**Trabajo Práctico N° 7**

**Ejercicio 1.**

*La base de datos “firmas.dta” contiene información financiera referida a dos tipos de firmas: 12 firmas consideradas de buena performance y 12 firmas no tan buenas en ese sentido. En la base, se podrán observar dos características relevadas en la totalidad de las firmas: EBITASS (ganancias después de impuestos e intereses sobre activos) y ROTC (retorno sobre el capital).*

**(a)** *Realizar un gráfico de las variables relevadas. Interpretar.*



**(b)** *Realizar un test de medias multivariado.*

Estadístico= 34,31241761346325.

P-value= 3,54122225462e-08.

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los grupos no tienen medias iguales.

**(c)** *Realizar un análisis discriminante.*

Linear discriminant analysis

Resubstitution classification summary

+---------+

| Key |

|---------|

| Number |

| Percent |

+---------+

| Classified

True group | 0 1 | Total

-------------+----------------+-------

0 | 12 0 | 12

| 100.00 0.00 | 100.00

| |

1 | 1 11 | 12

| 8.33 91.67 | 100.00

-------------+----------------+-------

Total | 13 11 | 24

| 54.17 45.83 | 100.00

| |

Priors | 0.5000 0.5000 |

Por lo tanto, todas las observaciones clasificadas en el grupo 0 están correctamente clasificadas, mientras que una de las observaciones clasificada en el grupo 1 está incorrectamente clasificada.



+-----------------------------------------+

| | Classification | Probabilities |

| | | |

| Obs | True Class. | 0 1 |

|-----+------------------+----------------|

| 1 | 0 0 | 0.9962 0.0038 |

| 2 | 0 0 | 0.9999 0.0001 |

| 3 | 0 0 | 0.9998 0.0002 |

| 4 | 0 0 | 1.0000 0.0000 |

| 5 | 0 0 | 0.9997 0.0003 |

|-----+------------------+----------------|

| 6 | 0 0 | 0.9999 0.0001 |

| 7 | 0 0 | 0.9950 0.0050 |

| 8 | 0 0 | 1.0000 0.0000 |

| 9 | 0 0 | 0.5373 0.4627 |

| 10 | 0 0 | 0.9788 0.0212 |

|-----+------------------+----------------|

| 11 | 0 0 | 0.9993 0.0007 |

| 12 | 0 0 | 0.9994 0.0006 |

| 13 | 1 1 | 0.0001 0.9999 |

| 14 | 1 1 | 0.0112 0.9888 |

| 15 | 1 1 | 0.0086 0.9914 |

|-----+------------------+----------------|

| 16 | 1 1 | 0.0000 1.0000 |

| 17 | 1 1 | 0.0000 1.0000 |

| 18 | 1 1 | 0.0004 0.9996 |

| 19 | 1 1 | 0.0013 0.9987 |

| 20 | 1 0 \* | 0.5727 0.4273 |

|-----+------------------+----------------|

| 21 | 1 1 | 0.0000 1.0000 |

| 22 | 1 1 | 0.0239 0.9761 |

| 23 | 1 1 | 0.0001 0.9999 |

| 24 | 1 1 | 0.0019 0.9981 |

+-----------------------------------------+

\* indicates misclassified observations

Classification functions

| group

| 0 1

-------------+----------------------

ebitass | 61.23745 2.55117

rotc | 21.02689 -1.404444

\_cons | -7.7876 -.0033742

-------------+----------------------

Priors | .5 .5

**(d)** *Obtener predicciones para toda la muestra y comparar con la información observada.*

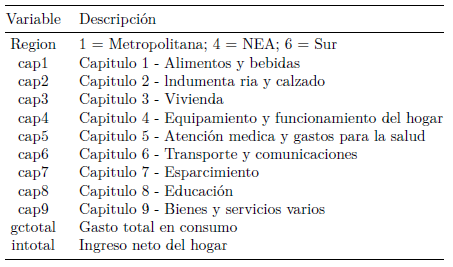
**(e)** *Considerar una firma con un EBITASS y un ROTC de 0,1. ¿En qué grupo se la clasificaría?*



Por lo tanto, considerando una firma con un EBITASS y un ROTC de 0,1, se la clasificaría en el grupo 0 (“Good performing companies”).

**Ejercicio 2.**

*La base de datos “muestra\_engh.dta” contiene información proveniente de la Encuesta Nacional de Gastos de los Hogares para tres regiones: Metropolitana, Noreste (NEA) y Sur. Considerar responder a las siguientes consignas con las variables desagregadas por capítulo de gasto.*



**(a)** *Realizar un test de igualdad de medias entre regiones.*

Estadístico= 72,80066506848998.

P-value= 1,50900286966e-08.

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que los grupos no tienen medias iguales.

**(b)** *Realizar un análisis discriminante.*

**(c)** *Obtener las predicciones a partir de la/s estimción/es de la función/es discriminate/s.*

**(d)** *Obtener las predicciones a partir de las distancias de Mahalanobis.*

----------------------------------

| predict2

| 1 2 3 Total

---------+------------------------

predict1 |

1 | 86 86

2 | 233 233

3 | 102 102

Total | 86 233 102 421

----------------------------------

**(e)** *Completar una tabla de predicción-realización.*

table1[3,3]

predict\_1 predict\_2 predict\_3

id\_1 44 60 16

id\_2 20 96 19

id\_3 22 77 67

Correct= 207.

Incorrect= 214.

Por lo tanto, 207 y 214 observaciones fueron clasificadas correcta e incorrectamente, respectivamente.

Linear discriminant analysis

Resubstitution classification summary

+---------+

| Key |

|---------|

| Number |

| Percent |

+---------+

| Classified

True id | 1 2 3 | Total

-------------+------------------------+-------

1 | 44 60 16 | 120

| 36.67 50.00 13.33 | 100.00

| |

2 | 20 96 19 | 135

| 14.81 71.11 14.07 | 100.00

| |

3 | 23 75 68 | 166

| 13.86 45.18 40.96 | 100.00

-------------+------------------------+-------

Total | 87 231 103 | 421

| 20.67 54.87 24.47 | 100.00

| |

Priors | 0.3333 0.3333 0.3333 |

**(f)** *¿Cómo se realizaría el análisis en caso de no poder suponer que las matrices de varianzas y covarianzas son iguales entre grupos? ¿Cuántos parámetros se deberían estimar?*

table2[3,3]

predict\_1 predict\_2 predict\_3

id\_1 36 72 12

id\_2 12 113 10

id\_3 22 100 44

Correct= 193.

Incorrect= 228.

Por lo tanto, 193 y 228 observaciones fueron clasificadas correcta e incorrectamente, respectivamente.

Quadratic discriminant analysis

Resubstitution classification summary

+---------+

| Key |

|---------|

| Number |

| Percent |

+---------+

| Classified

True id | 1 2 3 | Total

-------------+------------------------+-------

1 | 36 72 12 | 120

| 30.00 60.00 10.00 | 100.00

| |

2 | 12 113 10 | 135

| 8.89 83.70 7.41 | 100.00

| |

3 | 22 100 44 | 166

| 13.25 60.24 26.51 | 100.00

-------------+------------------------+-------

Total | 70 285 66 | 421

| 16.63 67.70 15.68 | 100.00

| |

Priors | 0.3333 0.3333 0.3333 |