GENERALIZACIÓN DE LA DISTANCIA DE MAHALANOBIS PARA EL ANALISIS DISCRIMINANTE LINEAL EN POBLACIONES CON MATRICES DE COVARIANZA DESIGUALES

Ricardo A. Leiva
Universidad Nacional de Cuyo
Argentina

Myriam Herrera
Universidad Nacional de San Juan
Argentina

RESUMEN

En este trabajo se define una distancia entre poblaciones multivariadas con diferentes matrices de varianza - covarianza que luego es utilizada para fundamentar la metodología de discriminación lineal entre dichas poblaciones.

PALABRAS CLAVES: Distancia de Mahalanobis, discriminación lineal y poblaciones multivariadas.

I - Introducción

El análisis Discriminante es un método del Análisis Estadístico Multivariado que permite separar observaciones vectoriales provenientes de diferentes poblaciones. Fisher (1936), bajo el supuesto de que eran dos poblaciones con la misma matriz de varianza - covarianza, convirtió a este problema multivariado en un problema univariado mediante la utilización de una combinación lineal del vector aleatorio involucrado. La determinación de esta combinación lineal la realizó exigiendo que fuera la que maximizaba la separación entre las poblaciones transformadas por ella. Asociada con este criterio existe una medida de la separación entre las poblaciones multivariadas originales que se mide usando la distancia de Mahalanobis (Mahalanobis 1936). Una propiedad de esta distancia de Mahalanobis es que cuando se la utiliza para establecer una medida de la separación entre dos poblaciones multivariadas, esta medida es "invariante" ante transformaciones lineales del vector aleatorio p variado X.

Cuando la suposición de igualdad entre las matrices de varianza - covarianza de las poblaciones involucradas se elimina, el problema de la discriminación entre ellas se convierte en un problema no lineal.

Sin embargo se puede encontrar una solución lineal aproximada para este problema. Precisamente, el objetivo de este trabajo es determinar una distancia (generalización de la de Mahalanobis) que sirva para medir la separación entre poblaciones multivariadas con diferentes matrices de varianza - covarianza y
que permita resolver el problema de la discriminación entre estas poblaciones vía la maximización de la separación entre las poblaciones transformadas.

El esquema de este trabajo será presentar en la sección 2 la solución de Fisher desde el punto de vista de la maximización de la separación de las poblaciones transformadas y en la sección 3 se resolverá el problema para el caso de dos poblaciones con diferentes matrices de varianza-covarianza. En la sección 4 se aplica la metodología propuesta a datos multivariados obtenidos de la bibliografía para los cuales se rechaza la hipótesis de homogeneidad de matrices de covarianza y en la sección 5 se realiza una simulación para comparar la solución propuesta en este trabajo con el problema mencionado con la de Fisher. En esta sección también se analizan distintas medidas de la distancia entre dos poblaciones multivariadas desde el punto de vista de su capacidad para predecir la bondad de la discriminación lineal entre dichas poblaciones. Finalmente, se presenta un apéndice con las tablas que muestran los resultados de la simulación realizada.

2 Análisis Discriminante Lineal de dos poblaciones con la misma matriz de varianza-covarianza

Supongamos que hay dos poblaciones \( \Pi_1 \) y \( \Pi_2 \) de objetos y que en cada uno de estos objetos se miden \( p \) características que conforman el vector aleatorio \( p \)-variado \( X = (X_1, \ldots, X_p)' \). Este vector \( X \) tiene las densidades \( f_{X}^{(1)} \) y \( f_{X}^{(2)} \) en las poblaciones \( \Pi_1 \) y \( \Pi_2 \) respectivamente. Indicaremos con \( \mu_1 \) y \( \mu_2 \) sus respectivos vectores medios y por \( \Sigma \) su matriz de varianza-covarianza común a la que supondremos definida positiva. Consideraremos también la distancia \( d \) de Mahalanobis entre los vectores \( p \)-dimensionales \( a \) y \( b \) pertenecientes a estas poblaciones que está definida por la siguiente ecuación

\[
d^2(a; b) = (a - b)\Sigma^{-1}(a - b).
\]  

(1)

Es claro que la definición de \( d \) dada por (1) satisface las propiedades definitorias de una distancia, a saber

1. \( d(a; b) \geq 0 \).

2. \( d(a; b) = 0 \iff a = b \).

3. \( d(a; b) = d(b; a) \).

4. \( d(a; b) + d(b; c) \geq d(a; c) \).

Con esta distancia se puede definir el concepto de separación entre las poblaciones \( \Pi_1 \) y \( \Pi_2 \) para el vector de características \( X \), a la que indicamos por \( sep(f_{X}^{(1)}; f_{X}^{(2)}) \), como el cuadrado de la distancia de Mahalanobis (1) entre sus medias \( \mu_1 \) y \( \mu_2 \), es decir por la ecuación

\[
sep(f_{X}^{(1)}; f_{X}^{(2)}) = d^2(\mu_1; \mu_2) = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2).
\]  

(2)

Entre las propiedades que tiene este concepto de separación es conveniente destacar las siguientes dos, cuyas demostraciones pueden consultarse en el Apéndice (Propiedad 1 y Propiedad 2):

1. La separación entre \( f_{X}^{(1)} \) y \( f_{X}^{(2)} \) es invariante ante transformaciones lineales de la forma \( Z = A'X + b \) donde \( A \) es una matriz \( p \times p \) no singular (\( \text{ran}(A) = p \)):
2. Para transformaciones lineales de la forma \( Z = A'X + b \) donde \( A \) es una matriz \( p \times m \) con \( \text{rang}(A) = m \) se verifica que

\[
\text{sep}(f^{(1)}_X : f^{(2)}_X) = \max_{A \in M(p,m)} \text{sep}(f^{(1)}_Z : f^{(2)}_Z),
\]

donde \( M(p,m) \) es el conjunto de todas las matrices de orden \( p \times m \) de rango \( m \).

Con esta definición de separación, la solución al problema de la discriminación entre las dos poblaciones obtenida por Fisher se puede presentar como la búsqueda de la combinación lineal \( Z = a'X + b \) para la cual \( \text{sep}(f^{(1)}_Z : f^{(2)}_Z) \) sea máxima. Ahora bien, como las medias y la varianza común de esta combinación lineal para las dos poblaciones consideradas son respectivamente

\[
\eta_1 = a'\mu_1 + b \\
\eta_2 = a'\mu_2 + b \\
\sigma^2 = a'\Sigma a,
\]

entonces el problema a resolver es

\[
\max_{a \in M(p,1)} \text{sep}(f^{(1)}_Z : f^{(2)}_Z) = \max_{a \in \mathbb{R}^p} (\eta_1 - \eta_2)'(\sigma^2)^{-1}(\eta_1 - \eta_2)
\]

\[
= \max_{a \in \mathbb{R}^p} \frac{|a'(\mu_1 - \mu_2)|^2}{a'\Sigma a}
\]

\[
= \max_{a \in \mathbb{R}^p} \frac{a'(\mu_1 - \mu_2)(\mu_1 - \mu_2)'a}{a'\Sigma a}
\]

(4)

Por lo tanto el problema queda resuelto por el siguiente teorema cuya demostración se puede encontrar en libros de análisis multivariado (ver Mardia et al. [1982]):

**Teorema 2.1** Sea \( B \) una matriz \( p \times p \) definida positiva y sea \( d \) un vector \( p \) variado fijo. Entonces para todo vector \( p \) variado \( x \neq 0 \) se verifica que

\[
\max_{x \in \mathbb{R}^p} \frac{(x'y)^2}{x'Bx} = d'B^{-1}d,
\]

(5)

donde el máximo es alcanzado en el punto \( x = kB^{-1}d \) para cualquier constante \( k \neq 0 \).

En consecuencia, usando el resultado dado por el Teorema 2.1, reemplazando en (5) \( x \) por \( a \), \( d \) por \( (\mu_1 - \mu_2) \) y \( B \) por \( \Sigma \) y teniendo en cuenta (2), se obtiene en (4) que

\[
\max_{a \in M(1,0)} \text{sep}(f^{(1)}_Z : f^{(2)}_Z) = \max_{a \in \mathbb{R}^p} \frac{|a'(\mu_1 - \mu_2)|^2}{a'\Sigma a}
\]

\[
= \frac{a'(\mu_1 - \mu_2)(\mu_1 - \mu_2)'a}{a'\Sigma a},
\]

y que este máximo es alcanzado cuando el vector \( a \) está dado por

\[
a = k \Sigma^{-1}(\mu_1 - \mu_2),
\]

es decir, la combinación lineal que maximiza la separación entre las poblaciones transformadas es

\[
Z = q_1(X) = (\mu_1 - \mu_2)\Sigma^{-1}X,
\]

(6)
donde se ha elegido $k = 1$ y $b = 0$.

La transformación $q_1$ dada por (6) se denomina función discriminante y es la que se emplea para clasificar un objeto, cuyos valores de las característicos medidas son $x_0$, como perteneciente a una de las dos poblaciones $\Pi_1$ o $\Pi_2$. Esta clasificación se realiza comparando el valor que toma la función discriminante en $x_0$ con el valor medio de las medias de las dos poblaciones transformadas. Concretamente, (3) da la expresión de la medias y varianza común de las poblaciones transformadas por la combinación lineal dada en (6), resulta

$$
\begin{align*}
q_1 &= (\mu_1 - \mu_2)^\Sigma^{-1} \mu_1 \\
q_2 &= (\mu_1 - \mu_2)^\Sigma^{-1} \mu_2 \\
\sigma^2 &= (\mu_1 - \mu_2)^\Sigma^{-1} (\mu_1 - \mu_2),
\end{align*}
$$

el punto medio $c_1$ entre estas dos medias es

$$
c_1 = \frac{q_1 + q_2}{2} = \frac{1}{2} (\mu_1 - \mu_2)^\Sigma^{-1} (\mu_1 + \mu_2). 
$$

Entonces, el criterio de discriminación para el punto $x_0$ es el siguiente

- $x_0 \in \Pi_1$ si $(\mu_1 - \mu_2)^\Sigma^{-1} x_0 \geq c_1$
- $x_0 \in \Pi_2$ si $(\mu_1 - \mu_2)^\Sigma^{-1} x_0 < c_1$.

Una propiedad importante del punto de corte $c_1$ es que bajo el supuesto de normalidad la probabilidad teórica de clasificar correctamente un objeto es la misma (máxima posible) para ambas poblaciones, es decir

$$
Pr[Z \ \text{claseficada en} \ \Pi_1 \ | \ X \in \Pi_1] = Pr[Z \ \text{claseficada en} \ \Pi_2 \ | \ X \in \Pi_2].
$$

Consideraremos ahora al punto $c_p$ dado por

$$
c_p = \frac{1}{2} (\mu_1 + \mu_2),
$$

es decir, $c_p$ es la versión $p$ dimensional del punto de corte $c_1$ (que es el transformado de $c_p$ por (6)). Este punto $c_p$ está caracterizado por ser aquel en el que se obtiene el mínimo de la suma de los cuadrados de las distancias de un punto a cada una de las medias poblacionales. Es decir $c_p$ es el punto que verifica que

$$
\min_{x \in \Pi} d^2(x; \mu_1) + d^2(x; \mu_2) = d^2(x; \mu_1) + d^2(x; \mu_2).
$$

Reemplazando en (10) la expresión de $c_p$ dada en (9), se tiene además que

$$
\min_{x \in \Pi} d^2(x; \mu_1) + d^2(x; \mu_2) = \frac{1}{2} d^2(\mu_1; \mu_2).
$$

La demostración de este resultado se puede consultar en la Propiedad 3 del Apéndice.

3 Análisis Discriminante Lineal de dos poblaciones con matrices de varianza-covarianza diferentes

Consideramos ahora el caso en que las dos poblaciones $\Pi_1$ y $\Pi_2$ de objetos son tales que el vector al azar $p$ variado $X = (X_1, \ldots, X_p)'$ tiene respectivamente densidades $f_X^{(1)}$ y $f_X^{(2)}$, medias $\mu_1$ y $\mu_2$ y matrices de
varianza-covarianzas $\Sigma_i$ y $\Sigma_j$. Existen entonces dos distancias de Mahalanobis posibles, a saber $d_1$ y $d_2$ definidas por

$$d_1^2(a; b) = (a - b)'\Sigma_i^{-1}(a - b)$$
$$d_2^2(a; b) = (a - b)'\Sigma_j^{-1}(a - b).$$

Utilizando estas dos distancias de Mahalanobis y siguiendo la idea dada en (11) que caracterizaba al punto de corte $c_{ij}$, definiremos las funciones $d_{1,2}$ y $d_{2,1}$ mediante la ecuación válida para $i \neq j \in \{1, 2\}$

$$\frac{1}{2} d_{ij}^2(a; b) = \min_{x \in \mathbb{R}^p} d_i^2(a; x) + d_j^2(b; x) = \min_{x \in \mathbb{R}^p} h_{ij}(x). \quad (12)$$

El punto $c_{ij}$ que minimiza el segundo miembro de (12) se obtiene resolviendo la ecuación que resulta de igualar a cero la derivada de $h_{ij}$, o sea

$$\frac{dh_{ij}}{dx} = 2\Sigma_i^{-1}(a - x) + \Sigma_j^{-1}(b - x) = 0,$$

de donde, como de la invertibilidad de las matrices individuales de varianza-covarianza se deduce la invertibilidad de su suma (o lo que es lo mismo, invertibilidad de la suma de sus inversas), se tiene

$$c_{ij} = (\Sigma_i^{-1} + \Sigma_j^{-1})^{-1}(\Sigma_i^{-1}a + \Sigma_j^{-1}b). \quad (13)$$

Otra forma de expresar $c_{ij}$ es a partir de las siguientes ecuaciones

$$\Sigma_i^{-1} + \Sigma_j^{-1} = \Sigma_i^{-1}(\Sigma_i + \Sigma_j)^{-1} + (\Sigma_i^{-1})\Sigma_j^{-1} = \Sigma_i^{-1}(\Sigma_i + \Sigma_j)\Sigma_j^{-1}$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j$$

de donde, reemplazando (14) y (15) en (13), resulta

$$c_{ij} = \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j^{-1}a + \Sigma_i(\Sigma_i + \Sigma_j)^{-1}\Sigma_j^{-1}b$$
$$= \Sigma_i(\Sigma_i + \Sigma_j)^{-1}a + \Sigma_i(\Sigma_i + \Sigma_j)^{-1}b. \quad (16)$$

Es decir, para $i \neq j \in \{1, 2\}$ se puede expresar $c_{ij} = B_1a + B_2b$ con $B_1 = B_2 = I$, donde $B_1 = \Sigma_i(\Sigma_i + \Sigma_j)^{-1}$ para $k = 1, 2$. Entonces, como $a - c_{ij} = B_1(a - b)$ y $b - c_{ij} = -B_2(a - b)$, reemplazando en (12) se tiene que para $i \neq j \in \{1, 2\}$ es

$$d_{ij}^2(a; b) = (a - b)'\left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1}(a - b). \quad (17)$$

La demostración de esta igualdad puede consultarse en la Propiedad 4 del Apéndice.

Es evidente a partir de (17) que se verifica que $d_{1,2} = d_{2,1}$. Esta igualdad conduce naturalmente a la definición de la función $d$ dada por

$$d^2(a; b) = (a - b)'\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(a - b). \quad (18)$$

Es fácil probar que
1. La función $d$ es una distancia pues satisface las cuatro propiedades mencionadas en la sección 2. Esta distancia definida por (18) fundamenta la sugerencia realizada por Park y Kshirsagar [1994] para modificar a la distancia entre poblaciones normales propuesta por Paranjpe y Gore [1994].

2. Los puntos $c_{12}$ y $c_{21}$, que no son en general iguales entre sí, verifican que

$$c_{12} + c_{21} = (B_2a + B_1b) + (B_1a + B_2b) = a + b.$$

En consecuencia, $\frac{B_1 + B_2}{2} = \frac{a + b}{2}$ y los cuatro puntos $a$, $b$, $c_{12}$, y $c_{21}$ son coplanares y vértices de un paralelogramo.

3. En el caso en que las matrices de varianza-covarianza son iguales, es decir $\Sigma_1 = \Sigma_2 = \Sigma$, la distancia definida por (18) coincide con la distancia de Mahalanobis dada en (1). Además en este caso los puntos de corte $c_{12}$ y $c_{21}$ son ambos iguales al punto $c_p$ definido por (9).

4. El punto de corte $c_{ij}$ depende no sólo del orden de los subíndices $i$ y $j$ sino también de los puntos $a$ y $b$. En particular se verifica que $a = b \iff c_{ij} = a \land c_{ij} = b \iff c_{ji} = a \lor c_{ji} = b$, y por ejemplo

$$a = c_{ij} = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}a + \Sigma_2(\Sigma_1 + \Sigma_2)^{-1}b$$

$$\iff \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}(b - a) = 0$$

$$\iff a = b.$$

5. De la propiedad anterior se deduce que si $a = b$ entonces $c_{ij} = c_{ji}$. Si $\Sigma_1 - \Sigma_2$ es no singular se verifica que $a \neq b \Rightarrow c_{ij} \neq c_{ji}$.

6. Cualesquiera sean $a$ y $b$ se verifica que $d^2(a; b) \leq 2 \min_{i \in \{1, 2\}} d_i^2(a; b)$. Esto se puede demostrar usando (18), (17) y (12), ya que

$$d^2(a; b) = d_{ij}^2(a; b) = 2 \min_{x \in \mathbb{R}^p} d_i^2(a; x) + d_j^2(b; x) \leq 2(d_i^2(a; x) + d_j^2(b; x)) \forall x \in \mathbb{R}^p.$$

En particular, eligiendo $x = b$ se obtiene $d^2(a; b) < 2d_j^2(a; b)$ y eligiendo $x = a$ se obtiene $d^2(a; b) < 2d_i^2(a; b)$.

7. Cuando $a = \mu_1$ y $b = \mu_2$, el punto $c_{12}$ adquiere una mayor relevancia que el punto $c_{21}$, ya que en la definición de $c_{12}$ (ver (12)) intervienen $d_i^2(\mu_1; x)$ y $d_j^2(\mu_2; x)$ que tienen más sentido que las $d_i^2(\mu_1; x)$ y $d_j^2(\mu_2; x)$ que intervienen en la definición de $c_{21}$. Debido a esta situación adoptaremos la notación $\gamma_2 = \gamma_{12} = \gamma_{21}$ que nos permitirá más adelante simplificar algunos enunciados. Más precisamente,

$$\gamma_2$$

es el punto donde se alcanza el

$$\min_{x \in \mathbb{R}^p} d_i^2(\mu_2; x) + d_j^2(\mu_2; x) \quad (19)$$

Usando esta distancia se puede definir el concepto de separación entre las poblaciones $\Pi_1$ y $\Pi_2$ mediante la ecuación

$$sep[f_{\Pi_1}^{(1)}; f_{\Pi_2}^{(2)}] = d^2(\mu_1; \mu_2) = (\mu_1 - \mu_2)' \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2). \quad (20)$$

69
Esta medida de la separación entre poblaciones multivariadas está vinculada a la medida de disimilaridad \(H\) de Hellinger (Rao y Varadarajan [1965]) definida para funciones densidad normales multivariadas \(f\) y \(g\), con, respectivamente, medias \(\mu_1\) y \(\mu_2\) y matrices de covarianza \(\Sigma_1\) y \(\Sigma_2\). Más precisamente, la medida \(H\) está dada por

\[
H = -\log \int f(x)g(x)dx = \frac{1}{4} \rho + \frac{1}{8} D^2, \tag{21}
\]

donde

\[
\rho = 2\log\left[\frac{1}{2}(\Sigma_1 + \Sigma_2)\right] - \log|\Sigma_1| - \log|\Sigma_2|,
\]

\[
D^2 = (\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mu_1 - \mu_2).
\]

Asimismo, la medida de separación dada en (20) está relacionada con la medida de la cercanía \(r\) propuesta por Matushita [1960] definida en general para \(f\) y \(g\) cualesquiera por

\[
r(f, g) = \int \left( f(x)\frac{1}{2} - g(x)\frac{1}{2} \right) dx.
\]

Esta noación surge de la distancia \(d^*\) dada por

\[
d^* (f, g) = \sqrt{\left( \int \left( f(x)\frac{1}{2} - g(x)\frac{1}{2} \right) \right)^2 dx},
\]

a través de la igualdad \(d^* (f, g) = 2[1 - r(f, g)]\) con \(0 \leq r(f, g) \leq 1\). En particular para funciones densidades \(f\) y \(g\), respectivamente, normales multivariadas \(N(\mu_1, \Sigma_1)\) y \(N(\mu_2, \Sigma_2)\), resulta dada por

\[
r(f, g) = \frac{1}{\sqrt{|\Sigma_1\Sigma_2|}} \exp - Q(\mu_1, \mu_2, \Sigma_1, \Sigma_2) \tag{22}
\]

donde el exponente \(Q = Q(\mu_1, \mu_2, \Sigma_1, \Sigma_2)\) dado por Matushita

\[
Q = \sum_{i=1}^{2} \sum_{j=1}^{2} (\mu_i - \mu_j)^T (\Sigma_i^{-1} + \Sigma_j^{-1})^{-1} \Sigma_i^{-1} \mu_i
\]

se reduce, usando (14), a

\[
Q = \frac{1}{2} (\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mu_1 - \mu_2) = \frac{1}{2} d^2 (\mu_1; \mu_2).
\]

Con la definición de separación dada en (20), para resolver el problema de la discriminación entre las dos poblaciones basta con encontrar la combinación lineal \(Z = aX + b\) para la cual la \(\text{sep}(f_x^{(1)}, f_x^{(2)})\) sea máxima. Resulta así que, (ver Propiedad 5 en el Apéndice), la función discriminante es la combinación lineal

\[
Z^* = (\mu_1 - \mu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} X.
\]

Notar entonces que

\[
\text{sep}(f_x^{(1)}, f_x^{(2)}) = (\gamma_1 - \gamma_2)^T \left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right)^{-1} (\gamma_1 - \gamma_2) = d^2 (\mu_1; \mu_2) = \text{sep}(f_x^{(1)}, f_x^{(2)}).
\]

70
Llamando $\Sigma_{1,2}$ a

$$\Sigma_{1,2} = \frac{\Sigma_1 + \Sigma_2}{2},$$

la función discriminante es la transformación lineal $T_{12}$ dada por

$$Z^* = T_{12}(X) = (\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}X.$$  \hfill (23)

Como lo hicimos con (7) en el caso de matrices de varianza-covarianza iguales, debemos ahora elegir razonablemente el punto de corte $c$ que servirá para establecer el criterio de discriminación entre las dos poblaciones. Existen varias posibles opciones para esta elección. Entre esas opciones se pueden mencionar las siguientes:

1. El punto $\eta_{12}$ que es el punto medio entre $\eta_1$ y $\eta_2$, o sea $\eta_{12} = \frac{\eta_1 + \eta_2}{2}$.

2. El punto $m_{12}$ que es la versión univariada del punto $\eta_{12}$.

3. Algún punto intermedio entre $\eta_{12}$ y $m_{12}$, o sea un punto de la forma $\lambda \eta_{12} + (1 - \lambda) m_{12}$ para $0 \leq \lambda \leq 1$.

4. El punto $m_{12}'$ dado por $m^* = \frac{\Sigma_1\mu_1 + \Sigma_2\mu_2}{\Sigma_1 + \Sigma_2}$.

Elijamos uno cualquiera de estos puntos para poder definir un hiperplano que sepa a las poblaciones multivariadas originales $\Pi_1$ y $\Pi_2$. Este hiperplano, que es perpendicular a $a' = (\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}$, está formado por todos los puntos $X$ que son transformados por $T_{12}$ en el punto elegido.

En particular, mencionemos algunas características de los puntos enumerados más arriba.

1. El punto $\eta_{12}$ es el transformado por $T_{12}$ del punto medio entre $\mu_1$ y $\mu_2$, ya que

$$T_{12}(\frac{\mu_1 + \mu_2}{2}) = (\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}(\frac{\mu_1 + \mu_2}{2}) = \frac{1}{2}(\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}\mu_1 + (\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}\mu_2 = \frac{\eta_1 + \eta_2}{2} = \eta_{12},$$

además se prueba que

$$\eta_{21} = T_{21}(\frac{\mu_1 + \mu_2}{2}) = -T_{12}(\frac{\mu_1 + \mu_2}{2}).$$  \hfill (24)

La ecuación del hiperplano mencionado más arriba es $(\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}(x - \frac{\eta_1 + \eta_2}{2}) = 0$. Al comparar este hiperplano con el que surge cuando se supone que $\Sigma_1 = \Sigma_2$ se observa que:

(a) Ambos hiperplano conciden cuando efectivamente $\Sigma_1 = \Sigma_2$.

(b) Ambos hiperplanos pasan por el punto $m_{12} = \frac{\eta_1 + \eta_2}{2}$, diferenciándose en la inclinación.

(c) Ambos tienen la tendencia de favorecer en la clasificación a la población con mayor concentración de datos (ver Sección 4).

2. Elegir para separar $f_{12}^{(1)}$ de $f_{12}^{(2)}$ al ponto $m_{12}$ por surgir este punto naturalmente en este trabajo, ya que es la versión univariada del punto $\gamma_{12}$ (igual al ci) que sirvió para definir la separación entre las poblaciones originales y que está dado por (10). Más precisamente, se tiene que

$$m_{12} = \frac{\sigma_1^2(a_1^2 + a_2^2)^{-1}b_1 + \sigma_2^2(a_1^2 + a_2^2)^{-1}b_2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^2(\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}\mu_1 + \sigma_2^2(\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}\mu_2}{2(\mu_1 - \mu_2)'\Sigma_{1,2}^{-1}(\mu_1 - \mu_2)}.$$

71
de donde restando y sumando $\mu_2$ al $\mu_1$ que aparece en el numerador del primer sumando del segundo miembro de la igualdad anterior, resulta:

$$m_{12} = (\mu_1 - \mu_2)\Sigma_{1,2}^{-1}y_2 + \frac{1}{2}(\mu_1 - \mu_2)^2\Sigma_{1,2}^{-1}\Sigma_{2,2}\Sigma_{1,2}^{-1}(\mu_1 - \mu_2).$$

O bien la siguiente expresión general que será usada más adelante:

$$m_{ij} = (\mu_i - \mu_j)\Sigma_{i,j}^{-1}y_j + \frac{1}{2}(\mu_i - \mu_j)^2\Sigma_{i,j}^{-1}\Sigma_{j,j}\Sigma_{i,j}^{-1}(\mu_i - \mu_j).$$

Es importante destacar que este punto $m_{12}$ es el mismo que se obtiene transformando el punto $\gamma_{12}$ mediante (23) y es fácil demostrar que el punto $m_{12}$ dado por (23) se reduce al dado en (7) cuando se verifica que $\Sigma_1 = \Sigma_2 = \Sigma$. Convien también notar que el transformado $m_{21}$ del punto $\gamma_{21}$ por la transformación $T_{21}(X) = (\mu_2 - \mu_1)\Sigma_{1,2}^{-1}(X)$ verifica que

$$m_{21} = T_{21}(\gamma_{21}) = -T_{12}(\gamma_{12}) = -m_{12}$$

Este punto de corte $m_{12}$, y su versión $p$-variada $\gamma_{13}$, adquirieron una significación especial, ya que a partir de ellos se puede definir un hiperplano que separa a las poblaciones multivariadas originales $\Pi_1$ y $\Pi_2$. Este hiperplano es el que siendo perpendicular a $d' = (\mu_1 - \mu_2)\Sigma_{1,2}^{-1}$ pasa por $\gamma_{12}$, o sea, su ecuación está dada por $d'(x - \gamma_{12}) = 0$ y, por lo tanto, es el hiperplano formado por todos los puntos $x$ que al aplicarles (23) se obtiene el mismo $m_{12} = \frac{d'x + d'y}{d'x + d'y}$ que al transformar $\gamma_{12}$. Si se compara este hiperplano con el que surge como separador de las poblaciones suponiendo $\Sigma_1 = \Sigma_2$ se puede observar que:

(a) Ambos hiperplanos coinciden cuando efectivamente $\Sigma_1 = \Sigma_2$.
(b) En las aplicaciones en las que $\Sigma_1 \neq \Sigma_2$, el uso de este hiperplano que proponemos como separador de las poblaciones multivariadas corrige, generalmente en demasía, la tendencia que tiene el otro hiperplano a favorecer en la clasificación a la población con mayor concentración de datos (ver Sección 4).

3. La elección de un punto intermedio de la forma $\lambda\gamma_{12} + (1 - \lambda) m_{12}$ para algún $\lambda$ tal que $0 \leq \lambda \leq 1$, queda justificado por lo mencionado más arriba respecto del comportamiento tendencioso de la discriminación cuando se elige alguno de los dos puntos anteriores.

4. El punto de corte $m_{12}$, bajo el supuesto de normalidad, satisface que la probabilidad teórica de clasificar correctamente un objeto en la misma (máxima posible) para ambas poblaciones como se estipulaba en (8). Esta deseable propiedad la cumple el punto $m_{12}$ dado por

$$m_{12} = \frac{\alpha_1\bar{y}_1 + \alpha_2\bar{y}_2}{\alpha_1 + \alpha_2}.$$ (27)

Para probar esto supongamos que se requiere que

$$Pr[X \text{ clasificado en } \Pi_1 | X \in \Pi_1] = Pr[X \text{ clasificado en } \Pi_2 | X \in \Pi_2],$$

esto es equivalente a pedir que los $Z$ que son el resultado de aplicar a los $X$ la transformación $T_{12}$, satisfagan la misma propiedad, es decir

$$Pr[Z = T_{12}(X) \text{ clasificado en } \Pi_1 | X \in \Pi_1] = Pr[Z = T_{12}(X) \text{ clasificado en } \Pi_2 | X \in \Pi_2]$$

72
Prueba de hipótesis para el análisis discriminante lineal en poblaciones con matrices de covarianzas desiguales

\[
Pr\left[ \frac{Z - \eta_1}{\sigma_1} \leq \frac{m_{12} - \eta_1}{\sigma_1} \mid X \in \Pi_1 \right] = Pr\left[ \frac{Z - \eta_2}{\sigma_2} \geq \frac{m_{12} - \eta_2}{\sigma_2} \mid X \in \Pi_2 \right]
\]

\[
\Phi\left( \frac{m_{12} - \eta_1}{\sigma_1} \right) = \Phi\left( \frac{m_{12} - \eta_2}{\sigma_2} \right)
\]

\[
\frac{m_{12} - \eta_1}{\sigma_1} = \frac{m_{12} - \eta_2}{\sigma_2}
\]

\[
\frac{m_{12}}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}} = \frac{\eta_1 + \eta_2}{\sigma_1 + \sigma_2}
\]

\[
m_{12} = \frac{\eta_1 + \eta_2}{\sigma_1 + \sigma_2}
\]

También en este caso se demuestra que el punto \( m_{12} \) de (27) se reduce al dado en (7) cuando se verifica que \( \Sigma_1 = \Sigma_2 = \Sigma \). Es fácil probar que el punto \( m_{12} \) se puede expresar como combinación lineal convexa de los puntos \( \eta_{12} \) y \( m_{12} \) dados en los incisos 1 y 2 anteriores. Más precisamente, si se elige

\[
\lambda = \frac{2\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2)^2},
\]

que claramente es \( 0 \leq \lambda \leq 1 \), se tiene que \( m_{12} = \lambda \eta_{12} + (1 - \lambda) m_{12} \).

En consecuencia, el hiperplano que "mejor" separa a las poblaciones multivariadas originales \( \Pi_1 \) y \( \Pi_2 \) es el que está perpendicular a \( a' = (\mu_1 - \mu_2)/\Sigma_{12}^{-1} \) pasa por \( \beta_{12} \), donde

\[
\beta_{12} = \lambda \mu_{12} + (1 - \lambda) \gamma_{12},
\]

con \( \lambda \) dado por (28). O sea, es el hiperplano de ecuación

\[
g_2(x) = (\mu_1 - \mu_2)/\Sigma_{12}^{-1}(x - \beta_{12}) = 0
\]

y, por lo tanto, está formado por todos los puntos \( x \) que al aplicárselos (23) se obtiene el mismo \( m_{12} \) = \( \Sigma_{12}^{-1}\mu_2 \) que al transformar \( \beta_{12} \). Si se compara este hiperplano con el que surge como separador de las poblaciones suponiendo \( \Sigma_1 = \Sigma_2 \) se puede observar que:

(a) Ambos hiperplanos coinciden cuando efectivamente \( \Sigma_1 = \Sigma_2 \).

(b) En las aplicaciones en las que \( \Sigma_1 \neq \Sigma_2 \), el uso de este hiperplano que proponemos como separador de las poblaciones multivariadas corregirá la tendencia que tiene el otro hiperplano a favorecer en la clasificación a la población con mayor concentración de datos (ver sección 4).

Este hiperplano que pasa por el punto \( \beta_{12} \) (y su versión univariada: el punto \( m_{12} \)) es el que se propone en este trabajo para separar las poblaciones multivariadas originales (respectivamente las poblaciones univariadas).

Como \( T_{12} \) y \( T_{21} \) son transformaciones lineales, usando (24) y (25), se tiene que

\[
m_{21}^* = T_{12}(\beta_{12}) = -T_{12}(\beta_{12}) = -m_{12}.
\]

Con esta resultando el criterio de discriminación para un punto \( x_0 \) es el siguiente
Clasifique \( x_0 \in \Pi_i \) si \( T_i(x_0) - m_{i2} > 0 \) \( \Leftrightarrow (\mu_i - \mu_j)^T \Sigma_{ii}^{-1} x_0 - m_{i2} > 0 \)
\( x_0 \in \Pi_j \) si \( T_j(x_0) - m_{j2} > 0 \) \( \Leftrightarrow (\mu_i - \mu_j)^T \Sigma_{jj}^{-1} x_0 - m_{j2} < 0 \).

Es decir, el espacio \( p \)-dimensional \( V \) de las observaciones se particiona en dos regiones disjuntas \( V_1 \) y \( V_2 \), donde \( V_i = \{ x : (\mu_i - \mu_j)^T \Sigma_{jj}^{-1} x_0 - m_{ij} > 0 \} \) para \( i \neq j : 1, 2 \), de tal forma que si una observación se encuentra en la región \( V_i \) es la clasifica como perteneciente a la población \( \Pi_i \).

En el caso de tener \( k \), con \( k \geq 2 \), poblaciones con matrices de covarianza desiguales, se particiona al espacio \( p \)-dimensional \( V \) de las observaciones en \( k \) regiones disjuntas \( V_1, \cdots, V_k \), donde, para \( i : 1, \cdots, k \), es
\[
V_i = \{ x : (\mu_i - \mu_j)^T \Sigma_{jj}^{-1} x_0 - m_{ij} > 0 \text{ con } 1 \leq j \leq k, j \neq i \}
\]
con
\[
\Sigma_{ij} = \frac{\Sigma_i + \Sigma_j}{2},
\]
y
\[
m_{ij} = T_i(\beta_{ij}),
\]
donde
\[
\beta_{ij} = \lambda \mu_{ij} + (1 - \lambda) \gamma_{ij}
\]
con
\[
\mu_{ij} = \frac{\mu_i + \mu_j}{2},
\]
y con \( \gamma_{ij} \) dado por (19). Es decir,
\[
m_{ij} = T_i(\beta_{ij}) = \lambda \mu_{ij} + (1 - \lambda) \left( (\mu_i - \mu_j)^T \Sigma_{jj}^{-1} \mu_j + (1 - \lambda) \left( (\mu_i - \mu_j)^T \Sigma_{ii}^{-1} \mu_i + \frac{1}{2} [(\mu_i - \mu_j)^T \Sigma_{ii}^{-1} \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} (\mu_i - \mu_j)] \right) \right),
\]
donde \( \lambda \) está dado en (38).

Cuando no se disponen de las medias y matrices de covarianza poblacionales, como generalmente ocurre en las aplicaciones, estos parámetros se reemplazan por sus respectivos estimadores inasiguados.

4 Ejemplo de aplicación de la metodología propuesta

El ejemplo que se mostrará se refiere a datos obtenidos de la bibliografía. Los datos (que son bivariados) corresponden a dos poblaciones para las que se reconstruyó la hipótesis de homogeneidad de matrices de covarianza. Por lo tanto, de acuerdo con lo expuesto en la introducción, bajo el supuesto de normalidad la función discriminante es una forma cuadrática, cuya estimación está dada por
\[
q(x) = -\frac{1}{2}(x - \hat{\Sigma}^{-1} x) - (\mu_1 - \mu_2)^T \hat{\Sigma}^{-1} x - \frac{1}{2} \ln \left( \frac{\hat{\Sigma}_1}{\hat{\Sigma}_2} \right) - \frac{1}{2} \left( \hat{\Sigma}_1 \hat{\Sigma}_2^{-1} \mu_1 - \hat{\Sigma}_2 \hat{\Sigma}_1^{-1} \mu_2 \right)
\]

Esta función y será graficada en el ejemplo para que sirva de referencia en la comparación de las funciones lineales que surgen como funciones discriminantes al aplicar el método propuesto en este trabajo y el método que presupone que \( \Sigma_1 = \Sigma_2 \). Para no complicar el gráfico, de los tres posibles hyperplanes (paralelos entre sí) propuestos sólo se mostrará el que pasa por el punto \( \beta_{12} \).

Ejemplo 4.1

Los datos usados en este ejemplo son datos (aproximados) extraídos del ejemplo 11.3.3 de la página 312 del libro de Marcin et al.[1982]. Estos datos fueron recolectados por Smith [1947] en dos grupos
de 25 individuos cada uno. El primer grupo de datos corresponde a información obtenida en individuos sicóticos, mientras que el segundo a individuos normales. En ambos grupos se realizaron mediciones en dos variables, datos que se muestran en la siguiente tabla:

<table>
<thead>
<tr>
<th>Sicóticos</th>
<th></th>
<th>Normales</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_1$</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>31</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>41</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Con estos datos resultan las siguientes estimaciones:

$$\hat{\beta}_1 = \begin{bmatrix} 12.70 \\ 35.20 \end{bmatrix}, \quad \hat{\Sigma}_1 = \begin{bmatrix} 36.27 & 13.05 \\ 13.05 & 290.50 \end{bmatrix}$$

$$\hat{\beta}_2 = \begin{bmatrix} 20.80 \\ 11.04 \end{bmatrix}, \quad \hat{\Sigma}_2 = \begin{bmatrix} 6.917 & 4.867 \\ 4.867 & 38.707 \end{bmatrix}$$

por lo que la estimación de la función discriminante dada en (29) es

$$\hat{g}_2(x) = (\hat{\beta}_1 - \hat{\beta}_2)' \hat{\Sigma}_1^{-1} \{ x - (\hat{\mu}_1 + 1 - \hat{\lambda}) \hat{\mu}_2 \} = (\hat{\beta}_1 - \hat{\beta}_2)' \hat{\Sigma}_1^{-1} \{ x - (\hat{\mu}_1 + 1 - \hat{\lambda}) \hat{\mu}_2 \}
\quad \text{(31)}$$

$$= -0.402 x_1 + 0.157 x_2 + 4.335$$

donde, en este caso,

$$\hat{\lambda} = \frac{2\hat{\beta}_1 \hat{\beta}_2}{(\hat{\beta}_1 + \hat{\beta}_2)^2} = 0.4402.$$

Igualando $\hat{g}_2(x)$ a 0 se obtiene la ecuación de la recta $r_2$

$$f_2(x_1) = 2.564 x_1 - 27.6536$$

Para calcular $q_1$, la función discriminante lineal bajo el supuesto $\Sigma_1 = \Sigma_2 = \Sigma$, dada en (6), se requiere el cálculo de la estimación $\hat{\Sigma}$ dada por

$$\hat{\Sigma} = \frac{(n_1 - 1) \hat{\Sigma}_1 + (n_2 - 1) \hat{\Sigma}_2}{n_1 + n_2 - 2} = \begin{bmatrix} 21.395 & 4.092 \\ 4.092 & 164.003 \end{bmatrix}.$$
de donde

\[ q_1(x) = (\mu_1 - \mu_2)^T \Sigma^{-1} x - \frac{1}{2} (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \]

\[ = -0.402 x_1 + 0.157 x_2 + 3.121, \]

que igualada a 0 es la recta \( r_1 \) de ecuación

\[ r_1(x_1) = 2.564 x_1 - 19.910. \]

Finalmente, la estimación de la función discriminante cuadrática (30) resulta

\[ q(x) = 0.065 x_1^2 + 0.021 x_1 x_2 + 0.012 x_2^2 - 3.206 x_1 - 0.621 x_2 + 34.839. \]

En el gráfico 4.1 se representan estas tres funciones discriminantes junto con los puntos que corresponden a las 50 observaciones. Además están representadas las curvas (elipsoides) de concentración del 60%. En este gráfico se puede observar que la recta discriminante \( r_1 \) que surge de \( q_1 \) clasifica bien a todos los puntos de los individuos normales y clasifica mal a cuatro puntos de los sicóticos; mientras que la recta \( r_2 \) que surge de \( q_2 \) clasifica mal dos puntos de los normales y dos de los sicóticos. Es decir, mientras que la recta \( r_1 \) que proviene de la teoría que presupone la igualdad de las matrices de covarianzas tiene una clara preferencia por los datos con mayor concentración, la recta \( r_2 \) que resulta de la propuesta hecha en este trabajo es más equitativa en la clasificación.

Gráfico 4.1: Datos del ejemplo 4.1 y funciones discriminantes lineales y cuadrática.

Elipses de concentración del 60%: Sicóticos (Sic.) y Normales (Nor.), hiperplanos: \( r_1 \) (Método de Fisher), \( r_2 \) (Método Propuesto) y elipses \( q \) (Método Cuadrático).
5 Simulación

Se realizó una simulación para comprobar las propiedades del hiperplano que se propone como separador de poblaciones con matrices de covarianza diferentes y, al mismo tiempo, analizar la capacidad de la distancia definida entre esas poblaciones para predecir la bondad del análisis discriminante lineal que surge del uso del mencionado hiperplano.

A primera vista hacer una simulación exhaustiva, que contemple el amplio espectro de variación de los parámetros involucrados, requeriría emplear tiempo y recursos que no se disponen. Sin embargo, este emprendimiento se simplifica notablemente si se tiene en cuenta que las clasificaciones (lineales) que surgen por ambos métodos son invariantes ante transformaciones lineales no singulares. Más precisamente, si se define la transformación lineal \( Z = L(X) = A'X + b \), con \( A \) no singular, donde \( X \sim (\mu_i, \Sigma_i) \) bajo \( \Pi_i \), entonces \( Z \sim (\nu_i, \Gamma_i) \) con \( \nu_i = A'\mu_i + b \) y \( \Gamma_i = A'\Sigma_i A \) para \( i = 1, 2 \). En consecuencia, usando cualquiera de los dos métodos de clasificación, si \( z_0 \) es un punto cualquiera (fijo) de \( \mathbb{R}^p \) tal que \( z_0 = L(x_0) \), entonces, para \( i = 1, 2 \), se verifica que \\
\[ x_0 \text{ se clasifica como perteneciente a la población transformada por } L \text{ de la } \Pi_i \iff x_0 \text{ y sólo si } x_0 \text{ se clasifica como perteneciente a la población } \Pi_i. \]

La matriz no singular \( A \) de la propiedad anterior se puede elegir de tal forma que \( A'\Sigma_1 A = I_{d_1} \) y \( A'\Sigma_2 A = D \) donde \( D \) es una matriz diagonal cuyos elementos en su diagonal principal son las raíces de la ecuación polinomial \( |\Sigma_2 - t \Sigma_1| = 0 \) (ver Graybill, F [1976], pág. 22). Por lo tanto, sin pérdida de generalidad, la simulación puede limitarse a considerar \( X \sim (\mu_1, I_{d_1}) \) cuando \( X \in \Pi_1 \) y \( X \sim (\mu_2, D) \) cuando \( X \in \Pi_2 \). Más aún, como la clasificación depende de la distancia entre las poblaciones y no de sus ubicaciones respecto de los ejes coordenados, siempre se puede elegir \( \mu_1 = 0 \) y \( \mu_2 \) con sólo su última componente distinta de 0.

No obstante la gran reducción del espacio de parámetros obtenida a partir del análisis previo, una simulación exhaustiva no es posible y se debe aceptar un compromiso entre lo deseable y lo realizable. Es por ello que la simulación realizada se limitó a considerar el caso de \( X \) vector bivariado con \( X \sim (0, I_{d_2}) \) bajo \( \Pi_1 \) y \( X \sim \left( \begin{array}{c} 0 \\ t \end{array} \right), \left( \begin{array}{cc} \delta_1 & 0 \\ 0 & \delta_2 \end{array} \right) \) bajo \( \Pi_2 \).

Con \( t, \delta_1, \delta_2 \in \mathbb{R}^n \).

La simulación usando el paquete Matlab consistió en los siguientes pasos

1. Se genera al azar un valor para \( \delta_1 \) y otro para \( \delta_2 \).
2. Se elige \( t \).
3. Se generan vectores bivariados \( E_1 \sim N(0, I_{d_1}) \) para \( i = 1, \ldots, 100 \).
4. Se eligen \( N_1 \) y \( N_2 \) con \( N_1 + N_2 = 100 \).
5. Los últimos \( N_2 \) datos generados en 3 se transforman usando una transformación apropiada, a saber, para \( i = N_1 + 1, \ldots, N \) se obtiene \\
\[ X_i^{(3)} = \left( \begin{array}{cc} \delta_1^{1/2} & 0 \\ 0 & \delta_2^{1/2} \end{array} \right) E_i + \left( \begin{array}{c} 0 \\ t \end{array} \right). \]
6. Se realizan las estimaciones $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\Sigma}_1$ y $\hat{\Sigma}_2$ con $\hat{\Sigma} = \frac{N_1\hat{\Sigma}_1 + N_2\hat{\Sigma}_2}{N_1 + N_2}$.

Usando estas estimaciones se clasifican los $N = 100$ datos generados usando tanto la metodología propuesta en este trabajo como la que presupone $\Sigma_1 = \Sigma_2 = \Sigma$.

7. Se cuentan las observaciones mal clasificadas en ambas poblaciones por cada uno de los dos métodos empleados.

8. Se calculan las distancias de Mahalanobis, Hellinger y $d^2$.

La siguiente es la descripción más detallada del método:

1. La matriz diagonal $\Sigma_2 = \text{diag}(\delta_1, \delta_2)$ se elige de tal manera que los datos de la población $\Pi_2$ sean más dispersos que los de la $\Pi_1$. Para ello, se genera inicialmente una matriz diagonal $B$ cuyo elemento lésimo de la diagonal principal es de la forma $1 + u_i$ con $u_i$ uniforme(0,1). La matriz $\Sigma_2$ se obtiene entonces por la relación $\Sigma_2 = A^2 I$, donde $A$ toma los valores 1, 2, 3, 4. De esta manera la matriz $\Sigma_2$ es para cada valor de $k$ una matriz acodada por $2^k I_{d_1}$ y por $2^{k+1} I_{d_2}$.

2. La segunda componente $t$ de $\mu_2$ toma los valores $c = 1, 2, 3, 4$. La elección de estos valores relativamente pequeños es para que se produzca una masa entre las observaciones de las poblaciones $\Pi_1$ y $\Pi_2$, de tal forma de obtener valores mal clasificados y poder comparar la eficiencia de los dos métodos.

3. La generación de vectores bivariados $N(0, I_{d_2})$ se realizó usando el corazon "randu" de Mathlib. El número $N = 100$ de vectores generados se divide en dos partes de igual tamaño para poder realizar $N_1$ y realizar estimaciones precisas y, a la vez, lo suficientemente pequeño como para no extender inoecesariamente la simulación.

4. $N_1$ toma los valores 25 a 50 en pasos de a.

5. Con la matriz de la población $\Pi_1$ se toman los subconjuntos $N_1$ vectores $\xi_1$ generados.

6. No se realiza la discriminación cuadrática por que el objetivo de este trabajo es comparar las dos metodologías lineales de clasificación.

7. Se calculan los porcentajes de observaciones mal clasificadas respecto del tamaño de cada muestra y respecto del total $N$.

8. La distancia de Mahalanobis $r = U \exp(-Q)$, como medida de la cercanía entre poblaciones dada en (22), tiene el defecto de tomar valores próximos a 0 para los valores de $Q$ que surgen de esta conclusión. En consecuencia, se adoptó como medida de Mahalanobis de la distancia entre las poblaciones a la cantidad $1 - U$. De esta forma se presentan tres distancias, la mencionada de Mahalanobis, $H$ de Hellinger, dada en (21) y la $d^2$ dada en (18). Los valores de estas distancias se calculan usando los parámetros teóricos así como también con las estimaciones que surgen para cada tamaño muestral.

Para cada selección de los pasos 1 y 2, se realizaron 500 repeticiones de los pasos 3 al 8. De esta manera, para cada $N_1$ se tuvieron 500 estimaciones de los porcentajes mencionados en 7 y de las distancias mencionadas en 8. De estos resultados, sólo se presentan en el Apéndice 2 dos grupos de tablas encabezadas por los valores de $k$ y $c$ usados. También se explica en cada caso la matriz $\Sigma_2$ generada.
Del análisis de los 18 grupos de tablas, resultados del total de la simulación realizada, se pueden extraer algunas conclusiones. En primer lugar, aunque no sorprendente pero sí destacable, se observa que la distancia d es la única de las tres consideradas que permite anticipar el comportamiento de las clasificaciones lineales. A mayor distancia menor porcentaje de observaciones mal clasificadas. Para mostrar esta característica de la distancia d se construyó la siguiente tabla en la que se muestran, ordenados de menor a mayor, los valores de porcentajes estimados totales de la primera fila de cada una de las tablas del apéndice. Junto a estos datos se presentan los correspondientes valores estimados de las tres distancias.

<table>
<thead>
<tr>
<th>(k; g)</th>
<th>% MAL CLASIF.</th>
<th>DISTANCIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Propuesto</td>
<td>Fisher</td>
</tr>
<tr>
<td>(4; 1)</td>
<td>36.03</td>
<td>36.67</td>
</tr>
<tr>
<td>(3; 1)</td>
<td>38.46</td>
<td>39.33</td>
</tr>
<tr>
<td>(2; 1)</td>
<td>37.69</td>
<td>38.25</td>
</tr>
<tr>
<td>(1; 1)</td>
<td>33.49</td>
<td>33.65</td>
</tr>
<tr>
<td>(4; 2)</td>
<td>32.61</td>
<td>33.58</td>
</tr>
<tr>
<td>(3; 2)</td>
<td>27.45</td>
<td>29.05</td>
</tr>
<tr>
<td>(2; 2)</td>
<td>24.10</td>
<td>25.69</td>
</tr>
<tr>
<td>(4; 4)</td>
<td>23.58</td>
<td>25.55</td>
</tr>
<tr>
<td>(4; 3)</td>
<td>22.82</td>
<td>24.98</td>
</tr>
<tr>
<td>(2; 3)</td>
<td>22.36</td>
<td>24.04</td>
</tr>
<tr>
<td>(3; 4)</td>
<td>18.38</td>
<td>21.41</td>
</tr>
<tr>
<td>(2; 2)</td>
<td>17.42</td>
<td>20.16</td>
</tr>
<tr>
<td>(3; 3)</td>
<td>15.36</td>
<td>16.53</td>
</tr>
<tr>
<td>(2; 4)</td>
<td>12.00</td>
<td>15.24</td>
</tr>
<tr>
<td>(1; 3)</td>
<td>9.01</td>
<td>9.54</td>
</tr>
<tr>
<td>(1; 4)</td>
<td>4.47</td>
<td>4.59</td>
</tr>
</tbody>
</table>

En la última columna de esta tabla se observa que salvo las filas correspondientes a (k; g) igual a (2; 2) y (4; 4), que debieron estar intercambias, los valores van de menor a mayor, correspondiendo a porcentajes de observaciones mal clasificadas que van de mayor a menor. Como puede observarse esto no ocurre con las columnas correspondientes a las otras dos distancias. Esto mismo puede observarse si en las tablas del Apéndice 2 se elige como porcentaje total estimado de observaciones mal clasificadas que corresponden a otro valor de \(N_1\).

El comportamiento de la distancia \(d^2\) es claro, ya que aumenta cuando las medias de la población se alejan y disminuye cuando aumenta la dispersión de los datos.

Con respecto al valor de \(\lambda\), en general, se observa que para una misma distancia entre las medias (c constante), cuanto mayor es la diferencia entre las concentraciones de las dos muestras menor es el valor de \(\lambda\) (más lejos está el punto de corte \(\beta_2\) del promedio de las medias poblacionales \(\mu_2\)). Sin embargo, para hacer un estudio más preciso de la vinculación entre los valores de \(\lambda\) y las características diferenciales de las matrices de covarianza de las dos poblaciones, se necesitaría programar una simulación diferente a la que se realizó en este trabajo apoyando a un objetivo distinto.

Finalmente, en lo que se refiere a la comparación entre los dos métodos lineales de discriminación, se puede concluir que con ambos se obtienen porcentajes de observaciones mal clasificadas similares.
(en la simulación realizada la diferencia entre los porcentajes nunca superó el 5%). Sin embargo, los porcentajes estimados de observaciones mal clasificadas en cada una de las muestras presentan grandes diferencias según el método que se adopte. Por un lado, el método propuesto, primero valores de cada fila encabezada por $N_1$ y $\lambda$ en el Apéndice 2, mantiene una equidad entre los porcentajes de cada muestra (no diferente del porcentaje total), mientras que el método de Fisher, segundos valores de cada fila encabezada por $N_1$ y $\lambda$ en el Apéndice 2, presenta una marcada preferencia hacia la muestra más concentrada. En general, para una misma separación entre las medias (c constante) a mayor diferencia de concentración entre las muestras (menor distancia $d^2$) se observa un aumento en la diferencia entre el porcentaje de las muestras y el porcentaje total. Estas diferencias se acrecentan a medida que el tamaño de la muestra más concentrada ($N_1$) disminuye respecto del tamaño de la muestra menos concentrada ($N_2$).

Teniendo en cuenta el porcentaje total de observaciones mal clasificadas, el método propuesto es ligeramente mejor que el de Fisher cuando el tamaño de la muestra más concentrada ($N_1$) es menor que el tamaño de la muestra menos concentrada ($N_2$). Esta pequeña superioridad aumenta cuanto mayor es la diferencia entre los tamaños muestrales.

**Bibliografía**


6 Apéndice 1

Propiedad 1: La separación entre $f_1^{(1)}$ y $f_2^{(2)}$ es invariante ante transformaciones lineales de la forma $Z = A'X + b$ donde $A$ es una matriz $p \times p$ no singular ($\text{rang}(A) = p$).

Demonstración: si se indican por $f_1^{(1)}$ y $f_2^{(2)}$ son las correspondientes funciones densidades de $Z$, entonces como

\[
\theta_1 = E[Z|\Pi_1] = A'\mu_1 + b
\]
\[
\theta_2 = E[Z|\Pi_2] = A'\mu_2 + b
\]
\[\Gamma = \text{cov}(Z, Z|\Pi_1) = \text{cov}(Z, Z|\Pi_2) = A'\Sigma A,
\]
entonces resulta que

\[
\text{sep}(f_1^{(1)}; f_2^{(2)}) = \delta_1(\theta_1; \theta_2) = (\theta_1 - \theta_2)'\Gamma^{-1}(\theta_1 - \theta_2)
\]
\[= [A'(\mu_1 - \mu_2)]'\Gamma^{-1}[A'(\mu_1 - \mu_2)]
\]
\[= (\mu_1 - \mu_2)'A(A'SA)^{-1}A'(\mu_1 - \mu_2)
\]
\[= (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)
\]

(32)

Propiedad 2: Para transformaciones lineales de la forma $Z = A'X + b$ donde $A$ es una matriz $p \times m$ con $\text{rang}(A) = m$ se verifica que

\[\text{sep}(f_1^{(1)}; f_2^{(2)}) = \max_{M \in M(p, m)} \text{sep}(f_1^{(1)}; f_2^{(2)}).
\]

donde $M(p, m)$ es el conjunto de todas las matrices de orden $p \times m$ de rango $m$.

Demostración: usando la notación dada por (32), entonces primero que la matriz de varianza-covarianza de $Z$ es no singular por ser $\text{rang}(\Gamma) = m$, por lo tanto como antes se tiene

\[
\text{sep}(f_1^{(1)}; f_2^{(2)}) = \delta_1(\theta_1; \theta_2) = (\theta_1 - \theta_2)'\Gamma^{-1}(\theta_1 - \theta_2)
\]
\[= [A'(\mu_1 - \mu_2)]'\Gamma^{-1}[A'(\mu_1 - \mu_2)]
\]
\[= (\mu_1 - \mu_2)'A(A'SA)^{-1}A'(\mu_1 - \mu_2).
\]

Ahora bien, como $\Sigma$ es definida positiva entonces existe una matriz $B$ no singular de orden $p \times p$ tal que $\Sigma = B'B$, entonces la ecuación anterior resulta

\[
\text{sep}(f_1^{(1)}; f_2^{(2)}) = (\mu_1 - \mu_2)'B^{-1}BA(A'B'BA)^{-1}A'B'B^{-1}(\mu_1 - \mu_2)
\]
\[= S'M(M'M)^{-1}M'S,
\]

donde se ha llamado $\delta$ a $B^{-1}(\mu_1 - \mu_2)$ y $M$ a $BA$. Por lo tanto el problema planteado consiste en probar que

\[
\max_{M \in M(p, m)} S'M(M'M)^{-1}M'S = \text{sep}(f_1^{(1)}; f_2^{(2)})
\]
\[= (\mu_1 - \mu_2)'(B'B)^{-1}(\mu_1 - \mu_2)
\]
\[= S'\delta.
\]
Como la matriz $P = M(M'M)^{-1}M'$ es una matriz simétrica e idempotente, o sea $P$ es una matriz de proyección ortogonal, se verifica que

$$\delta' MM^{-1} \delta \leq \delta' \delta,$$

ya que

$$\delta' M(M'M)^{-1}M' \delta = \delta' P \delta \leq \|\delta\|^2 + P\|\delta\| \cos\theta \leq \|\delta\|^2 = \delta' \delta.$$

Por lo tanto, el problema se reduce a encontrar una matriz $M \in M(p,m)$ tal que verifique la igualdad en (33). Es claro que esto es equivalente al siguiente problema: Dadas $\delta \in \mathbb{R}^p$ y $\omega \in \mathbb{R}^m$ fijos, encontrar una matriz $M \in M(p,m)$ tal que $\delta = M\omega$, y ya que, si esto es el caso, reemplazando en el lado izquierdo de (33) $\delta$ por $M\omega$ resulta

$$\delta' M(M'M)^{-1}M' \delta = (M\omega)' M(M'M)^{-1}M'(M\omega) = \omega' M'M\omega = \delta' \delta.$$

Como este problema tiene infinitas soluciones cuando $m < p$ la propiedad está demostrada.

Propiedad 3: $c_\rho$ es el punto que verifica que

$$\min_{x \in \mathbb{R}^p} \delta^2(\mu_1; x) + \delta^2(\mu_2; x) = \delta^2(\mu_1; c_\rho) + \delta^2(\mu_2; c_\rho),$$

se tiene además que

$$\min_{x \in \mathbb{R}^p} \delta^2(\mu_1; x) + \delta^2(\mu_2; x) = \frac{1}{2} \delta^2(\mu_1; \mu_2).$$

Demostración: indiquemos con $h$ a la función dada por

$$h(x) = \delta^2(\mu_1; x) + \delta^2(\mu_2; x) = (\mu_1 - x)' \Sigma^{-1}(\mu_1 - x) + (\mu_2 - x)' \Sigma^{-1}(\mu_2 - x).$$

Derivando $h$ e igualando a cero se tiene

$$\frac{dh}{dx} = 2[\Sigma^{-1}(\mu_1 - x) + \Sigma^{-1}(\mu_2 - x)] = 2[\Sigma^{-1}(\mu_1 + \mu_2) - 2x] = 0,$$

de donde el mínimo se obtiene en el punto $x$ tal que $2x = (\mu_1 + \mu_2)$, o sea en $c_\rho$.

Finalmente, como $\mu_1 - c_\rho = \frac{1}{2} (\mu_1 - \mu_j)$, para $i \neq j \in \{1, 2\}$, reemplazando esta expresión en el segundo miembro de (10) resulta (11). Convie ne notar aquí que esta propiedad dada en (11) es válida para cualquier par de puntos a y b, no necesariamente para las medias $\mu_1$ y $\mu_2$.

Propiedad 4: $d^2_{ij}(a; b) = (a - b)' (E_{ij} + E_{ji})^{-1} (a - b)$.

Demostración: $c_{ij} = B_{ja} + B_{b}$ con $B_1 + B_2 - I$ donde $B_k = \Sigma_k (\Sigma_k + \Sigma_j)^{-1}$ para $k = 1, 2$. Entonces, como $a - c_{ij} = B_i(a - b)$ y $b - c_{ij} = -B_j(a - b)$, reemplazando en (12) se tiene

$$\frac{1}{2} d^2_{ij}(a; b) = \min_{x \in \mathbb{R}^p} \delta^2(\mu_1; x) + \delta^2(\mu_2; x)$$

$$= \delta^2(\mu_1; c_{ij}) + \delta^2(\mu_2; c_{ij})$$

$$= (a - c_{ij})' \Sigma^{-1}(a - c_{ij}) + (b - c_{ij})' \Sigma^{-1}(b - c_{ij})$$

$$= (a - b)' B_{ij} \Sigma_{ij}^{-1} B_{ij}(a - b) + (a - b)' B_{kj} \Sigma_{kj}^{-1} B_{kj}(a - b)$$

$$= (a - b)' (E_{ij} + E_{ji})^{-1} (a - b).$$
\[
\begin{align*}
&= (a - b)(\Sigma_i + \Sigma_j)^{-1}\Sigma_i\Sigma_i^{-1}\Sigma_i(\Sigma_i + \Sigma_j)^{-1} + \\
&\quad (\Sigma_i + \Sigma_j)^{-1}\Sigma_i\Sigma_i^{-1}\Sigma_i(\Sigma_i + \Sigma_j)^{-1}(a - b) \\
&= (a - b)(\Sigma_i + \Sigma_j)^{-1}\Sigma_i(\Sigma_i + \Sigma_j)^{-1} + \\
&\quad (\Sigma_i + \Sigma_j)^{-1}\Sigma_i(\Sigma_i + \Sigma_j)^{-1}(a - b) \\
&= (a - b)(\Sigma_i + \Sigma_j)^{-1}(a - b),
\end{align*}
\]
de donde se obtiene que para \(i \neq j \in \{1, 2\}\) es
\[
d^2_{ij}(a; b) = 2(a - b)(\Sigma_i + \Sigma_j)^{-1}(a - b) = (a - b) \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1}(a - b).
\]

Propiedad 5: La \(sep(f_{ij}^{(1)}; f_{ij}^{(2)})\) es máxima para cualquier combinación lineal \(Z = a'X + b\), donde 
\(a = k \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2)\), con \(k \neq 0\).

Demostración: para \(i = 1\) y \(2\), la media \(\eta_i\) y la varianza \(\sigma_i^2\) de una tasa combinación lineal \(Z\) cuando el vector \(X\) proviene de la población \(\Pi_2\) se calculan como en (3), a saber:
\[
\begin{align*}
\eta_1 &= (\mu_1 - \mu_2)^{\prime}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}\mu_1 \\
\eta_2 &= (\mu_1 - \mu_2)^{\prime}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}\mu_2 \\
\sigma_1^2 &= (\mu_1 - \mu_2)^{\prime}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}\Sigma_1\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2), \\
\sigma_2^2 &= (\mu_1 - \mu_2)^{\prime}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}\Sigma_2\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2). \\
\end{align*}
\]
En consecuencia, el problema se resuelve utilizando el teorema 2.1 de la siguiente manera
\[
\begin{align*}
\max_{a \in \mathbb{R}^{(1 \times 2)}} sep(f_{ij}^{(1)}; f_{ij}^{(2)}) &= \max_{a \in \mathbb{R}^{(1 \times 2)}} d^2(\eta_1; \eta_2) \\
&= \max_{a \in \mathbb{R}^{(1 \times 2)}} (\eta_1 - \eta_2)^{\prime}\left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right)^{-1}(\eta_1 - \eta_2) \\
&= \max_{a \in \mathbb{R}^{(1 \times 2)}} \frac{(a'\mu_1 - \mu_2)^2}{a'\Sigma_1a} \\
&= (\mu_1 - \mu_2)^{\prime}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2) \\
&= d^2(\mu_1; \mu_2) = sep(f_{ij}^{(1)}; f_{ij}^{(2)}). \\
a &= k \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(\mu_1 - \mu_2),
\end{align*}
\]
para cualquier \(k \neq 0\).
7 Apéndice 2

Caso $k = 2$ y $c = 3$.

Matriz aleatoria $\Sigma_1 = \begin{bmatrix} 1.9826 & 0 \\ 0 & 1.6015 \end{bmatrix}$

<table>
<thead>
<tr>
<th>$N_1$</th>
<th>1 - U</th>
<th>H</th>
<th>$d^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.3183</td>
<td>0.7634</td>
<td>3.0968</td>
</tr>
<tr>
<td>30</td>
<td>0.3342</td>
<td>0.7792</td>
<td>3.2614</td>
</tr>
<tr>
<td>35</td>
<td>0.3318</td>
<td>0.7799</td>
<td>3.2748</td>
</tr>
<tr>
<td>40</td>
<td>0.3281</td>
<td>0.7828</td>
<td>3.3027</td>
</tr>
<tr>
<td>45</td>
<td>0.3255</td>
<td>0.7831</td>
<td>3.3095</td>
</tr>
<tr>
<td>50</td>
<td>0.3245</td>
<td>0.7853</td>
<td>3.3153</td>
</tr>
</tbody>
</table>

| $N_1$ | $\lambda$ | MUESTRA 1 | | MUESTRA 2 | | TOTAL |
|-------|------------|-----------|-----------|-----------|--------|
| 25    | 0.4417     | 17.3040   | | 17.4640   | | 17.4240|
|       | 8.6480     | 23.9973   | | 17.4240   | | 20.1600|
| 30    | 0.4425     | 17.3067   | | 17.5486   | | 17.4760|
|       | 8.6667     | 24.0086   | | 17.4760   | | 19.4000|
| 35    | 0.4441     | 17.1543   | | 17.4031   | | 17.3160|
|       | 8.6343     | 23.9138   | | 17.3160   | | 18.5660|
| 40    | 0.4449     | 17.2950   | | 17.2833   | | 17.2840|
|       | 8.5250     | 23.5867   | | 17.2840   | | 17.8020|
| 45    | 0.4454     | 17.2622   | | 17.3273   | | 17.2980|
|       | 8.5333     | 24.1236   | | 17.2980   | | 17.1680|
| 50    | 0.4456     | 17.3800   | | 17.4280   | | 17.4040|
|       | 8.4540     | 24.1680   | | 17.4040   | | 16.3100|
Caso $k = 1$ y $c = 4$.

Matriz aleatoria $\Sigma_1 = \begin{bmatrix} 1.2749 & 0 \\ 0 & 1.4565 \end{bmatrix}$

<table>
<thead>
<tr>
<th>$N_1$</th>
<th>1 - $U$</th>
<th>$H$</th>
<th>$d^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1345</td>
<td>1.4862</td>
<td>11.2430</td>
</tr>
<tr>
<td>25</td>
<td>0.1535</td>
<td>1.5416</td>
<td>11.6693</td>
</tr>
<tr>
<td>30</td>
<td>0.1900</td>
<td>1.5639</td>
<td>11.9081</td>
</tr>
<tr>
<td>35</td>
<td>0.1486</td>
<td>1.5933</td>
<td>11.6130</td>
</tr>
<tr>
<td>40</td>
<td>0.1468</td>
<td>1.5381</td>
<td>11.6474</td>
</tr>
<tr>
<td>45</td>
<td>0.1452</td>
<td>1.5361</td>
<td>11.6390</td>
</tr>
<tr>
<td>50</td>
<td>0.1442</td>
<td>1.5378</td>
<td>11.6577</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N_1$</th>
<th>$\lambda$</th>
<th>MUESTRA 1</th>
<th>MUESTRA 2</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.4961</td>
<td>4.1840</td>
<td>4.5680</td>
<td>4.4720</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0940</td>
<td>4.5733</td>
<td>4.5860</td>
</tr>
<tr>
<td>30</td>
<td>0.4964</td>
<td>4.1600</td>
<td>4.4600</td>
<td>4.3700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3533</td>
<td>4.6600</td>
<td>4.6680</td>
</tr>
<tr>
<td>35</td>
<td>0.4967</td>
<td>4.3200</td>
<td>4.5231</td>
<td>4.4520</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3143</td>
<td>4.5492</td>
<td>4.6260</td>
</tr>
<tr>
<td>40</td>
<td>0.4968</td>
<td>4.3300</td>
<td>4.4700</td>
<td>4.4140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1000</td>
<td>4.9533</td>
<td>4.6120</td>
</tr>
<tr>
<td>45</td>
<td>0.4970</td>
<td>4.3280</td>
<td>4.5345</td>
<td>4.4420</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.8800</td>
<td>5.1127</td>
<td>4.5580</td>
</tr>
<tr>
<td>50</td>
<td>0.4970</td>
<td>4.3840</td>
<td>4.6580</td>
<td>4.4260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.6780</td>
<td>5.2800</td>
<td>4.4780</td>
</tr>
</tbody>
</table>