa, se clasificará como **No compra** (grupo 0), y si tiene una puntuación discriminante positiva, se clasificará como **Sí Compra**.

Estadísticos de clasificación

Probabilidades previas para los grupos

compra	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
No compra	.500	10	10.000
Sí compra	.500	10	10.000
Total	1.000	20	20.000

Los dos grupos son de igual tamaño, pues tienen el mismo número de elementos, luego la probabilidad de pertenencia a priori a cada uno de los grupos es la misma (en este caso, sería equivalente haber seleccionado en las opciones de Clasificación que se calcularan las probabilidades previas de pertenencia a un grupo de forma proporcional a su tamaño).

Coeficientes de la función de clasificación

	compra	
	No compra	Sí compra
Salario mensual	.007	.012
(Constante)	-5.678	-15.929

Funciones discriminantes lineales de Fisher

De la expresión de estos coeficientes, se deduce que las funciones de clasificación son:

$$\mathbf{F_0 = 0.007 \cdot salario - 5.678}$$
$$\mathbf{F_1 = 0.012 \cdot salario - 15.929}.$$

Estas funciones de clasificación nos proporcionan la tercera opción de construcción de una regla de clasificación. En este caso, para cada caso del conjunto de análisis, se calculan los valores de sendas funciones de clasificación, y el caso se asigna al grupo cuya función de clasificación tome un mayor valor.

Los estadísticos de clasificación son los que aparecen en la tabla siguiente, en la que se aprecia que hay tres casos mal clasificados (los que están señalados con dos asteriscos):

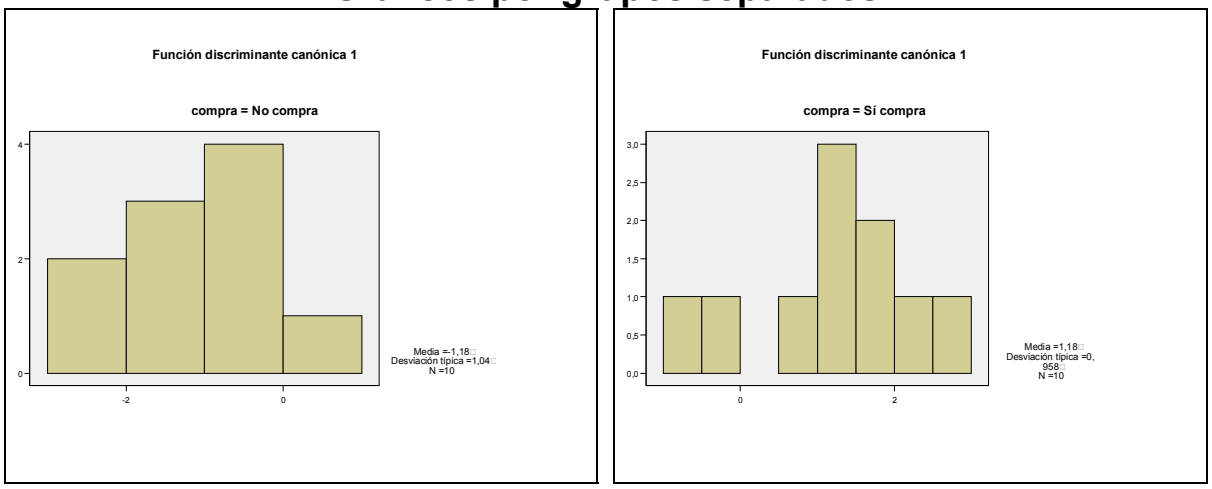
Estadísticos por casos

			Grupo mayor					Segundo grupo mayor			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	Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1
Original	1	0	0	.738	1	.881	.112	1	.119	4.115	-.847
	2	0	0	.894	1	.923	.018	1	.077	4.973	-1.049
	3	0	1**	.530	1	.787	.394	0	.213	3.010	.553
	4	0	0	.952	1	.950	.004	1	.050	5.872	-1.242
	5	0	0	.676	1	.978	.174	1	.022	7.731	-1.599
	6	0	0	.125	1	.998	2.348	1	.002	15.173	-2.714
	7	0	0	.651	1	.849	.204	1	.151	3.651	-.729
	8	0	0	.073	1	.999	3.209	1	.001	17.257	-2.973
	9	0	0	.595	1	.823	.283	1	.177	3.352	-.650
	10	0	0	.538	1	.792	.380	1	.208	3.050	-.565
	11	1	1	.882	1	.959	.022	0	.041	6.307	1.330
	12	1	1	.981	1	.945	.001	0	.055	5.696	1.205
	13	1	1	.404	1	.992	.696	0	.008	10.221	2.016
	14	1	0**	.573	1	.811	.318	1	.189	3.237	-.618
	15	1	1	.777	1	.893	.081	0	.107	4.322	.898
	16	1	1	.841	1	.963	.040	0	.037	6.573	1.382
	17	1	1	.741	1	.973	.109	0	.027	7.255	1.512
	18	1	1	.160	1	.998	1.975	0	.002	14.199	2.587
	19	1	0**	.316	1	.604	1.006	1	.396	1.849	-.178
	20	1	1	.618	1	.981	.249	0	.019	8.188	1.680

**. Caso mal clasificado

Los gráficos de los grupos por separado son (conviene recordar que el cero es la puntuación de corte discriminante):

Gráficos por grupos separados



Por último, la herramienta que permite analizar la bondad de la clasificación realizada es, la matriz de confusión, que se presenta en la siguiente tabla, bajo el título de Resultados de la clasificación:

Resultados de la clasificación[a]

			Grupo de pertenencia pronosticado		
		compra	No compra	Sí compra	Total
Original	Recuento	No compra	9	1	10
		Sí compra	2	8	10
	%	No compra	90.0	10.0	100.0
		Sí compra	20.0	80.0	100.0

a. Clasificados correctamente el 85.0% de los casos agrupados originales.

Como se ve, la clasificación obtenida es bastante buena (únicamente el 15% de los casos no están bien clasificados).

Se puede hacer lo mismo con la variable **edad**, y se comprueba que el error de clasificación es ligeramente superior (el 20% de las observaciones están mal clasificadas).

Vamos a pasar ahora a comprobar cómo se puede mejorar la clasificación si utilizamos las dos variables posibles de clasificación:

Discriminante

Se muestran los estadísticos descriptivos por grupo (hay dos grupos) para cada variable:

Estadísticos de grupo

				N válido (según lista)	
compra		Media	Desv. típ.	No ponderados	Ponderados
No compra	Salario mensual	1458.5390	480.60643	10	10.000
	Edad (en años)	30.0000	4.44722	10	10.000
Sí compra	Salario mensual	2550.0300	442.51025	10	10.000
	Edad (en años)	39.7000	5.22919	10	10.000
Total	Salario mensual	2004.2845	718.10950	20	20.000
	Edad (en años)	34.8500	6.86160	20	20.000

Análisis 1
Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

compra	Rango	Logaritmo del determinante
No compra	2	15.331
Sí compra	2	15.457
Intra-grupos combinada	2	15.426

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

Resultados de la prueba

M de Box		.578
F	Aprox.	.169
	gl1	3
	gl2	58320.000
	Sig.	.917

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2.854[a]	100.0	100.0	.861

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Se observa que el poder de discriminación con dos variables clasificadoras es mayor (en este caso, la correlación canónica es 0.861).

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	.260	22.933	2	.000

Los coeficientes estandarizados de las funciones discriminantes canónicas son:

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función
	1
Salario mensual	.784
Edad (en años)	.677

y, por tanto, la función discriminante canónica con coeficientes estandarizados, es (el punto de corte es el cero):

$$\mathbf{D_{CE} = 0.784\ Z_{salario} + 0.677\ Z_{edad}.}$$

En la matriz de estructura se presentan las correlaciones intra-grupo entre las variables discriminantes y las funciones discriminantes canónicas tipificadas, y es muy útil para analizar qué variable tiene más importancia en la discriminación:

Matriz de estructura

	Función
	1
Salario mensual	.737
Edad (en años)	.623

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.

Los coeficientes de la función canónica discriminante (no tipificados) son:

Coeficientes de las funciones canónicas discriminantes

	Función
	1
Salario mensual	.002
Edad (en años)	.140
(Constante)	-8.263

Coeficientes no tipificados

y la función canónica discriminante tiene la siguiente expresión:

$$\mathbf{D = 0.002\ salario + 0.140\ edad - 8.263}.$$

Otra vez el punto de corte discriminante de esta función será cero, ya que la función discriminante tiene el mismo valor, pero cambiado de signo, en los centroides de los dos grupos analizados:

Funciones en los centroides de los grupos

	Función
compra	1
No compra	-1.603
Sí compra	1.603

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

Estadísticos de clasificación

Probabilidades previas para los grupos

compra	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
No compra	.500	10	10.000
Sí compra	.500	10	10.000
Total	1.000	20	20.000

Los coeficientes de las funciones de clasificación (en este caso hay dos funciones de clasificación, una para cada grupo):

Coeficientes de la función de clasificación

	compra	
	No compra	Sí compra
Salario mensual	.008	.013
Edad (en años)	1.324	1.771
(Constante)	-26.236	-52.720

Funciones discriminantes lineales de Fisher

Así:

$$\mathbf{F_0 = 0.008\ salario + 1.324\ edad - 26.236}$$
$$\mathbf{F_1 = 0.013\ salario + 1.771\ edad - 52.720}$$

y

$$\mathbf{F_1 - F_0 = D\text{-}C = u_1 X_1 + u_2 X_2 + ... + u_p X_p - C}$$

será la función discriminante; en este caso:

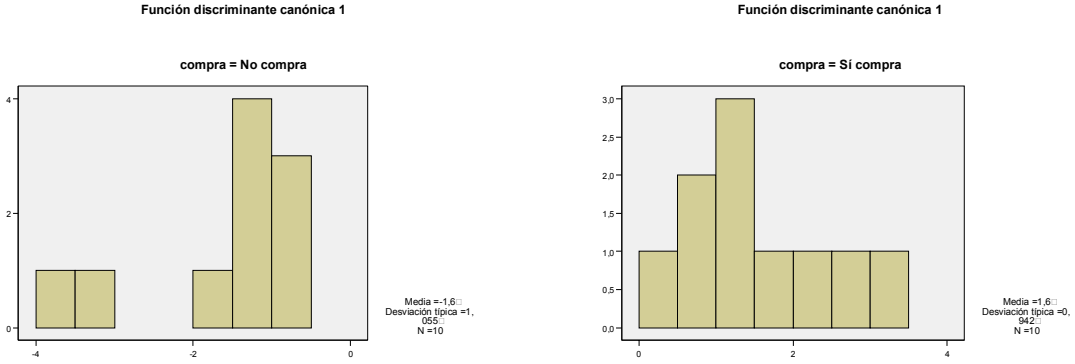
$$\mathbf{F_1 - F_0 = 0.005\ salario + 0.447\ edad - 26.484}$$

Para analizar los resultados del procedimiento de análisis discriminante, conviene observar los estadísticos de clasificación:

Estadísticos por casos

			Grupo mayor					Segundo grupo mayor			Puntuaciones discriminantes
				P(D>d | G=g)							
	Número de caso	Grupo real	Grupo pronosticado	p	gl	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1
Original	1	0	0	.337	1	.887	.921	1	.113	5.042	-.643
	2	0	0	.753	1	.998	.099	1	.002	12.388	-1.917
	3	0	0	.601	1	.970	.273	1	.030	7.197	-1.080
	4	0	0	.610	1	.971	.261	1	.029	7.260	-1.092
	5	0	0	.711	1	.981	.137	1	.019	8.036	-1.232
	6	0	0	.029	1	1.000	4.742	1	.000	28.974	-3.780
	7	0	0	.440	1	.935	.597	1	.065	5.916	-.830
	8	0	0	.123	1	1.000	2.382	1	.000	22.548	-3.146
	9	0	0	.890	1	.991	.019	1	.009	9.409	-1.465
	10	0	0	.446	1	.937	.581	1	.063	5.968	-.840
	11	1	1	.564	1	.999	.333	0	.001	14.303	2.179
	12	1	1	.952	1	.995	.004	0	.005	10.665	1.663
	13	1	1	.210	1	1.000	1.572	0	.000	19.882	2.856
	14	1	1	.801	1	.987	.063	0	.013	8.722	1.351
	15	1	1	.749	1	.984	.102	0	.016	8.323	1.282
	16	1	1	.359	1	.900	.841	0	.100	5.236	.686
	17	1	1	.906	1	.992	.014	0	.008	9.533	1.485
	18	1	1	.089	1	1.000	2.895	0	.000	24.073	3.304
	19	1	1	.193	1	.723	1.697	0	.277	3.619	.300
	20	1	1	.494	1	.950	.467	0	.050	6.358	.919

Gráficos por grupos separados



En este caso, no se ha encontrado ningún caso mal clasificado según lafunción lineal discriminante obtenida, con lo que se puede decir que la regla de clasificación definida consigue unos resultados óptimos.

Resultados de la clasificación[a]

			Grupo de pertenencia pronosticado		
		compra	No compra	Sí compra	Total
Original	Recuento	No compra	10	0	10
		Sí compra	0	10	10
	%	No compra	100.0	.0	100.0
		Sí compra	.0	100.0	100.0

a. Clasificados correctamente el 100.0% de los casos agrupados originales.

Para ilustrar el análisis discriminante con más de dos grupos vamos a analizar los datos sobre el modelo de coche elegido por 60 clientes de este mismo concesionario (fichero **Gama_coche.sav**). Las variables que aparecen en dicho fichero son las siguientes:

- **modelo:** Gama del modelo elegido por el comprador de un vehículo en el concesionario. Esta variable es categórica (con una escala de medida nominal), y presenta los siguientes valores:
  - **1:** Gama alta.
  - **2:** Gama media.
  - **3:** Gama básica.
- **salario:** Salario mensual.
- **edad:** Edad (en años).

Pues bien, con la información de este fichero se va a realizar el análisis discriminante considerando las siguientes opciones:

- Como variable de agrupación la variable **modelo**, definiendo su rango entre 1 y 3.
- Como variables independientes el salario mensual y la edad.
- Como **Estadísticos**, pedimos que nos muestre las *medias*, el *contraste M de Box* y los *coeficientes de la función de Fisher y no tipificados*.
- En las opciones de **clasificar**, elegimos que calcule las probabilidades previas según el *tamaño de grupo* (conviene fijarse que hay diferencias importantes de tamaño entre los grupos que define esta variable modelo), y que se muestren los *resultados para cada caso*, así como la *tabla de resumen* y todos los gráficos posibles.

Los resultados que proporciona el procedimiento de discriminante son los siguientes:

Discriminante

Primero aparecerán las medidas descriptivas para cada variable en cada grupo de partida:

Estadísticos de grupo

modelo		Media	Desv. típ.	N válido (según lista)	
				No ponderados	Ponderados
Gama alta	Salario mensual	3452.9310	393.69383	10	10.000
	Edad (en años)	45.6000	5.75809	10	10.000
Gama media	Salario mensual	2568.5200	464.29635	20	20.000
	Edad (en años)	26.7000	3.32613	20	20.000
Gama básica	Salario mensual	1748.4243	371.39246	30	30.000
	Edad (en años)	35.2667	5.00988	30	30.000
Total	Salario mensual	2305.8740	751.83846	60	60.000
	Edad (en años)	34.1333	7.91366	60	60.000

La prueba de la M de Box sobre la igualdad de matrices de varianzas-covarianzas en los tres grupos resulta aceptar la hipótesis nula, según se muestra en estas tablas (el p-valor es 0.151):

Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

modelo	Rango	Logaritmo del determinante
Gama alta	2	15.452
Gama media	2	14.391
Gama básica	2	15.055
Intra-grupos combinada	2	15.074

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

Resultados de la prueba

M de Box		10.090
F	Aprox.	1.571
	gl1	6
	gl2	7439.868
	Sig.	.151

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Resumen de las funciones canónicas discriminantes

Si consideramos las variables tipificadas, las dos funciones discriminantes canónicas tendrían los coeficientes:

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función	
	1	2
Salario mensual	.846	-.557
Edad (en años)	.686	.745

y su matriz de estructura será:

Matriz de estructura

	Función	
	1	2
Salario mensual	.736*	-.677
Edad (en años)	.550	.835*

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.

lo cual nos dice qué variable tiene más importancia en la discriminación en una función discriminante u otra (en este caso, el salario mensual en la primera y la edad en la segunda).

Los coeficientes de las funciones canónicas discriminantes son:

Coeficientes de las funciones canónicas discriminantes

	Función	
	1	2
Salario mensual	.002	-.001
Edad (en años)	.147	.160
(Constante)	-9.810	-2.314

Coeficientes no tipificados

Por tanto, las dos funciones canónicas discriminantes son:

$$\mathbf{D_1 = 0.002\ salario + 0.147\ edad - 9.810}$$
$$\mathbf{D_2 = -\ 0.001\ salario + 0.160\ edad - 2.314}$$

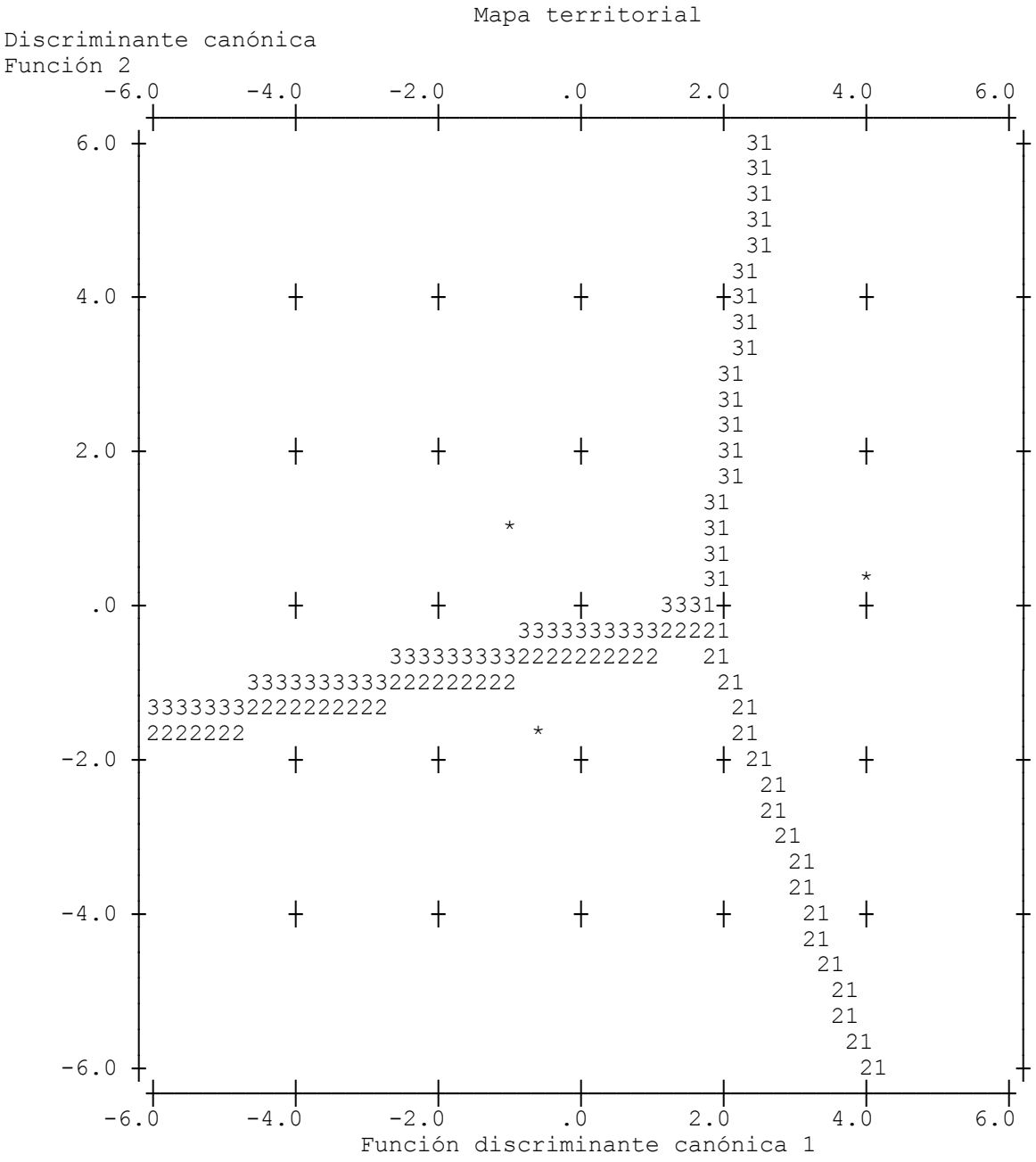
Las puntuaciones de los centroides de ambos grupos con respecto a las funciones discriminantes son las siguientes (conviene darse cuenta de que en este caso no existe un único punto de corte discriminante, pues el conjunto de datos está dividido en tres grupos).

Funciones en los centroides de los grupos

	Función	
modelo	1	2
Gama alta	4.068	.269
Gama media	-.550	-1.549
Gama básica	-.989	.943

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

Sin embargo, SPSS muestra el mapa territorial, que delimita, en el plano definido por los valores que toman las dos funciones discriminantes canónicas (con valores de las variables no estandarizados), las áreas que se asignan a cada grupo. En concreto, el área situada a la derecha de la función discriminante 1 es la correspondiente al grupo 1, que elige el modelo de gama alta. La parte que queda a la izquierda, se reparte entre los grupos 2 y 3, correspondiendo la parte superior al grupo 3 (individuos que eligen la gama básica), y la parte inferior al grupo 2.



Símbolos usados en el mapa territorial

Símbol	Grupo	Etiqu
1	1	Gama alta
2	2	Gama media
3	3	Gama básica
*		Indica un centroide de grupo

En cuanto a los estadísticos de clasificación, se tiene:

Estadísticos de clasificación

Probabilidades previas para los grupos

modelo	Previas	Casos utilizados en el análisis	
		No ponderados	Ponderados
Gama alta	.167	10	10.000
Gama media	.333	20	20.000
Gama básica	.500	30	30.000
Total	1.000	60	60.000

Ahora tenemos que calcular el valor de tres funciones de clasificación, y clasificaremos a cada individuo en aquél grupo cuya función de clasificación resulte tomar el mayor valor:

Coeficientes de la función de clasificación

	modelo		
	Gama alta	Gama media	Gama básica
Salario mensual	.025	.018	.014
Edad (en años)	2.458	1.487	1.821
(Constante)	-101.416	-44.258	-44.899

Funciones discriminantes lineales de Fisher

De esta forma, las funciones de clasificación resultan:

$$\mathbf{F_1 = 0.025\ salario + 2.458\ edad - 101.416}$$
$$\mathbf{F_2 = 0.018\ salario + 1.487\ edad - 44.258}$$
$$\mathbf{F_3 = 0.014\ salario + 1.821\ edad - 44.899}$$

y nos permiten clasificar a un caso en aquél grupo cuya función de clasificación resulte ser mayor.

Hay cinco casos mal clasificados, como se muestra en la siguiente tabla de Estadísticos por caso, comprobándose como las probabilidades de pertenencia son mayores para la pertenencia al grupo mayor, y también que las puntuaciones discriminantes son las que sitúan a cada caso en el mapa territorial mostrado anteriormente:

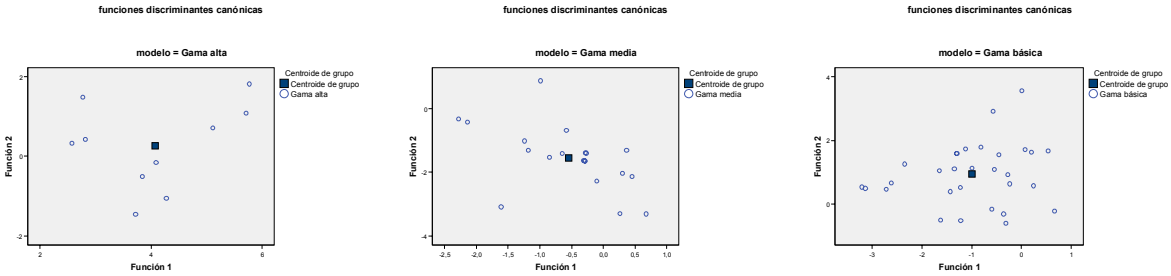
Estadísticos por casos

			Grupo mayor					Segundo grupo mayor			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	Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1	Función 2
Original	1	1	1	.325	2	.979	2.249	3	.013	13.037	2.569	.330
	2	1	1	.213	2	.999	3.095	2	.001	18.236	3.719	-1.456
	3	1	1	.070	2	1.000	5.310	3	.000	46.473	5.772	1.820
	4	1	1	.206	2	.989	3.160	3	.011	14.421	2.770	1.483
	5	1	1	.525	2	1.000	1.290	3	.000	37.299	5.114	.710
	6	1	1	.185	2	1.000	3.380	3	.000	44.949	5.714	1.087
	7	1	1	.721	2	1.000	.656	2	.000	20.389	3.844	-.510
	8	1	1	.410	2	1.000	1.784	2	.000	23.536	4.275	-1.051
	9	1	1	.913	2	1.000	.181	2	.000	23.459	4.089	-.157
	10	1	1	.448	2	.994	1.604	3	.004	14.705	2.811	.428
	11	2	2	.968	2	.957	.065	3	.043	7.086	-.309	-1.631
	12	2	2	.963	2	.958	.075	3	.042	7.160	-.293	-1.641
	13	2	2	.683	2	.762	.762	3	.238	3.900	-1.242	-1.016
	14	2	2	.985	2	.917	.030	3	.083	5.634	-.649	-1.406
	15	2	2	.156	2	.999	3.717	3	.001	19.535	.266	-3.295
	16	2	2	.618	2	.988	.964	3	.012	10.528	.303	-2.034
	17	2	2	.685	2	.649	.756	3	.351	2.799	-.581	-.680
	18	2	2	.962	2	.959	.078	3	.041	7.186	-.288	-1.644
	19	2	2	.953	2	.925	.095	3	.075	5.928	-.286	-1.388
	20	2	2	.100	2	1.000	4.601	3	.000	20.843	.676	-3.308
	21	2	2	.957	2	.933	.088	3	.067	6.153	-.846	-1.534
	22	2	3**	.193	2	.731	3.295	2	.269	4.485	-2.283	-.330
	23	2	3**	.998	2	.969	.005	2	.031	6.065	-.989	.875
	24	2	2	.637	2	.930	.904	3	.070	6.897	.368	-1.305
	25	2	2	.950	2	.928	.102	3	.072	6.014	-.267	-1.401
	26	2	2	.509	2	.991	1.351	3	.009	11.544	.454	-2.133
	27	2	2	.795	2	.873	.459	3	.127	5.117	-1.185	-1.311
	28	2	3**	.202	2	.671	3.196	2	.329	3.809	-2.144	-.422
	29	2	2	.175	2	.998	3.491	3	.002	16.625	-1.611	-3.087
	30	2	2	.693	2	.992	.734	3	.008	11.168	-.104	-2.280
	31	3	3	.421	2	.994	1.730	2	.006	11.049	.075	1.716
	32	3	3	.389	2	.992	1.891	2	.008	10.690	.200	1.634
	33	3	3	.773	2	.995	.515	2	.005	10.413	-1.296	1.591
	34	3	3	.800	2	.985	.446	2	.015	7.965	-1.648	1.051
	35	3	3	.132	2	1.000	4.057	2	.000	19.910	-.572	2.913
	36	3	3	.436	2	.895	1.659	2	.104	5.144	.246	.575
	37	3	3	.020	2	1.000	7.847	2	.000	26.423	.006	3.562
	38	3	3	.720	2	.992	.656	2	.008	9.600	-.451	1.548
	39	3	3	.376	2	.993	1.956	2	.007	11.119	-2.351	1.258
	40	3	3	.201	2	.960	3.207	2	.040	8.761	-2.716	.469
	41	3	3	.505	2	.666	1.368	2	.334	1.936	-.593	-.158
	42	3	3	.256	2	.974	2.724	2	.026	9.139	-2.615	.660
	43	3	3	.373	2	.550	1.973	2	.450	1.565	-.359	-.312
	44	3	3	.724	2	.996	.647	2	.004	11.116	-1.125	1.736
	45	3	3	.079	2	.972	5.088	2	.028	11.405	-3.207	.536
	46	3	3	.768	2	.995	.529	2	.005	10.468	-1.306	1.597
	47	3	2**	.196	2	.501	3.256	3	.496	4.086	.667	-.217
	48	3	3	.780	2	.919	.496	2	.081	4.535	-1.428	.392
	49	3	3	.775	2	.962	.509	2	.038	6.164	-.276	.919
	50	3	3	.336	2	.518	2.182	2	.482	1.515	-1.220	-.516
	51	3	3	.286	2	.567	2.506	2	.433	2.236	-1.624	-.507
	52	3	3	.895	2	.978	.222	2	.022	6.986	-.543	1.095
	53	3	3	.924	2	.985	.158	2	.015	7.710	-1.348	1.111
	54	3	3	.718	2	.925	.662	2	.075	4.868	-.236	.635
	55	3	3	.089	2	.968	4.833	2	.032	10.872	-3.141	.492
	56	3	3	.690	2	.996	.742	2	.004	11.187	-.812	1.786
	57	3	3	.888	2	.934	.238	2	.066	4.742	-1.231	.520
	58	3	2**	.620	2	.632	.955	3	.368	2.844	-.310	-.601
	59	3	3	.240	2	.990	2.858	2	.009	11.532	.538	1.668
	60	3	3	.982	2	.983	.036	2	.017	7.379	-.990	1.132

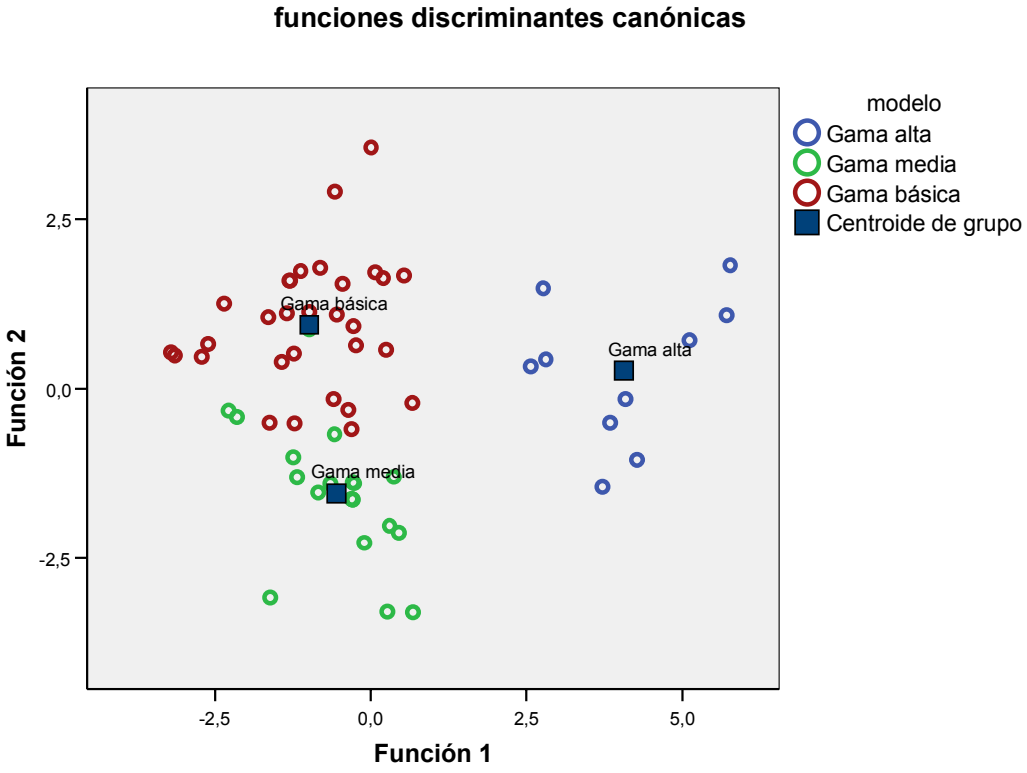
**. Caso mal clasificado

Representando los grupos por separado, tendremos tres gráficos (uno por cada grupo), en los cuales se ven algunos puntos que se alejan demasiado del centroide del grupo al que pertenecen.

Gráficos por grupos separados



Si ahora los dibujamos todos juntos y ampliamos el gráfico para apreciar mejor los distintos grupos, tenemos:



Analizando la matriz de confusión, comprobamos que hay cinco casos mal clasificados, que representan el 8.3% de los mismos:

Resultados de la clasificación[a]

			Grupo de pertenencia pronosticado			
		modelo	Gama alta	Gama media	Gama básica	Total
Original	Recuento	Gama alta	10	0	0	10
		Gama media	0	17	3	20
		Gama básica	0	2	28	30
	%	Gama alta	100.0	.0	.0	100.0
		Gama media	.0	85.0	15.0	100.0
		Gama básica	.0	6.7	93.3	100.0

a. Clasificados correctamente el 91.7% de los casos agrupados originales.