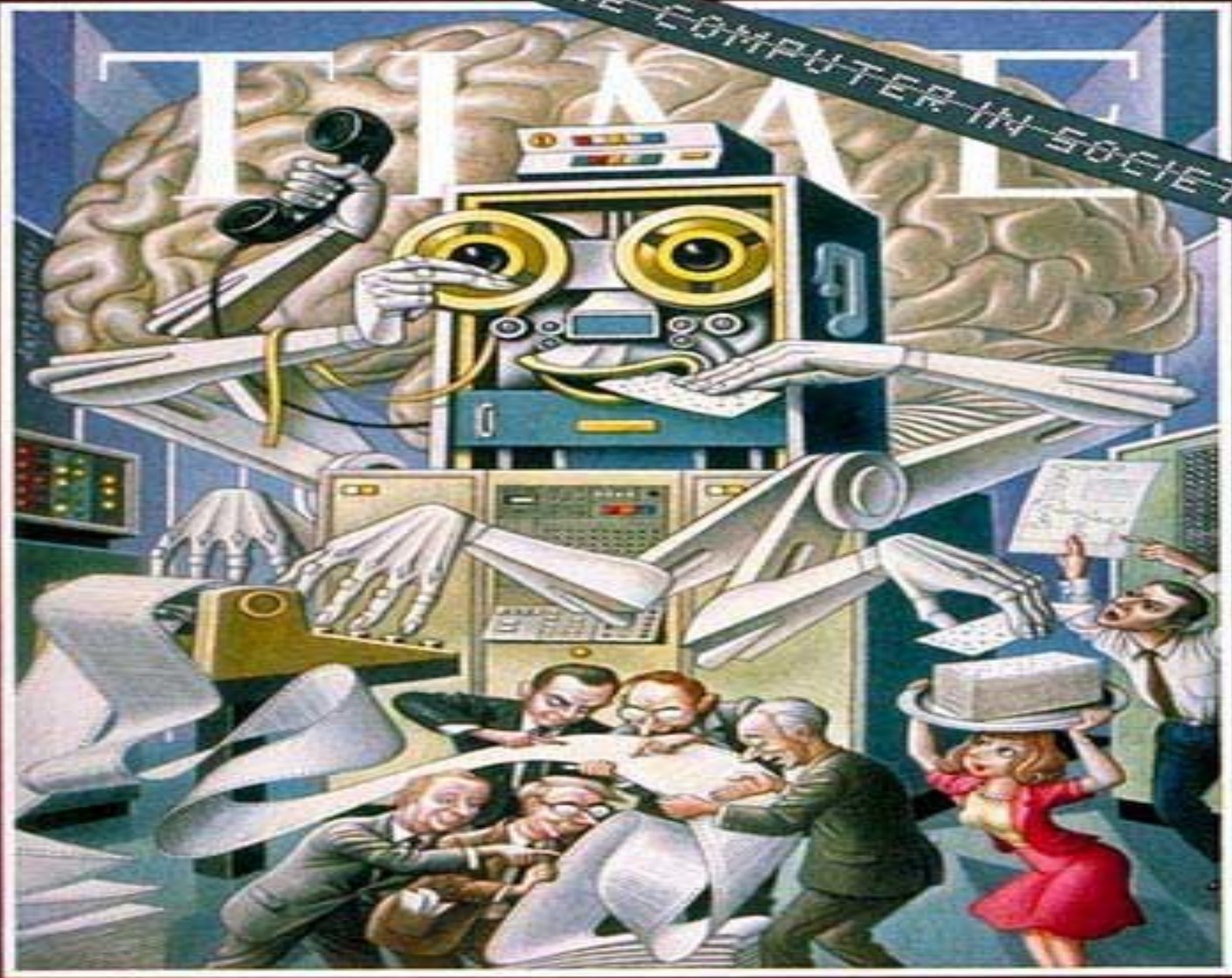


THE COMPUTER IN SOCIETY



# Análisis de Correlación Canónica y Análisis discriminante



# Análisis de Regresión Múltiple

## ***TÉCNICAS DE DEPENDENCIA***

Variable/s Dependiente/s	Variable/s Independiente/s	Técnica
1 Nivel de medición Continuo	1 Nivel de medición Continuo	REGRESIÓN SIMPLE
1 Nivel de medición Continuo	> 1 Nivel de medición Continuo	REGRESIÓN MÚLTIPLE
> 1 Nivel de medición Continuo	> 1 Nivel de medición Continuo	CORRELACIÓN CANÓNICA
1 Nivel de medición Nominal	> 1 Nivel de medición Continuo	ANÁLISIS DISCRIMINANTE
1 Nivel de medición Nominal (v. Dicotómica)	> 1 Nivel de medición Continuo	REGRESIÓN LOGÍSTICA
1 Nivel de medición Continuo	1 Nivel de medición Nominal	ANOVA Unidireccional
1 Nivel de medición Continuo	> 1 Nivel de medición Nominal	ANOVA Factorial o Bidireccional
> 1 Nivel de medición Continuo	1 Nivel de medición Nominal	MANOVA Unidireccional
> 1 Nivel de medición Continuo	1 Nivel de medición Nominal (v. Dicotómica)	T de Hotelling
> 1 Nivel de medición Continuo	> 1 Nivel de medición Nominal	MANOVA Factorial o Bidireccional

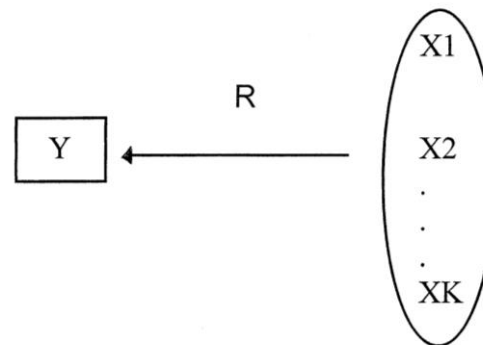
# Análisis de Correlación canónica

- Es una extensión del Análisis de Regresión Múltiple al caso de más de una variable dependiente

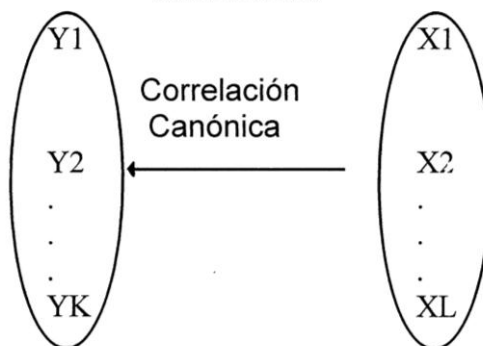


# Análisis de Correlación canónica

## CONCEPTO DE CORRELACIÓN MÚLTIPLE



## CONCEPTO DE CORRELACIÓN CANÓNICA



# Análisis de Correlación canónica

## VARIABLES CRITERIO

$$q = [V_1, V_2, \dots, V_q]$$

## VARIABLES PREDICTORAS

$$p = [U_1, U_2, \dots, U_p]$$

Estos coeficientes deben hacer máxima la correlación entre W y Z:

$$W = V_1 Y_1 + V_2 Y_2 + \dots + V_q Y_q$$

$$Z = U_1 X_1 + U_2 X_2 + \dots + U_p X_p$$

## VARIABLES CRITERIO

$$W_1 = V_{11} Y_1 + V_{12} Y_2 + \dots + V_{1q} Y_q$$

$$W_2 = V_{21} Y_1 + V_{22} Y_2 + \dots + V_{2q} Y_q$$

## VARIABLES PREDICTORAS

$$Z_1 = U_{11} X_1 + U_{12} X_2 + \dots + U_{1p} X_p$$

$$Z_2 = U_{21} X_1 + U_{22} X_2 + \dots + U_{2p} X_p$$



# Análisis de Correlación canónica

- Tal que, las variables canónicas de cada grupo no están relacionadas entre sí y
- cada una de las variables canónicas de cada grupo, únicamente se relaciona con la correspondiente del otro grupo.
- Es decir,  $Z_1$  no se relaciona con  $Z_2$ , ni  $W_1$  con  $W_2$  y  $Z_1$  únicamente se asocia a  $W_1$  y  $Z_2$  a  $W_2$



# Análisis de Correlación canónica

El nº de variables canónicas para cada grupo es = al nº de variables del grupo más pequeño

COEFFICIENTS FOR CANONICAL VARIABLES FOR FIRST SET OF VARIABLES					
		CNVRF1 <sub>1</sub>	CNVRF2 <sub>2</sub>	CNVRF3 <sub>3</sub>	CNVRF4 <sub>4</sub>
FUMA1	3	0.037854	-0.976451	-0.965493	-0.900841
FUMA2	5	-1.093216	-0.646536	1.341055	0.182999
FUMA3	7	1.191154	-0.173899	-0.033369	1.451945
FUMA4	12	-0.704060	1.285695	-0.660196	-0.214955

STANDARDIZED COEFFICIENTS FOR CANONICAL VARIABLES FOR FIRST SET OF VARIABLES (THESE ARE THE COEFFICIENTS FOR THE STANDARDIZED VARIABLES - MEAN ZERO, STANDARD DEVIATION ONE.)					
		CNVRF1 <sub>1</sub>	CNVRF2 <sub>2</sub>	CNVRF3 <sub>3</sub>	CNVRF4 <sub>4</sub>
FUMA1	3	0.043	-1.104	-1.092	-1.019
FUMA2	5	-1.160	-0.686	1.423	0.194
FUMA3	7	1.383	-0.202	-0.039	1.686
FUMA4	12	-0.898	1.641	-0.842	-0.274

COEFFICIENTS FOR CANONICAL VARIABLES FOR SECOND SET OF VARIABLES					
		CNVR1 <sub>1</sub>	CNVR2 <sub>2</sub>	CNVR3 <sub>3</sub>	CNVR4 <sub>4</sub>
CONCENTR	1	-0.441692	0.745510	-0.470381	-0.163811
MOLESTIA	2	0.801410	0.461495	-0.605503	-0.739549
SUENO	4	-0.250790	0.581216	-0.685988	0.615867
TENSION	6	-0.692552	-0.380734	0.421877	0.448775
ACTIVIDA	8	0.140028	0.204741	1.501592	-0.685341
IRRITABL	9	0.090000	-0.795294	0.425982	1.137464
CANSANCI	10	-0.327905	-0.616257	-0.246355	0.172116
SATISFAC	11	-0.402041	-0.595032	-0.971468	-0.795208

STANDARDIZED COEFFICIENTS FOR CANONICAL VARIABLES FOR SECOND SET OF VARIABLES (THESE ARE THE COEFFICIENTS FOR THE STANDARDIZED VARIABLES - MEAN ZERO, STANDARD DEVIATION ONE.)					
		CNVR1 <sub>1</sub>	CNVR2 <sub>2</sub>	CNVR3 <sub>3</sub>	CNVR4 <sub>4</sub>
CONCENTR	1	-0.474	0.800	-0.505	-0.176
MOLESTIA	2	0.781	0.450	-0.590	-0.721
SUENO	4	-0.257	0.595	-0.702	0.630
TENSION	6	-0.687	-0.378	0.418	0.445
ACTIVIDA	8	0.143	0.208	1.529	-0.698
IRRITABL	9	0.070	-0.622	0.333	0.890
CANSANCI	10	-0.313	-0.588	-0.235	0.164
SATISFAC	11	-0.339	-0.501	-0.818	-0.670

Figura 148. Coeficiente de las variables canónicas (BMDP)





# Análisis de Correlación canónica

Interpretación de resultados:

**Eigenvalue o valor propio:** índice de la varianza del grupo de las variables criterio o dependientes que viene explicada por el conjunto de las predictoras

**Correlación Canónica:** indica la relación entre los dos grupos de Variables. El coeficiente Eta denota esta correlación

$$\text{Eta}^2 = \frac{\text{SC entre}}{\text{SC Total}}$$



# Análisis de Correlación canónica

## Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon. Cor.	Squared Cor.
1	.37511	60.48650	60.48650	.52229	.27278
2	.16453	26.53017	87.01667	.37588	.14128
3	.06134	9.89039	96.90706	.24040	.05779
4	.01918	3.09294	100.00000	.13719	.01882

## Dimension Reduction Analysis

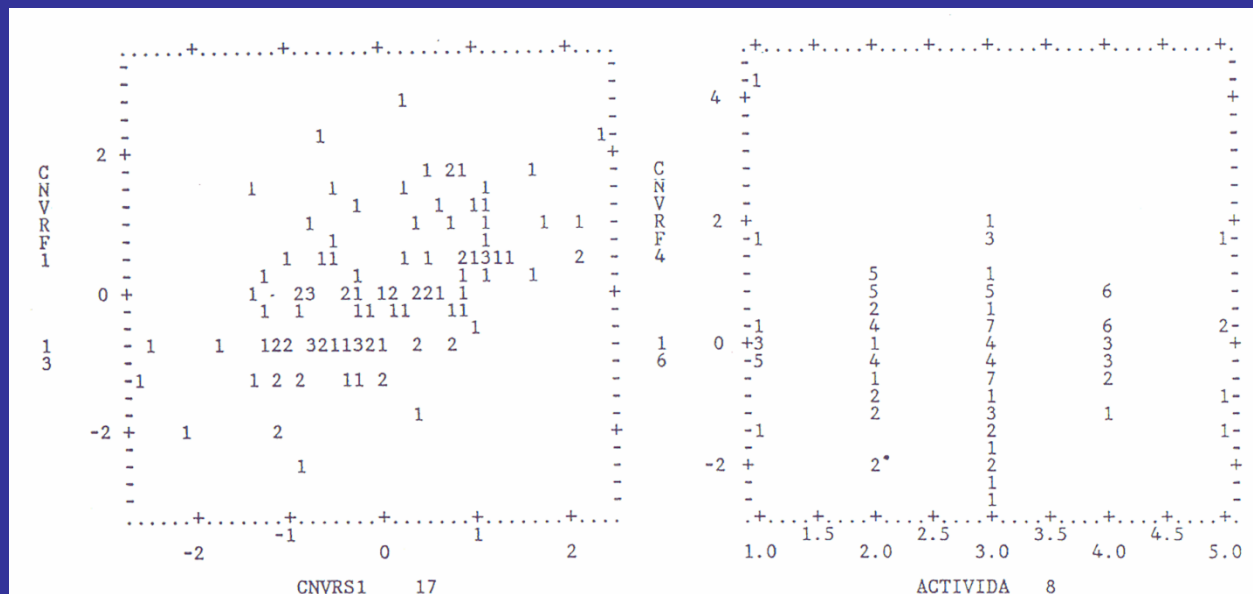
Roots	Wilks Lambda	F	Hypoth. DF	Error DF	Sig. of F
* 1 TO 4	.57731	1.82225	32.00	363.00	.005
** 2 TO 4	.79386	1.13539	21.00	284.82	.310
*** 3 TO 4	.92448	.66741	12.00	200.00	.781
4 TO 4	.98118	.38745	5.00	101.00	.856

\* Todas las correlaciones canónicas excepto la 1ª son = 0

\*\* Todas las correlaciones canónicas excepto la 1ª y la 2ª son = 0...



# Análisis de Correlación canónica



# Análisis Discriminante

Definición: Dada una variable dependiente cualitativa y un conjunto de independientes métricas se trata de obtener una combinación lineal (función discriminante) de éstas últimas con el objeto de entender las diferencias entre los grupos y poder realizar futuras predicciones de otras unidades de análisis, respecto a su grupo de pertenencia



# Análisis Discriminante

El objetivo del Análisis Discriminante es determinar si en función de las variables disponibles los grupos quedan suficientemente discriminados. Esta sería la explicación de la diferencia entre los grupos.



# Análisis Discriminante

## Pasos en el Análisis Discriminante

- Selección de las variables: a partir del mínimo

**Lambda de Wilks** = SC Intra/CS total

**F** de entrada 3,84 **F** de salida 2,71

**Tolerancia** nivel mínimo de tolerancia

- Obtención de las funciones discriminantes: estas son variables canónicas, es decir, combinaciones lineales de las variables originales. El número de funciones discriminantes es  $c = \min(K-1, p)$  y todas ellas deben estar incorrelacionadas



# Análisis Discriminante

## ■ Regla de Bayés

Cálculo de las puntuaciones discriminantes y las probabilidades de pertenencia a grupos que permita clasificar a las unidades de análisis

## ■ Matriz de confusión

En esta se indican los casos que pertenecen a cada grupo antes y después de establecerlos a partir de la función discriminante, y se indica el % de los Correctamente clasificados, de modo que se pueda comprobar la eficacia de la función discriminante



# Análisis Discriminante

```
----- DISCRIMINANT ANALYSIS -----  
On groups defined by REAPARIC Tiempo de reaparición de los síntomas  
  
312 (Unweighted) cases were processed.  
46 of these were excluded from the analysis.  
4 had missing or out-of-range group codes.  
42 were excluded by the select= variable.  
266 (Unweighted) cases will be used in the analysis.  
  
Number of cases by group  
  
      Number of cases  
REAPARIC Unweighted Weighted Label  
      1          34      34,0 4 Meses  
      2         126     126,0 8 Meses  
      3          86      86,0 12 Meses  
      4          20      20,0 16 Meses  
  
Total          266     266,0  
  
Stepwise variable selection  
Selection rule: minimize Wilks' Lambda  
Maximum number of steps..... 10  
Minimum tolerance level..... ,00100  
Minimum F to enter..... 3,84000  
Maximum F to remove..... 2,71000  
  
Canonical Discriminant Functions  
  
Maximum number of functions..... 3  
Minimum cumulative percent of variance... 100,00  
Maximum significance of Wilks' Lambda.... 1,0000  
  
Prior probabilities  
  
Group   Prior   Label  
      1   ,12782   4 Meses  
      2   ,47368   8 Meses  
      3   ,32331  12 Meses  
      4   ,07519  16 Meses  
  
Total   1,00000
```





# Análisis Discriminante

Variables not in the Analysis after Step 0

Variable	Tolerance	Minimum Tolerance	F to Enter	Wilks' Lambda
RESPUEST	1,0000000	1,0000000	119,2211726	,4228101
ALCOHOL	1,0000000	1,0000000	44,4888130	,6625088
CAFE	1,0000000	1,0000000	8,2001686	,9141645
ANTIACID	1,0000000	1,0000000	10,8238502	,8897294
TABACO	1,0000000	1,0000000	34,3737504	,7175698

At step 1, RESPUEST was included in the analysis.

Wilks' Lambda	Signif. Between Groups	Degre of Freedom	Equivalent F
,42281	,0000	1 3	119,22117
			262,0

Variables in the Analysis after Step 1

Variable	Tolerance	F to Remove	Wilks' Lambda
RESPUEST	1,0000000	119,2212	

Variables not in the Analysis after Step 1

Variable	Tolerance	Minimum Tolerance	F to Enter	Wilks' Lambda
ALCOHOL	,6271971	,6271971	121,9304489	,1760609
CAFE	,8596197	,8596197	23,2331063	,3336972
ANTIACID	,8781665	,8781665	24,3424190	,3303726
TABACO	,9846206	,9846206	30,3845512	,3133673

At step 2, ALCOHOL was included in the analysis.

Wilks' Lambda	Signif. Between Groups	Degrees of Freedom	Equivalent F
,17606	,0000	2 3	120,34225
			522,0

Variables in the Analysis after Step 2

Variable	Tolerance	F to Remove	Wilks' Lambda
RESPUEST	,6271971	240,3769	,6625088
ALCOHOL	,6271971	121,9304	,4228101

Variables not in the Analysis after Step 2

Variable	Tolerance	Minimum Tolerance	F to Enter	Wilks' Lambda
CAFE	,7352454	,5364509	,1870899	,1756816
ANTIACID	,7623120	,5444524	,8307264	,1743893
TABACO	,8091121	,5153993	55,5711445	,1072753



# Análisis Discriminante

At step 3, TABACO was included in the analysis.

		Degrees of Freedom	Signif.	Between Groups
Wilks' Lambda	,10728	3 3	262,0	
Approximate F	105,65657	9	632,9	,0000

Variables in the Analysis after Step 3

Variable	Tolerance	F to Remove	Wilks' Lambda
RESPUEST	,5430811	272,2577	,4442739
ALCOHOL	,5153993	166,4996	,3133673
TABACO	,8091121	55,5711	,1760609

Variables not in the Analysis after Step 3

Variable	Tolerance	Minimum Tolerance	F to Enter	Wilks' Lambda
CAFE	,7274911	,4644436	,8019598	,1062880
ANTIACID	,7540924	,4698825	1,2842028	,1057030

F level or tolerance or VIN insufficient for further computation.

Summary Table

Action	Vars	Wilks'		
Step Entered	Removed	in	Lambda	Sig. Label
1	RESPUEST	1	,42281	,0000 Tiempo de respuesta al tratamiento
2	ALCOHOL	2	,17606	,0000 Consumo de Alcohol (gramos diarios)
3	TABACO	3	,10728	,0000 Paciente ha dejado de fumar

Canonical Discriminant Functions

Fcn	Eigenvalue	Pct of Variance	Cum Pct	Canonical Corr	After Fcn	Wilks' Lambda	Chi-square	df	Sig
					0	,107275	583,761	9	,0000
1*	6,3915	96,26	96,26	,9299	1	,792926	60,675	4	,0000
2*	,1782	2,68	98,94	,3889	2	,934260	17,782	1	,0000
3*	,0704	1,06	100,00	,2564					

\* Marks the 3 canonical discriminant functions remaining in the analysis.

*Con las 3 funciones que quedan la más discriminante quitando la 1ª y 2ª ... 5*



# Análisis Discriminante

## Standardized canonical discriminant function coefficients

	Func 1	Func 2	Func 3
RESPUEST	1,26224	,17854	,46500
ALCOHOL	1,19442	,31794	-,64228
TABACO	-,64938	,86963	,24077

Case Number	Mis Val	Actual Sel Group	Highest Group	Probability P(D/G)	Probability P(G/D)	2nd Highest Group	Probability P(G/D)	Discrim Scores
1	YES	3	3	,6991	,9737	4	,0262	-3,1389 ,5715 ,1466
27	NO	2 **	3	,2848	,5658	2	,4341	-,4793 1,0285 ,5394
33	YES	2 **	3	,2488	,6326	2	,3674	-1,0162 -1,4317 ,1707
117	NO	UNGRPD	2	,7240	,9963	3	,0022	1,2512 -,8282 -1,0485
312	YES	2	2	,0340	,9949	3	,0033	1,3450 -,5523 -2,9219



# Análisis Discriminante

Standardized canonical discriminant function coefficients:

	Func 1	Func 2	Func 3
RESPUEST	,128485	,19242	,48137
ALCOHOL	,122802	,33896	-,61951
TABACO	-,67233	,86105	,22733

Classification results -

Actual Group	No. of Cases	Predicted Group Membership			
		1	2	3	4
Group 1 4 Meses	36	34 94,4%	2 5,6%	0 ,0%	0 ,0%
Group 2 8 Meses	153	4 2,6%	143 93,5%	6 3,9%	0 ,0%
Group 3 12 Meses	97	0 ,0%	5 5,2%	90 92,8%	2 2,1%
Group 4 16 Meses	22	0 ,0%	0 ,0%	3 13,6%	19 86,4%
Ungrouped cases	4	1 25,0%	2 50,0%	1 25,0%	0 ,0%

Percent of "grouped" cases correctly classified: 92,86%



# Análisis Discriminante

Classification results for cases selected for use in the analysis -

Actual Group	No. of Cases	Predicted Group Membership			
		1	2	3	4
Group 1 4 Meses	34	32 94,1%	2 5,9%	0 ,0%	0 ,0%
Group 2 8 Meses	126	4 3,2%	118 93,7%	4 3,2%	0 ,0%
Group 3 12 Meses	86	0 ,0%	3 3,5%	82 95,3%	1 1,2%
Group 4 16 Meses	20	0 ,0%	0 ,0%	3 15,0%	17 85,0%

Percent of "grouped" cases correctly classified: 93,61%

Classification results for cases not selected for use in the analysis -

Actual Group	No. of Cases	Predicted Group Membership			
		1	2	3	4
Group 1 4 Meses	2	2 100,0%	0 ,0%	0 ,0%	0 ,0%
Group 2 8 Meses	27	0 ,0%	25 92,6%	2 7,4%	0 ,0%
Group 3 12 Meses	11	0 ,0%	2 18,2%	9 81,8%	0 ,0%
Group 4 16 Meses	2	0 ,0%	0 ,0%	1 50,0%	1 50,0%

Percent of "grouped" cases correctly classified: 88,10%

