

Análisis Discriminante

Autor: Dr. Cs. Gustavo Ruiz Aranibar¹

Resumen

La presente investigación presenta los resultados preliminares y finales de la clasificación, aplicada a la minería, teniéndose muestras analizadas, cuyos contenidos corresponden a 13 metales, teniéndose tres grupos de acuerdo al lugar de origen de cada muestra, y reagrupados hasta obtener los grupos con las muestras correspondientes.

Palabras clave: análisis discriminante, función discriminante, multivariante, distancia de Mahalanobis, variables, matriz de varianza-covarianza, centroide.

1. Introducción

El Análisis Multivariante (AM) es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado. Su razón de ser, radica en un mejor entendimiento del fenómeno objeto de estudio, obteniendo información que con los métodos estadísticos univariantes y bivariantes no se pueden conseguir. El AM, estudia, analiza, representa e interpreta los datos que resulten de observar un número $p > 1$ de variables estadísticas sobre una muestra de n individuos.

2. Conocimientos previos de matemáticas y estadística

1. Matriz de datos
$$X = x_{ij} = \begin{bmatrix} x_{11} & \dots & \dots & x_{1n} \\ x_{21} & \dots & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{p1} & \dots & \dots & x_{pn} \end{bmatrix}$$

2. Suma de columnas
$$X = x_j = \sum_{i=1}^p x_{ij}$$

3. Matriz de suma de cuadrados y productos
$$Z = z_{i,j} = \sum_{i=1}^p x_{i,l} * x_{l,j}$$

4. Vector de promedios
$$\bar{X} = \bar{x}_j = \left(\sum_{i=1}^p x_{ij} / n \right) = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_n \end{bmatrix}$$

5. Matriz de suma de cuadrados y suma de productos de desviaciones ²

$$A = a_{ii} = SCE_i = \sum_{\alpha=1}^n (x_{i\alpha} - \bar{x}_i)^2 = \sum_{\alpha=1}^n x_{i\alpha}^2 - \frac{1}{n} \left(\sum_{\alpha=1}^n x_{i\alpha} \right)^2$$

¹ Se agradece a la UAGRM por la beca otorgada con fondos del IDH, para cursar y culminar exitosamente el Doctorado en Ciencias en Educación Superior. Especializado en Estadística. Profesor de Estadística, Matemáticas y Computación

² En inglés: sums of squares and products of desviates. En francés: sommes des carrés et des produits des écarts

$$A = a_{ij} = SPE_{ij} = \sum_{\alpha=1}^n (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = \sum_{\alpha=1}^n x_{i\alpha}x_{j\alpha} - \frac{1}{n} \left(\sum_{\alpha=1}^n x_{i\alpha} \right) \left(\sum_{\alpha=1}^n x_{j\alpha} \right)$$

6. Covarianza $\text{cov}(x,y) = \frac{1}{n} \left(\sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i) \right)$

7. Matriz de varianzas y covarianzas² $S = s_{ij} = \begin{bmatrix} s_{11} & \dots & \dots & s_{1m} \\ s_{21} & \dots & \dots & s_{2m} \\ \dots & \dots & \dots & \dots \\ s_{n1} & \dots & \dots & s_{nm} \end{bmatrix} = \begin{bmatrix} a_{11}/n & \dots & \dots & a_{1m}/n \\ a_{21}/n & \dots & \dots & a_{2m}/n \\ \dots & \dots & \dots & \dots \\ a_{n1}/n & \dots & \dots & a_{nm}/n \end{bmatrix}$

8. Vector de desviaciones estándar $D = d_i = [\sqrt{a_{ii}}]$

9. Matriz de coeficientes de correlación $R = r_{ij} = [s_{ij} / (d_i * d_j)] = \begin{bmatrix} 1 & r_{12} & \dots & \dots & s_{1p} \\ r_{21} & 1 & \dots & \dots & r_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & \dots & 1 \end{bmatrix}$

10. Matriz inversa. Existen diferentes métodos para encontrar la inversa de una matriz, estando entre ellas los métodos de: Gauss-Jordán, Monte Carlo, partición de matrices, o hallando el cociente de la matriz adjunta por su determinante: $S^{-1} = \frac{Adj S}{|S|}$

11. Probabilidad total condicional $P(A) = \sum_{i=1}^n P(H_i)P(A|H_i)$

12. Fórmula de Bayes $P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}$

3. Matriz de correlaciones

Cuando existen altas correlaciones entre las p variables discriminantes, el investigador debe cuidarse de interpretaciones erróneas de los coeficientes de las funciones discriminantes, porque las variables relacionadas están compartiendo el peso en la función. En investigaciones se recomienda eliminar variables altamente correlacionadas, ya que la presencia de éstas puede generar ciertas limitaciones en el análisis.

4. Descripción del análisis discriminante (AD)

El AD, fue propuesto por R. Fisher, cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos, en el caso de que existan, explicar en qué sentido se dan y proporcionar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados, es una técnica de clasificación donde el objetivo es obtener una función capaz de clasificar a un nuevo individuo a partir del conocimiento de los valores de ciertas variables discriminadoras.

El AD, permite describir, seleccionar las variables que más influyen en el problema, construir una función a partir de estas variables y predecir en qué grupo se clasifica un nuevo individuo, el cual ha sido evaluado

2 En inglés: variance-covariance matrix, dispersión matrix. En francés: matrice des variances et covariances ou matrice de dispersion



en dicha función. Se ve a este procedimiento como un modelo de predicción de una variable respuesta categórica (variable grupo) a partir de p variables explicativas generalmente continuas (variables clasificatorias).

Los pasos a seguir para llevar a cabo un AD, comprenden:

- Plantear el problema a resolver por el AD.
- Analizar si existen diferencias significativas entre los grupos.
- Establecer el número y composición de las dimensiones de discriminación entre los grupos analizados.
- Determinar qué variables clasificatorias explican la mayor parte de las diferencias observadas.
- Construir procedimientos sistemáticos de clasificación de objetos de procedencia desconocida en los grupos analizados.
- Evaluar la significación estadística y práctica de los resultados obtenidos en el proceso de clasificación.

Como cualquier otra técnica estadística, la aplicación del AD ha de ir precedida de una comprobación de los supuestos asumidos por el modelo, el AD se apoya en los siguientes supuestos: a) Normalidad multivariante b) Igualdad de matrices de varianza-covarianza c) Linealidad d) Ausencia de multicolinealidad y e) Singularidad.

5. Análisis discriminante lineal (ADL)

En el AD, el punto de partida es un conjunto de objetos clasificados en dos o más grupos, de estos objetos, se conocen sus variables atributo. Al reconocer de antemano la existencia de estos grupos, parece lógico pensar que existen variables cuyo valor numérico determina la pertenencia a uno u otro grupo. Los objetivos del AD son: a) La identificación de variables atributo que mejor discriminen entre los grupos y la evaluación del poder discriminante de cada una de ellas. b) Asignar, con un cierto grado de riesgo, un objeto del que no se conoce su clasificación y del que se conocen las variables atributo.

Como técnica de análisis de dependencia, el ADL permite obtener un modelo lineal de causalidad en el cual la variable dependiente puede ser métrica o categórica, y las variables independientes son métricas, continuas y determinan a qué grupo pertenecen los objetos. Se trata de encontrar relaciones lineales entre las variables que mejor discriminen a los grupos iniciales de objetos. Además, se trata de definir una regla de decisión que asigne un nuevo objeto a uno de los grupos prefijados.

Entre las ventajas del ADL se tiene:

- La técnica ADL es fácil de aplicar especialmente si se tiene el programa informático.
- Las probabilidades de pertenencia a un grupo dado son determinadas por el programa.
- Está disponible en muchos programas estadísticos.

Entre las desventajas del ADL se mencionan:

- Las suposiciones de normalidad e igualdad de varianzas no siempre se cumplen en las variables del modelo.
- La clasificación de nuevas observaciones no es muy eficiente a medida que se incrementa el número de variables del modelo.
- Seleccionar las variables antes de aplicar el ADL.
- Requiere que se especifiquen los grupos del conjunto de entrenamiento del modelo con clases prefijadas.

6. Análisis discriminante clásico

El AD de Fisher tiene por principio definir, para dos poblaciones y p variables, una función lineal:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

$$y = \beta_0 + \beta x$$

Permite afectar a una de las dos poblaciones, todo individuo caracterizado por un valor negativo de y , y a la otra población todo individuo por un valor positivo de y , de todas maneras el riesgo de clasificación errónea³, (clasificación de un individuo perteneciente a una población, dentro de la otra población) sea lo más pequeña posible.

De una manera general, la investigación de la función discriminante (FD) se realiza suponiendo que las dos poblaciones poseen distribuciones normales a p dimensiones, de iguales varianzas y covarianzas, y que los parámetros de esas poblaciones son conocidas o pueden ser estimadas a partir de muestras suficientemente grandes. En estas condiciones se puede demostrar que la función asegura el riesgo de clasificación erróneo mínimo, sea:

$$y = b_0 + bx$$

$$y = [x - (\bar{x}_1 + \bar{x}_2)/2]' \hat{\Sigma}^{-1} (\bar{x}_1 + \bar{x}_2)$$

donde:

$$y = (n_1 + n_2 - 2)[x - (\bar{x}_1 + \bar{x}_2)/2]' (A_1 + A_2)^{-1} (\bar{x}_1 + \bar{x}_2)$$

Los valores promedio de la variable y , relativos a las dos muestras son en consecuencia:

$$\bar{y}_1 = (\bar{x}_1 + \bar{x}_2)' \hat{\Sigma}^{-1} (\bar{x}_1 + \bar{x}_2)/2 \quad y \quad \bar{y}_2 = -\bar{y}_1$$

La diferencia existente entre los dos promedios:

$$\bar{y}_1 - \bar{y}_2 = (\bar{x}_1 + \bar{x}_2)' \hat{\Sigma}^{-1} (\bar{x}_1 + \bar{x}_2)$$

No es otro en valor absoluto, que el cuadrado de la distancia generalizada existente, en el sentido de Mahalanobis, entre las dos muestra:

$$|\bar{y}_1 - \bar{y}_2| = D^2$$

Por otra parte se demuestra, siempre en las mismas condiciones, que los valores de y poseen para cada población una distribución aproximadamente normal, donde la varianza puede ser estimada por: $|\bar{y}_1 - \bar{y}_2|$ o D^2 .

Esta probabilidad es en efecto aquella de observar un valor negativo de y para un individuo de la población de promedio positivo $(D^2)/2$ o un valor positivo de y para un individuo de la población de promedio negativo $(-D^2)/2$. En consecuencia la probabilidad de clasificación errónea puede ser calculada como sigue con la ayuda de tablas de la distribución normal reducida, donde ϕ designa la función de repartición de esta distribución.

$$P = \left(u > \frac{D^2/2}{\sqrt{D^2}} \right) = 1 - \phi \left(\frac{D^2}{2} \right) = 1 - \phi \left(\frac{1}{2} \sqrt{|\bar{y}_1 - \bar{y}_2|} \right)$$

La función y , que es así definida posee una cierto número de propiedades importantes. Ella es no solamente aquella que minimiza el riesgo de clasificación errónea, pero también, en relación con la prueba T^2 de Hotelling aquella que vuelve máximo la relación de la varianza de y entre las poblaciones de la varianza de y dentro las poblaciones. Dentro del espacio a p dimensiones correspondiente a las p variables observadas, la ecuación: $b_0 + bx = 0$

Es aquella de un hiperplano, donde todos los puntos son tales que las funciones de densidad de probabilidad de dos poblaciones son iguales: $f_1(x) = f_2(x)$.

Este hiperplano divide el espacio a p dimensiones en dos espacios, el uno engloba todos los puntos para los cuales: $f_1(x) > f_2(x)$, y el otro todos los puntos para los cuales: $f_1(x) < f_2(x)$.

Entre el AD y la RM, existe una relación simple, entre la distancia generalizada de Mahalanobis y el coeficiente de correlación múltiple (CCM), teniéndose:

$$D^2 = \frac{(n_1 + n_2)(n_1 + n_2 - 2)R^2}{n_1 n_2 (1 - R^2)}$$

Cuando los efectos de las dos muestras son iguales:

$$D^2 = \frac{4(n - 1)R^2}{n(1 - R^2)}$$

Estas relaciones permiten calcular las probabilidades de clasificación errónea a partir de los CCM.

7. Enfoque de Fisher del análisis discriminante

Encuentra una buena FD que sea una combinación lineal de las variables originales. Geométricamente: Se busca una buena dirección sobre la que se proyectará los datos de los grupos conocidos y de los que se desea clasificar. Se clasifica en función de qué grupo está más cerca en esa dirección.

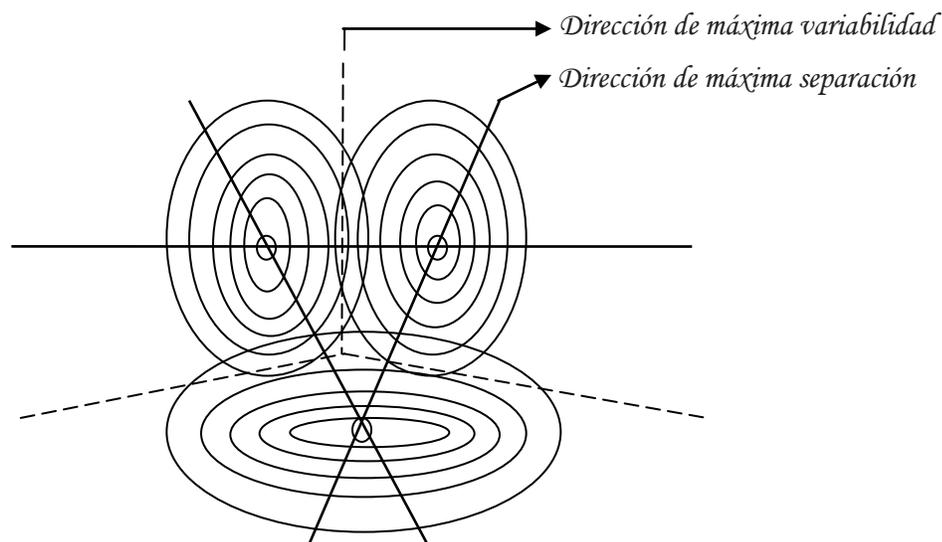


Fig. 1. Representación de las direcciones máximas considerando tres grupos

En la figura, se observa las direcciones de separación y variabilidad de las proyecciones de tres grupos de muestras, las diferentes comparaciones se realizan considerando solo dos a dos los grupos. Una buena dirección tiene que separar bien las medias y tener en cuenta la variabilidad.

8. Distancia de Mahalanobis

La distancia de Mahalanobis utiliza la inversa de la matriz de covarianza (MC), de determinada muestra, de pertenecer a un determinado colectivo, cabe señalar que la distancia de Mahalanobis es adimensional. La distancia D^2 de Mahalanobis es una medida de distancia generalizada y se basa en la distancia euclídea generalizada al cuadrado que se adecúa a varianzas desiguales; la regla de selección en este procedimiento es maximizar la distancia D^2 de Mahalanobis.

Se define la distancia de Mahalanobis como:
$$D_i^2 = (Y_i - \bar{Y})^t S^{-1} (Y_i - \bar{Y}).$$

La distancia de Mahalanobis se distribuye según una $\chi_{p,\alpha}^2$ siendo p el número de variables. (α = dado por el investigador, n = número de variables, $p = n-2$).

Si $\chi_{p,\alpha}^2$ es $<$ al Chi cuadrado teórico no existe observación atípica.

La distancia de Mahalanobis se usa también para calcular la probabilidad de pertenencia a cada grupo usando la regla de Bayes (para eso es necesario proporcionar una distribución a priori). En el resultado que arroja el programa informático⁴ del AD, se tiene la probabilidad de pertenencia al grupo asignado para cada observación.

9. Regla de Fisher con más de dos grupos

La regla de Fisher se extiende al caso en el que se desea clasificar un nuevo dato y exista tres o más grupos. Si se considera los datos utilizados por Fisher⁵ del año 1936, que consistió en medir los pétalos de 3 especies de flores, cada una de ellas con 4 variables, teniéndose una matriz 150 x 4, las tres especies son: iris setosa, iris virginica e iris versicolor.

Utilizando el programa computacional, y observando los resultados, el clasificador ha cometido tan sólo 3 errores: dos datos de la clase versicolor han sido asignados a virginica, y un dato de virginica a versicolor. En la clase setosa resulta que se clasifican todas, lo que significa que pertenecen a la especie iris setosa, no así las muestras 21 y 34 de la especie iris virginica corresponden a la especie iris versicolor y la muestra 34 de la especie iris virginica corresponde a la especie iris versicolor.

Aplicación. La técnica del AD tiene numerosas aplicaciones, utilizándose para abordar problemas complejos en diferentes disciplinas. Una aplicación del ADL a la minería, sería considerar los datos de muestras analizadas por su composición química correspondiendo a tres grupos: área productiva, no productiva y de prospección. Cada muestra compuesta de 13 variables, que corresponde a su composición química, siendo esta información fuente para cada muestra de los tres grupos:

4 Desarrollado por el autor de la presente investigación.

5 Multivariate Analysis, Maurice Kendal, página 40.

Análisis Discriminante

| Grp. 1 | Ti ^a | Mn ^b | Ag ^c | Ba ^a | Co ^d | Cr ^a | Cu ^c | Ni ^a | Pb ^a | Sr ^a | V ^a | Zn ^a | Au ^c |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|
| 1 | 7280 | 1300 | 30,0 | 720 | 30 | 150 | 73 | 50 | 70 | 60 | 70 | 190 | 0,02 |
| 2 | 10300 | 1200 | 0,7 | 1280 | 20 | 160 | 25 | 50 | 70 | 90 | 50 | 50 | 0,02 |
| 3 | 6500 | 700 | 1,0 | 1070 | 20 | 200 | 48 | 70 | 100 | 210 | 50 | 170 | 0,01 |
| 4 | 7000 | 1500 | 0,7 | 760 | 30 | 160 | 70 | 40 | 110 | 240 | 40 | 250 | 0,01 |
| 5 | 5100 | 1000 | 0,5 | 740 | 20 | 140 | 39 | 50 | 80 | 50 | 60 | 130 | 0,02 |
| 6 | 10600 | 2100 | 0,3 | 980 | 30 | 50 | 25 | 30 | 70 | 150 | 160 | 110 | 0,01 |
| 7 | 14200 | 2000 | 0,2 | 690 | 30 | 70 | 25 | 50 | 60 | 160 | 70 | 180 | 0,01 |
| 8 | 9700 | 900 | 0,2 | 680 | 35 | 70 | 38 | 30 | 70 | 90 | 110 | 250 | 0,01 |
| 9 | 2300 | 1500 | 0,2 | 710 | 5 | 110 | 50 | 20 | 70 | 80 | 30 | 12 | 0,01 |
| 10 | 12100 | 6300 | 0,1 | 1520 | 30 | 30 | 24 | 30 | 80 | 320 | 160 | 190 | 0,02 |
| 11 | 3000 | 1100 | 0,2 | 510 | 5 | 30 | 15 | 30 | 30 | 240 | 30 | 50 | 0,02 |
| 12 | 7500 | 2400 | 0,7 | 690 | 30 | 30 | 31 | 10 | 100 | 210 | 40 | 280 | 0,03 |
| 13 | 7800 | 1800 | 4,0 | 730 | 55 | 40 | 24 | 30 | 20 | 90 | 320 | 90 | 0,01 |
| 14 | 6900 | 1500 | 1,0 | 326 | 30 | 50 | 25 | 10 | 90 | 70 | 200 | 70 | 0,04 |
| 15 | 11200 | 3100 | 1,5 | 660 | 50 | 40 | 20 | 40 | 50 | 140 | 280 | 90 | 0,01 |
| 16 | 5200 | 1400 | 0,8 | 680 | 35 | 50 | 42 | 20 | 50 | 30 | 150 | 150 | 0,01 |
| 17 | 5100 | 1500 | 0,9 | 700 | 25 | 60 | 67 | 40 | 80 | 40 | 190 | 90 | 0,01 |
| 18 | 10500 | 2900 | 0,4 | 1640 | 25 | 20 | 21 | 30 | 30 | 320 | 90 | 200 | 0,01 |
| 19 | 11500 | 3200 | 0,7 | 710 | 30 | 30 | 15 | 20 | 20 | 260 | 270 | 180 | 0,01 |
| 20 | 7100 | 1800 | 0,9 | 490 | 75 | 50 | 8 | 10 | 30 | 80 | 180 | 100 | 0,02 |

| Grp. 2 | Ti ^a | Mn ^b | Ag ^c | Ba ^a | Co ^d | Cr ^a | Cu ^c | Ni ^a | Pb ^a | Sr ^a | V ^a | Zn ^a | Au ^c |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|
| 1 | 4820 | 500 | 0,1 | 160 | 20 | 70 | 30 | 10 | 0 | 720 | 140 | 200 | 0,01 |
| 2 | 3040 | 500 | 0,2 | 150 | 20 | 30 | 82 | 10 | 20 | 1580 | 160 | 70 | 0,01 |
| 3 | 890 | 600 | 0,1 | 50 | 10 | 10 | 61 | 10 | 0 | 340 | 40 | 50 | 0,02 |
| 4 | 2100 | 500 | 0,1 | 100 | 15 | 30 | 77 | 10 | 0 | 650 | 90 | 80 | 0,02 |
| 5 | 5060 | 700 | 0,3 | 140 | 20 | 50 | 154 | 20 | 0 | 1240 | 140 | 80 | 0,01 |
| 6 | 1980 | 700 | 0,1 | 80 | 15 | 20 | 63 | 20 | 0 | 720 | 80 | 110 | 0,00 |
| 7 | 3220 | 600 | 0,2 | 160 | 20 | 30 | 45 | 20 | 10 | 1100 | 120 | 60 | 0,01 |
| 8 | 3280 | 800 | 0,2 | 90 | 15 | 10 | 40 | 30 | 20 | 1480 | 70 | 40 | 0,00 |
| 9 | 2020 | 700 | 0,1 | 80 | 15 | 20 | 104 | 20 | 0 | 420 | 80 | 70 | 0,00 |
| 10 | 4600 | 700 | 0,3 | 160 | 20 | 60 | 48 | 10 | 20 | 780 | 150 | 50 | 0,02 |
| 11 | 3100 | 500 | 0,2 | 100 | 15 | 30 | 65 | 10 | 20 | 710 | 100 | 40 | 0,01 |
| 12 | 3020 | 600 | 0,2 | 90 | 15 | 10 | 69 | 0 | 30 | 1310 | 110 | 30 | 0,02 |
| 13 | 1860 | 500 | 0,1 | 70 | 10 | 20 | 63 | 0 | 10 | 480 | 80 | 50 | 0,00 |
| 14 | 2800 | 700 | 0,1 | 110 | 15 | 20 | 58 | 10 | 20 | 730 | 120 | 80 | 0,01 |
| 15 | 1040 | 1600 | 0,1 | 20 | 5 | 10 | 37 | 0 | 10 | 140 | 30 | 80 | 0,01 |
| 16 | 4640 | 800 | 0,3 | 220 | 15 | 20 | 121 | 20 | 20 | 1200 | 210 | 160 | 0,00 |
| 17 | 4990 | 900 | 0,3 | 190 | 20 | 40 | 59 | 20 | 30 | 480 | 230 | 120 | 0,02 |
| 18 | 2830 | 800 | 0,2 | 120 | 15 | 20 | 40 | 10 | 20 | 690 | 140 | 60 | 0,00 |
| 19 | 4500 | 700 | 0,2 | 140 | 20 | 30 | 82 | 20 | 10 | 710 | 170 | 70 | 0,00 |
| 20 | 2900 | 600 | 0,1 | 80 | 15 | 10 | 99 | 0 | 0 | 760 | 80 | 90 | 0,01 |

| Grp. 3 | Ti ^a | Mn ^b | Ag ^c | Ba ^a | Co ^d | Cr ^a | Cu ^c | Ni ^a | Pb ^a | Sr ^a | V ^a | Zn ^a | Au ^c |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|
| 1 | 4260 | 800 | 0,3 | 180 | 20 | 60 | 128 | 30 | 30 | 460 | 110 | 80 | 0,02 |
| 2 | 6500 | 1200 | 0,5 | 380 | 30 | 40 | 72 | 50 | 20 | 320 | 90 | 160 | 0,01 |
| 3 | 12200 | 5200 | 1,5 | 630 | 25 | 80 | 39 | 40 | 90 | 210 | 200 | 180 | 0,01 |
| 4 | 1080 | 1600 | 0,2 | 80 | 5 | 10 | 102 | 0 | 10 | 160 | 30 | 80 | 0,00 |
| 5 | 3820 | 500 | 0,2 | 170 | 25 | 40 | 60 | 20 | 10 | 1100 | 160 | 40 | 0,02 |
| 6 | 1020 | 2400 | 0,1 | 20 | 0 | 10 | 28 | 0 | 0 | 1320 | 20 | 60 | 0,00 |

^a Determinado lo más cercano a 10 ppm. ^b Determinado lo más cercano a 100 ppm.

^c Determinado a la fracción de ppm. ^d Determinado lo más cercano a 5 ppm.

Resultados. Procesando esta información se observó que las muestras 9 y 15 del segundo grupo corresponden al tercer grupo, la muestra 5 del tercer grupo corresponde al segundo grupo; obteniéndose la clasificación final con los siguientes resultados:

Cuadrado generalizado de Mahalanobis = 279,88

En esta aplicación se tienen tres grupos, con 13 variables, por consiguiente se tendrán tres funciones discriminantes lineales, que son las siguientes:

| | Función discriminante 1 | Función discriminante 2 | Función discriminante 3 |
|-------------|-------------------------|-------------------------|-------------------------|
| Constante | -16,71816 | -15,47526 | -16,03613 |
| Variable 1 | -0,00053 | -0,00020 | -0,00051 |
| Variable 2 | 0,00000 | 0,00092 | 0,00571 |
| Variable 3 | -0,05420 | -0,25450 | -0,40001 |
| Variable 4 | 0,01443 | -0,00121 | -0,00972 |
| Variable 5 | 0,24835 | 0,04631 | 0,12189 |
| Variable 6 | 0,01905 | 0,05187 | 0,05311 |
| Variable 7 | 0,03823 | 0,09533 | 0,13617 |
| Variable 8 | 0,03740 | 0,03363 | 0,20224 |
| Variable 9 | 0,06527 | -0,05170 | -0,08632 |
| Variable 10 | 0,00537 | 0,01480 | 0,00704 |
| Variable 11 | 0,01844 | 0,03457 | -0,00345 |
| Variable 12 | 0,02378 | 0,03199 | 0,02631 |
| Variable 13 | 237,12320 | 235,17480 | 259,0354 |

Evaluación de funciones de clasificación para cada observación

| Grupo 1 Nº obs. | Probabilidad asociada con la más grande función discriminante | Función más grande |
|--------------------|--|-----------------------|
| 1 | 1,00000 | 1 |
| 2 | 1,00000 | 1 |
| 3 | 1,00000 | 1 |
| 4 | 0,99998 | 1 |
| 5 | 0,99992 | 1 |
| 6 | 1,00000 | 1 |
| 7 | 0,99550 | 1 |
| 8 | 1,00000 | 1 |
| 9 | 0,99889 | 1 |
| 10 | 0,99999 | 1 |
| 11 | 0,80447 | 1 |
| 12 | 1,00000 | 1 |
| 13 | 0,99999 | 1 |
| 14 | 0,99995 | 1 |
| 15 | 0,99922 | 1 |
| 16 | 0,99999 | 1 |
| 17 | 0,99999 | 1 |
| 18 | 1,00000 | 1 |
| 19 | 0,75983 | 1 |
| 20 | 1,00000 | 1 |

| Grupo 2 Nº obs. | Probabilidad asociada con la más grande función discriminante | Función más grande |
|--------------------|--|-----------------------|
| 1 | 0,99975 | 2 |
| 2 | 1,00000 | 2 |
| 3 | 0,89424 | 2 |
| 4 | 0,98131 | 2 |
| 5 | 0,66340 | 2 |
| 6 | 0,99896 | 2 |
| 7 | 0,99694 | 2 |
| 8 | 0,99907 | 2 |
| 9 | 0,98893 | 2 |
| 10 | 0,99997 | 2 |
| 11 | 0,96769 | 2 |
| 12 | 0,99295 | 2 |
| 13 | 0,99993 | 2 |
| 14 | 0,99616 | 2 |
| 15 | 0,99687 | 2 |
| 16 | 0,98649 | 2 |
| 17 | 0,96505 | 2 |

| Grupo 3 N° obs. | Probabilidad asociada con la más grande función discriminante | Función más grande |
|--------------------|--|-----------------------|
| 1 | 0,96951 | 3 |
| 2 | 0,99741 | 3 |
| 3 | 0,99959 | 3 |
| 4 | 0,99940 | 3 |
| 5 | 0,55793 | 3 |

10. Conclusiones

Las muestras estarán bien clasificadas, si todas pertenecen al mismo grupo, siendo este el fin perseguido por la técnica del AD, pero puede ocurrir que en cada grupo existan muestras que pertenezcan a otros grupos, si esto acontece se debe sacar estas muestras de los grupos correspondientes, lo cual lo realiza el programa informático. Los resultados muestran la probabilidad de pertenecer al grupo correspondiente.

Colaboración

Dr. Manuel Febrero Bande. Dr. en Estadística e Investigación Operativa (1985-1990)
Universidad de Santiago de Compostela - Santiago de Compostela - España

Prof. M. Nilda Avilés de Ruiz. Lic. en Idiomas. Universidad Autónoma Gabriel René Moreno. Santa Cruz
– Bolivia (Agosto, 2005)

Bibliografía

1. Dagnelie Pierre, analyse statistique á plusieurs variables. Les presses agronomiques de Gembloux, Bélgica, 1975 (2da. Edición), pp. 362 – Xiv.
2. Kendall s. Maurice, multivariate analysis. Charles Griffin & co. Ltd., Londres, Inglaterra, 1975, pp. 210 – Xi.
3. Davis c. John, statistics and data analisis in geology. John Wiley & Sons, New york, Estados Unidos, 1973, pp. 550 – Vii.
4. Ruiz Aranibar Gustavo⁶. Libreria científica de programas informáticos, La Paz -Bolivia.

Pensamiento: Quien se decide a: enseñar, escribir o investigar, nunca debe dejar de: aprender, estudiar o producir intelectualmente, para divulgar y exponer este conocimiento

6 N° 8035 Calle 20 y Av. Ballivian, Calacoto, La Paz – Bolivia Tel. 591-22772162 Cel, 67111778
gustavoruiz432@hotmail.com.bo ruizaranibargustavo@gmail.com.bo Blog: Gustavo Ruiz Aranibar