

Análisis Discriminante como un ACP particular

Dr. Oldemar Rodríguez Rojas

1 de octubre de 2022

Índice general

1. Análisis Discriminante como un ACP particular	2
1.1. Los datos	2
1.2. Un ejemplo del problema a tratar	3
1.3. Varianza intra-clases y varianza inter-clases	4
1.4. El poder de discriminación de una variable	6
1.5. Representaciones gráficas	8

Capítulo 1

Análisis Discriminante como un ACP particular

1.1. Los datos

Los datos constan de dos juegos de variables:

1. El primer juego de variables indica los valores tomados por n individuos para m_1 variables numéricas. Estos individuos forman las n filas de una matriz X_1 , de las cuales las m_1 columnas son las variables. Suponemos además que cada una de las variables está centrada.
2. El segundo juego de datos describe la pertenencia de cada uno de los n individuos a una de las m_2 modalidades de una variable cualitativa; este segundo juego de datos consta de la matriz disyuntiva completa X_2 con n filas y m_2 columnas.

Ejemplo 1. Para la tabla de datos cuyas primeras 7 filas se muestran en la Figura 1.1 se tiene la matriz X_1 que se muestra en la Figura 1.2 y la matriz X_2 que se muestra en la Figura 1.3. Ver el ejemplo completo en el archivo [HTML](#).

	RT1	RT2	RT3	RT4	RT5	VC
A1	9.0	4.6	2.0	0.1	25.8	A
B1	4.0	3.3	0.4	0.6	32.4	B
C1	1.4	1.0	1.1	0.5	23.5	C
A2	10.0	6.7	3.9	0.2	15.7	A
B2	8.2	6.2	2.1	0.2	17.0	B
C2	2.5	5.4	0.8	0.4	16.3	C
A3	10.0	6.3	4.0	0.6	25.8	A

Figura 1.1: Tabla de Datos.

	RT1	RT2	RT3	RT4	RT5
A1	9.0	4.6	2.0	0.1	25.8
B1	4.0	3.3	0.4	0.6	32.4
C1	1.4	1.0	1.1	0.5	23.5
A2	10.0	6.7	3.9	0.2	15.7
B2	8.2	6.2	2.1	0.2	17.0
C2	2.5	5.4	0.8	0.4	16.3

Figura 1.2: Matriz X_1 .

	VC.A	VC.B	VC.C
A1	1	0	0
B1	0	1	0
C1	0	0	1
A2	1	0	0
B2	0	1	0
C2	0	0	1

Figura 1.3: Matriz X_2 .

Además se definen, n_g como el número de individuos en el grupo g ; D_G como la matriz diagonal cuyo elemento de orden g es el número de individuos del grupo g , es decir, n_g . Denotamos por G la matriz de dimensión $m_2 \times m_1$ cuya fila g está formada por las coordenadas del centro de grupo g ; así $G = (D_G)^{-1} (X_2)^t X_1$.

Ejemplo 2. Para la tabla de datos cuyas primeras 7 filas se muestran en la Figura 1.1 se tiene la siguiente matriz de centros de gravedad G que se muestra en la Figura 1.4.

	RT1	RT2	RT3	RT4	RT5
A	10.90	6.59	3.29	0.36	21.34
B	6.70	5.84	1.74	0.96	22.22
C	4.04	4.81	0.82	0.81	21.66

Figura 1.4: Matriz de Centros de Gravedad.

1.2. Un ejemplo del problema a tratar

A partir de los valores tomados por un individuo para las m_1 variables numéricas se trata de determinar a cuál de las m_2 clases de la variable categórica pertenece.

Suponga que $m_1 = 2$, es decir, tenemos solo 2 variables numéricas y que $m_2 = 2$, es decir, la variable categórica tiene solo 2 clases.

Estos datos están representados en la Figura 1.5, estando constituidos los ejes por las variables numéricas; los individuos del primer grupo están representados por la letra A y los del segundo grupo por la letra B .

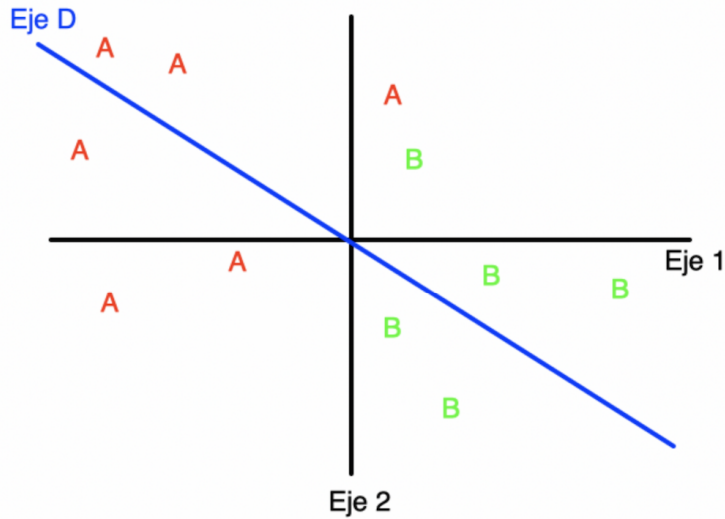


Figura 1.5: Ejemplo de un eje discriminante.

El **Eje Discriminante** discrimina inequívocamente entre los dos grupos. Si proyectamos ortogonalmente a los individuos en este eje, los individuos que tienen una coordenada positiva en el eje pertenecen al grupo B y aquellos que tienen una coordenada negativa pertenecen al grupo A .

Por tanto, aquí es posible determinar la pertenencia de un individuo dado a una clase a partir del valor tomado por este individuo para una combinación lineal de variables iniciales.

El problema del Análisis Discriminante es precisamente la construcción de una o más combinaciones de variables originales que permitan discriminar los grupos; estas combinaciones lineales se denominan *Variables Discriminantes*.

1.3. Varianza intra-clases y varianza inter-clases

Considere el caso general nuevamente. Sea z^1 la primera variable discriminante:

$$z^1 = X_1 u^1,$$

donde:

- u^1 es un vector columna con m_1 filas, siendo el primer factor de discriminación.

- z^1 se define para cada uno de los n individuos y z_i^1 denota el valor que toma z^1 para el individuo i .
- \bar{z}_g^1 es el promedio de los valores de z_i^1 para el grupo g .

Es posible definir 3 tipos de varianzas:

- La **Varianza Total** V_T^1 es la varianza de los n individuos:

$$V_T^1 = \frac{1}{n} \sum_{i=1}^n (z_i^1)^2.$$

- La varianza de los individuos que pertenecen a una misma clase g es la varianza intra-clase de un grupo g y se denota por $V_{I_g}^1$:

$$V_{I_g}^1 = \frac{1}{n_g} \sum_{i \in g} (z_i^1 - \bar{z}_g^1)^2.$$

Así, la **Varianza Intra-Clases** V_I^1 es la suma de las varianzas intra-clases de los m_2 clases, siendo el peso de cada una de las clases la frecuencia relativa de la clase, es decir, $\frac{n_g}{n}$, entonces:

$$V_I^1 = \sum_{g=1}^{m_2} \frac{n_g}{n} \frac{1}{n_g} \sum_{i \in g} (z_i^1 - \bar{z}_g^1)^2 = \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} (z_i^1 - \bar{z}_g^1)^2 = \frac{1}{n} \sum_{g=1}^{m_2} \sum_{i \in g} (z_i^1 - \bar{z}_g^1)^2.$$

- La **Varianza Inter-Clases** V_E^1 es la varianza de los valores promedio de las clases, donde cada uno de estos medios se ve afectado por la frecuencia relativa de la clase correspondiente:

$$V_E^1 = \sum_{g=1}^{m_2} \frac{n_g}{n} (\bar{z}_g^1)^2.$$

Teorema 1. [Teorema de Fisher]

Varianza Total = Varianza Intra-Clases + Varianza Inter-Clases, es decir:

$$V_T^1 = V_I^1 + V_E^1.$$

Prueba:

$$\begin{aligned}
V_I^1 &= \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} (z_i^1 - \bar{z}_g^1)^2 = \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} \left((z_i^1)^2 - 2(z_i^1 \bar{z}_g^1) + (\bar{z}_g^1)^2 \right) \\
&= \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} (z_i^1)^2 - 2 \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} z_i^1 \bar{z}_g^1 + \sum_{g=1}^{m_2} \frac{1}{n} \sum_{i \in g} (\bar{z}_g^1)^2 \\
&= \sum_{i=1}^n \frac{1}{n} (z_i^1)^2 - 2 \sum_{g=1}^{m_2} \frac{n_g}{n} \left(\sum_{i \in g} \frac{1}{n_g} z_i^1 \right) \bar{z}_g^1 + \frac{1}{n} \sum_{g=1}^{m_2} n_g (\bar{z}_g^1)^2 \\
&= V_T^1 - 2 \sum_{g=1}^n \frac{n_g}{n} (\bar{z}_g^1)^2 + V_E^1 \\
&= V_T^1 - 2V_E^1 + V_E^1 = V_T^1 - V_E^1.
\end{aligned}$$

◇

1.4. El poder de discriminación de una variable

Para un valor dado de la varianza total, una variable discrimina los grupos mucho mejor si para esta variable se tiene que:

- Los individuos que pertenecen al mismo grupo tienen valores lo más cercanos posible, es decir, la varianza intra-clase es lo más pequeña posible.
- Las medias de los diferentes grupos son lo más diferentes posible, en otras palabras, que la varianza inter-clases es lo más grande posible.

Como bien sabemos ambas cosas suceden al mismo tiempo gracias al Teorema de Fisher, es decir, como sumadas las varianza intra-clase y la varianza inter-clases dan una constante (la varianza total) entonces cuando una se maximiza la otra se minimiza.

Definición 1. El **Poder de Discriminación** de una variable se define como el cociente entre la varianza inter-clases y la varianza total.

Así, el objetivo del Análisis Discriminante, en un primer paso, es determinar la variable sintética que tiene un poder de discriminación máximo.

Teorema 2. La variable discriminante z^1 que tiene el mayor Poder de Discriminación se escribe como $z^1 = X_1 u^1$; donde el factor discriminante u^1 es el vector propio de $((X_1)^t X_1)^{-1} G^t D_G G$ asociado con su valor propio máximo β_1 . El Poder de Discriminación z^1 es igual a β_1 .

Para poder realizar la prueba ocupamos del siguiente Lema.

Lema 1. Sean A y B dos matrices cuadradas $m \times m$ simétricas y sea A una matriz definida positiva. Entonces el vector $y \in \mathbb{R}^n$ que resuelve el siguiente problema de optimización:

$$\begin{cases} \max y^t B y \\ \text{sujeto a } y^t A y = 1 \end{cases}$$

es el vector propio a^1 de $A^{-1}B$ de norma 1 asociado al valor propio más grande β_1 .

Prueba del Lema: Se omite (se ve en el curso de optimización).

Observación: El problema:

$$\begin{cases} \max y^t B y \\ \text{sujeto a } y^t A y = 1 \end{cases}$$

es equivalente a:

$$\begin{cases} \max \frac{y^t B y}{y^t A y} \\ y \neq 0 \end{cases}$$

Prueba del Teorema:

- Se debe maximizar el cociente $\frac{V_E^1}{V_T^1}$.
- Las coordenadas de los centros de los grupos vienen dadas por el vector $G u^1$, y por lo tanto la varianza inter-clases es $\frac{1}{n} (u^1)^t G^t D_G G u^1$.
- La varianza total está dada por: $\frac{1}{n} (u^1)^t (X_1)^t X_1 u^1$.
- Entonces el cociente que se debe maximizar es:

$$\frac{(u^1)^t G^t D_G G u^1}{(u^1)^t (X_1)^t X_1 u^1},$$

por lo que el resultado se obtiene usando el lema anterior.

En la tarea completar los detalles y explicar por qué el Poder de Discriminación z^1 es igual a β_1 .

◇

En general se tiene el siguiente teorema:

Teorema 3. En la etapa k , el Análisis Discriminate se calcula una variable discriminante $z^k = X_1 u^k$ que tenga el máximo poder de discriminación bajo la restricción: $(u^k)^t (X_1)^t X_1 u^r = 0$ para $r = 1, \dots, k-1$. El factor discriminante u^k es el vector propio de $(X_1^t X_1)^{-1} G^t D_G G$ asociado al k -ésimo valor propio más grande β_k . El poder discriminante de z^k es igual a β_k .

Prueba del Teorema: Análogo al teorema anterior.

Teorema 4. El Análisis Discriminante es un ACP de la matriz G cuando el espacio de los individuos está provisto con la métrica de *Mahalanobis* $\left(\frac{1}{n} (X_1)^t X_1\right)^{-1}$.

Prueba: Parte de la tarea, debe repasar del curso anterior el ACP con otras métricas.

1.5. Representaciones gráficas

El plano principal y el círculo de correlaciones se representan de la misma forma que en el ACP, como se muestra en las Figuras 1.6 y 1.7:

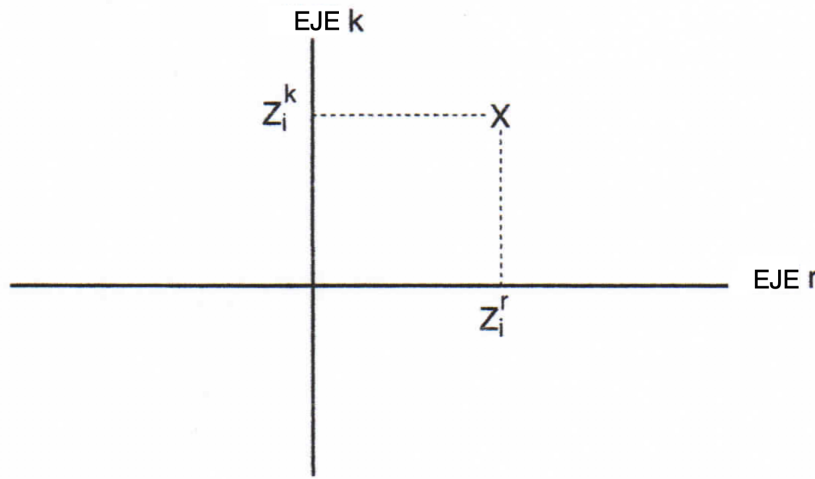


Figura 1.6: Representación de los individuos.

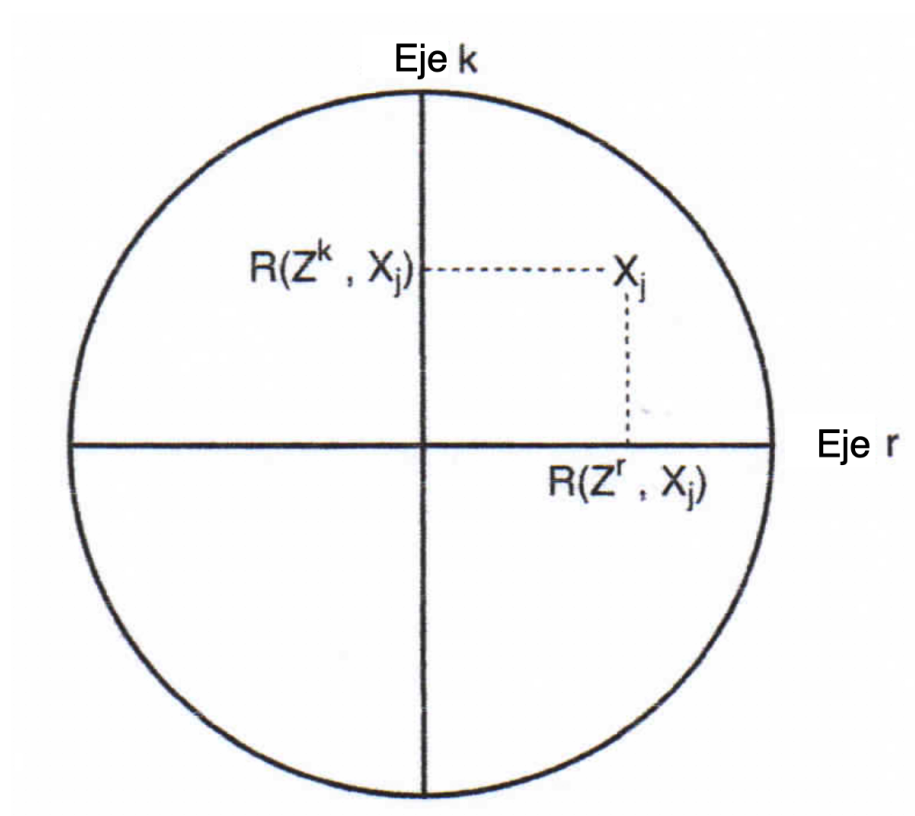


Figura 1.7: Representación de las variables.