PROPUESTA Y EVALUACIÓN DE UNA REGLA DE CLASIFICACIÓN CUANDO LOS DOS O TRES PRIMEROS COMPONENTES PRINCIPALES EXPLICAN UNA VARIABILIDAD ADMISIBLE

POR:
ALVARO BURBANO MONTENEGRO

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ - COLOMBIA
2009

PROPUESTA Y EVALUACIÓN DE UNA REGLA DE CLASIFICACIÓN CUANDO LOS DOS O TRES PRIMEROS COMPONENTES PRINCIPALES EXPLICAN UNA VARIABILIDAD ADMISIBLE

POR:
ALVARO BURBANO MONTENEGRO

Trabajo de grado presentado como requisito para optar al titulo de Magister en Ciencias Estadísticas

DIRECTOR:
JORGE HUMBERTO MAYORGA M.Sc.

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ - COLOMBIA
2009

Nota de Aceptación

El trabajo de tesis desarrollado bajo el nombre de "Propuesta y evaluación de una regla de clasificacíon cuando los dos tres componentes principales explican una variabilidad admisible ." Presentado como requisito parcial para optar al titulo de Magíster en Ciencias Estadísticas fue aprobado por su director y jurado calificador.

JORGE HUMBERTO MAYORGA M.Sc.
Director

Presidente del jurado

Firma del jurado

Firma del jurado

Bogota D.C, 9 de diciembre de 2009

Dedicatoria

A la memoria de mi madre Graciela, mis hermanos y sobrinos.

Agradecimientos

Agradezco especialmente a Jorge Humberto Mayorga, mi director, por sus sugerencias y constante ayuda durante la investigación.

UNIVERSIDAD DE NARIÑO, Departamento de Matemáticas y Estadística.

A todas aquellas personas que de una u otra forma colaboraron en la realización de esta investigación.

Índice

1. Introducción **1**

2. Elementos conceptuales preliminares **2**

2.1. Procedimientos de clasificacion 2

2.2. Discriminación en dos poblaciones normales 5

2.3. Evaluación de la regla de clasificación 7

3. Propuesta de una nueva regla de clasificación **10**

3.1. Propuesta inicial 10

3.2. Estructura de la propuesta inicial 11

3.2.1. Simulación de la nueva regla de clasificación para dos poblaciones 13

3.2.2. Modificación de la regla inicial 14

4. Exploración del desempeño de la nueva regla de clasificación **16**

4.1. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la primera pauta. 18

4.2. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la segunda pauta. 23

4.3. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la tercera pauta. 30

5. Conclusiones **36**

A. Programas de simulación para la clasificación de unidades estadísticas usando la regla de puntajes lineales discriminantes y la regla propuesta para las diferentes pautas **39**

B. Procedimiento para encontrar el vector de la población dos a partir de una distancia determinada con relación a el vector de la población **50**

C. Resultados de la simulación de la clasificación de unidades estadísticas por la regla propuesta y la regla de puntajes lineales discriminante para dos y tres componentes principales **52**

D. Resultados de la simulación de la clasificación de unidades estadísticas por la regla propuesta y la regla de puntajes lineales discriminante **65**

Índice de figuras

4.1. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la probabilidad a priori . . . 19

4.2. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides . 20

4.3. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides 21

4.4. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y número de variables 22

4.5. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides . 23

4.6. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides 24

4.7. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la probabilidad a priori . . . 25

4.8. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides . 26

4.9. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides 27

4.10. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y número de variables 28

4.11. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides . 29

4.12. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides 29

4.13. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides 30

4.14. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y el porcentaje de variabilidad 31

4.15. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el probabilidad a priori y el número de variables . . . 32

4.16. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la distancia y el porcentaje de variabilidad 33

4.17. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el porcentaje de variabilidad y el número de variables 34

4.18. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la distancia y el número de variables 35

Capítulo 1

Introducción

El problema de la clasificación está presente en un amplio rango de actividades humanas. En su forma más general el término puede cubrir cualquier contexto en el que se toma una decisión o se realiza una predicción con base en una información disponible en ese momento. La toma de decisiones para otorgar un crédito, de acuerdo a una información financiera, hasta la asignación de un tratamiento a un paciente cuyo diagnostico se hizo con base en una información previa, son dos situaciones, a manera de ejemplo, que pueden entenderse en este contexto.

Particularmente el análisis discriminante es una técnica de clasificación de elementos u objetos en las que se presume la existencia de dos o más clases, describe diferencias existentes entre esas clases en base a valores que toman ciertos rasgos sobre objetos de cada una de esas clases, clasifica nuevos objetos en algunas de esas clases preexistentes en función de los valores que toman ciertas variables para esos objetos.

Este trabajo propone un procedimiento de discriminación estadística, cuya regla de clasificación se evaluo frente a la regla lineal discriminante, cuando con dos o tres componentes principales se retiene una variabilidad aceptable y la clasificación se circunscribe a dos poblaciones.

Se recurrió a la simulación como un medio experimental para llevar a cabo el cotejo de la propuesta frente a la regla lineal discriminante, como etapa previa a una evaluación analítica, acopiando de esta manera hallazgos que dicten derroteros para los análisis teóricos posteriores que finalmente certifiquen la idoneidad clasificatoria de la propuesta.

El primer capítulo presenta una síntesis de la clasificación bajo normalidad, el segundo presenta la estructura y modificaciones de la propuesta, el tercer capítulo explora el desempeño de la regla propuesta frente a la regla de puntajes lineales discriminantes en tres escenarios diferentes, y finalmente se presenta las conclusiones que dan lugar el presente trabajo.

Capítulo 2

Elementos conceptuales preliminares

La discriminación y la clasificación son técnicas multivariadas usadas para separar distintos conjuntos de objetos u observaciones y para la ubicación de los mismos a diferentes grupos previamente definidos. La función de separar objetos muchas veces sirve para ubicarlos, así mismo una regla de ubicación o clasificación puede sugerir un procedimiento de discriminación. Según Afifi y Clark (1996), Habbema y Hermans (1977) se pueden distinguir dos enfoques diferentes al problema de clasificación: Cuando los grupos están bien definidos y se trata de determinar un criterio para etiquetar cada individuo como perteneciente a algún grupo a partir de los valores de una serie limitada de variables. En este caso las técnicas más utilizadas se conocen como: Análisis discriminante(aprendizaje supervisado). El segundo enfoque es el que tiene que ver cuando a priori no se conocen los grupos y lo que precisamente se desea es establecerlos a partir de los datos que se tengan (por ejemplo el caso taxonómico); las técnicas más utilizadas para este propósito se conocen como Análisis de *cluster*.

2.1. Procedimientos de clasificacion

Según Brown, Fearn, Haque (1999). existen varios métodos de clasificación dependiendo del número de grupos a clasificar, de las hipótesis hechas acerca del comportamiento de las variables en cada grupo (normalidad conjunta, homoscedasticidad) así como del criterio utilizado para llevar a cabo dicha clasificación.

Inicialmente cuando la clasificación involucra solamente dos poblaciones con las funciones de densidad de probabilidad $f_1(x)$ y $f_2(x)$ del vector aleatorio p×1 de X para las poblaciones π_1 y π_2 respectivamente. Un objeto con medidas asociadas a X debe de ser asignado a las poblaciones π_1 o π_2. Sea Ω el espacio de todas las posibles observaciones de X y R_1 el conjunto de todos los valores de X para los cuales se clasifican los objetos en la población π_1 y $R_2 = \Omega - R_1$ los valores restantes de X para los cuales se clasifican los objetos en la población π_2. Como cada objeto debe ser asignado a una y sola una de las poblaciones, los conjuntos R_1 y R_2 son mutuamente excluyentes.

La regla como mecanismo clasificador puede equivocarse en la ubicación de unidades y por lo consiguiente se debe de considerar la probabilidad condicional $P(2|1)$ de clasificar un objeto en π_2 cuando en realidad pertenece a π_1 . De acuerdo con Seber (1984), Huberty (1994), Johnson y Wichern (1999) entre otros, estas probabilidades están dadas por:

$$P(2|1) = P(X \in R_2|\pi_1) = \int_{R_2=\Omega-R_1} f_1(x)dx \tag{2.1}$$

y similarmente la probabilidad condicional $P(1|2)$ de clasificar un objeto en π_1 cuando en realidad pertenece a π_2 es:

$$P(1|2) = P(X \in R_1|\pi_2) = \int_{R_1} f_2(x)dx \tag{2.2}$$

Se considera $p_i = p(\pi_i)$ $i = 1, 2$ a las probabilidades a priori de que la observación considerada pertenezca a cada grupo. Por consiguiente la probabilidad de clasificar incorrectamente una unidad de la población π_1 es por tanto:

$$P(X \in R_1|\pi_2)P(\pi_2) = P(2|1)P_1 \tag{2.3}$$

de igual manera la probabilidad de clasificar incorrectamente la unidad en la población π_2 es:

$$P(X \in R_2|\pi_1)P(\pi_1) = P(1|2)P_2 \tag{2.4}$$

Los aciertos de la regla pueden igualmente ser cuantificados por medio de las probabilidades de correcta clasificación:

$$P(X \in R_1|\pi_1)P(\pi_1) = P(1|1)P_1 \quad y \quad P(X \in R_2|\pi_2)P(\pi_2) = P(2|2)P_2 \tag{2.5}$$

A menudo se clasifica en términos de las probabilidades de la clasificación incorrecta, de acuerdo con King L (1970), O'Gorman y Woolson (1991), la regla de clasificación puede tener otros elementos en su configuración relativos a los costos y éstos se pueden denotar como sigue en la siguiente matriz de costos. Johnson y Wichern (1999). Los costos son: 0 para una clasificación correcta, $C(1|2)$ cuando una observación que pertenece a π_2 es clasificada incorrectamente en π_1 y $C(2|1)$ cuando una observación que pertenece a π_1 es clasificada incorrectamente en π_2 .

Globalmente a la regla de clasificación se le puede determinar una medida de eficiencia a través de los costos de clasificación incorrecta que se puede denominar como ECM (costo esperado de clasificación incorrecta).

$$ECM = C(2|1)P(2|1)P_1 \ + \ C(1|2)P(1|2)P_2 \tag{2.6}$$

La regla será admisible para un ECM mínimo. Las regiones que minimizan al ECM están definidas por los valores de x (valores observados del vector aleatorio X) para los cuales las siguientes desigualdades se mantienen:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq [\frac{C(1|2)}{C(2|1)}][\frac{P_2}{P_1}] \tag{2.7}$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < [\frac{C(1|2)}{C(2|1)}][\frac{P_2}{P_1}] \tag{2.8}$$

En consecuencia la nueva observación con información x_0, es clasificada según la siguiente regla: τ :*"Ubíquese la unidad estadística con información x_0, en la población π_1 si:*

$$\frac{f_1(x_0)}{f_2(x_0)} \geq [\frac{C(1|2)}{C(2|1)}][\frac{P_2}{P_1}] \tag{2.9}$$

.

En caso contrario: Ubíquese en la población π_2".

Cuando las probabilidades a priori son desconocidas a menudo se toman como iguales y si los costos de clasificación incorrecta son iguales o si la proporción de las probabilidades a priori son iguales al inverso de la proporción de los costos de clasificación incorrecta entonces:

$$R_1 : \frac{f_1(x_0)}{f_2(x_0)} \geq 1 \tag{2.10}$$

$$R_2 : \frac{f_1(x_0)}{f_2(x_0)} < 1 \tag{2.11}$$

Otra medida de eficiencia de una regla de clasificación es la probabilidad total de clasificación incorrecta denotada por TPM y que se puede calcular como:

$$TPM = P_1 \int_{R_2} f_1(x)dx + P_2 \int_{R_1} f_2(x)dx \tag{2.12}$$

Igualmente es razonable admitir una regla de clasificación con TPM mínimo. La regla de clasificación que minimiza la TPM es equivalente a la regla que minimiza a ECM para el caso en el cual $C(1|2) = C(2|1)$ es decir que la regla será:

τ :*"Ubíquese la unidad estadística con información x_0, en la población π_1 si:*

$$\frac{f_1(x_0)}{f_2(x_0)} \geq \frac{P_2}{P_1} \tag{2.13}$$

.

En caso contrario: Ubíquese en la población π_2".

El teorema de Bayes permite en este contexto establecer las prioridades a posteriori.

$$P(\pi_1|x_0) = \frac{P_1 f_1(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \qquad\qquad P(\pi_2|x_0) = \frac{P_2 f_2(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \tag{2.14}$$

En consecuencia la regla de clasificación basada en estas probabilidades a posteriori será:

τ :*"Ubíquese la unidad estadística con información x_0, en la población π_1 si:*

$$P(\pi_1|x_0) \geq P(\pi_2|x_0). \tag{2.15}$$

En caso contrario: Ubíquese en la población π_2".

2.2. Discriminación en dos poblaciones normales

En la práctica los procesos de clasificación basados en poblaciones normales, son los que más predominan debido a que son simples y altamente eficientes entre una gran variedad de modelos de población y están incluidos en la mayoría de programas estadísticos que forman parte del software actual. Se asume que $f_1(x)$ y $f_2(x)$ son densidades normales multivariadas, la primera con vector de medias μ_1 y matriz de covarianzas $\sum_1$, la segunda con vector de medias μ_2 y matriz de covarianzas $\sum_2$.

La clasificación sustentada por el modelo normal es corrientemente utilizada y parte por asumir según Peña (2002), que para la población π_i , el vector X se distribuye $N \sim (\mu_i, \sum_i)$ $i = 1,\ 2$ esto es:

$$f_i(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\sum_i|^{\frac{1}{2}}} \quad exp \quad \{-\frac{1}{2}(x-\mu_i)' {\sum}_i^{-1} (x-\mu_i)\} \tag{2.16}$$

Un caso especial lo constituye cuando las matrices cuyas covarianzas son iguales, lo cual conduce al uso de la función discriminante lineal. De igual manera existe el caso para el cual las matrices de covarianzas son diferentes, lo cual conduce al uso de la función discriminante cuadrática. Critchey y Ford (1985)

Debido a que la propuesta que se pretende desarrollar en el presente trabajo conducirá a una nueva regla de clasificación para ser evaluada frente a la regla que se basa en el puntaje lineal discriminante, se hará especial énfasis en la homoscedasticidad y en el puntaje lineal discriminante.

En efecto, cuando se tiene que: $\sum_1 \ = \ \sum_2 \ = \ \sum$, y suponiendo que las densidades de $X' = (X_1\ , X_2\ , ..., X_p)$ para las poblaciones π_1 y π_2 están dadas por la ecuación 2.16 para i = 1 , 2, las regiones que minimizan ECM son:

$$R_1 : exp[-\frac{1}{2}(x-\mu_1)'\sum^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)'\sum^{-1}(x-\mu_2)] \geq (\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.17}$$

$$R_2 : exp[-\frac{1}{2}(x-\mu_1)'\sum^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)'\sum^{-1}(x-\mu_2)] < (\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.18}$$

Debido a que las cantidades en 2.17 y 2.18 son no negativas para todo vector particular x de X se puede tomar logaritmo natural conservando el orden de la desigualdad y consecuentemente:

$$R_1 : [(\mu_1-\mu_2)'\sum^{-1}x_0 - \frac{1}{2}(\mu_1-\mu_2)'\sum^{-1}(\mu_1+\mu_2)] \geq ln(\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.19}$$

$$R_2 : [(\mu_1-\mu_2)'\sum^{-1}x_0 - \frac{1}{2}(\mu_1-\mu_2)'\sum^{-1}(\mu_1+\mu_2)] < ln(\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.20}$$

Dadas estas regiones R_1 y R_2 se puede construir la correspondiente regla de clasificación:

τ :*"Ubíquese la unidad estadística con información x_0, en la población π_1 si:*

$$(\mu_1-\mu_2)'\sum^{-1}x_0 - \frac{1}{2}(\mu_1-\mu_2)'\sum^{-1}(\mu_1+\mu_2) \geq C, \tag{2.21}$$

siendo $C = ln[\frac{c(1|2)P_2}{c(2|1)P_1}]$,en caso contrario: Ubíquese en la población π_2".

Como la anterior regla supone el conocimiento de los vectores μ_1, μ_2 y de la matriz $\sum$ se constituye en una situación ideal.

En la práctica, la información de p variables en n_1 individuos que pertenecen realmente a la población π_1 y la información de las mismas p variables en n_2 individuos que pertenecen realmente a la población π_2, se representa por medio de dos vectores de datos X_1 y X_2. Expresando estos vectores como: $X_1' = (X_{11}, X_{12}, ..., X_{1n_1})$ y $X_2' = (X_{21}, X_{22}, ..., X_{2n_2})$, los vectores de medias muestrales y las matrices de covarianzas muestrales están determinados por:

$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$, $\overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$, $S_1 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)(X_{1i}-\overline{X}_1)'$ *y*
$S_2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\overline{X}_2)(X_{2i}-\overline{X}_2)'$.

Cuando se supone homoscedasticidad se estima $\sum$ por medio de la matriz:

$$S_p = \frac{n_1-1}{n_1+n_2-2}S_1 \quad + \quad \frac{n_2-1}{n_1+n_2-2}S_2, \tag{2.22}$$

la cual combina las matrices S_1 y S_2 para derivar un único estimador insesgado de $\sum$. La regla de clasificación a partir de la muestra o la regla *plug-in* sustituye los vectores y la matriz en mención por sus respectivos vectores de estimaciones. En consecuencia esta regla se enuncia así:

τ :*"Ubíquese la unidad estadística con información x_0, en la población π_1 si:*

$$(\overline{x}_1 - \overline{x}_2)' S_p^{-1} x_0 - \frac{1}{2}(\overline{x}_1 - \overline{x}_2)' S_p^{-1}(\overline{x}_1 + \overline{x}_2) \geq C \tag{2.23}$$

siendo $C = ln[\frac{c(1|2)P_2}{c(2|1)P_1}]$,*en caso contrario: Ubíquese en la población π_2".*

2.3. Evaluación de la regla de clasificación

La forma de juzgar el desempeño de un procedimiento de clasificación, es evaluando la magnitud del error en su función clasificadora, es decir las probabilidades de clasificación incorrecta. Cuando se asume un modelo probabilistico y se conoce la totalidad de sus parámetros, las probabilidades de clasificación incorrecta pueden ser calculadas con relativa facilidad. Sin embargo en la práctica cuando esto no es factible, se hace necesario entonces concentrarse en las proporciones del error, asociadas con la función de clasificación en la muestra. Morrison (1969).

Hay tres tipos de error que pueden ser asociados a las reglas de clasificación: error verdadero, error actual, y error aparente. El error verdadero o también conocido como error óptimo es la frecuencia de clasificación incorrecta usando una regla de clasificación que asume conocidos los parámetros de la población. Tanto el error actual como el error aparente emplean una regla que usa los estimadores de los parámetros desconocidos, el primero se refiere a la frecuencia de clasificación incorrecta, mientras que el segundo se refiere a la proporción de la muestra clasificada incorrectamente. Una vez la regla de clasificación es construida, se hace necesario tener una medida de su desempeño. La probabilidad total de clasificación incorrecta es:

$$TPM = p_1 \int_{R_2} f_1(x)dx \; + \; p_2 \int_{R_1} f_2(x)dx \tag{2.24}$$

El valor más pequeño de esta cantidad, se obtiene al escoger las regiones R_1 y R_2 y es llamado proporción de error óptima: (OER)

$$OER = p_1 \int_{R_2} f_1(x)dx \; + \; p_2 \int_{R_1} f_2(x)dx \tag{2.25}$$

donde R_1 y R_2 son determinados teniendo en cuenta que los costos de clasificación incorrecta son iguales por tanto:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{p_2}{p_1} \qquad R_2 : \frac{f_1(x)}{f_2(x)} < \frac{p_2}{p_1} \tag{2.26}$$

El desempeño de la función de clasificación de la muestra puede ser evaluado en principio calculando la tasa de error actual (AER).

$$AER = p_1 \int_{\widehat{R}_2} f_1(x)dx \ + \ p_2 \int_{\widehat{R}_1} f_2(x)dx \tag{2.27}$$

donde $\widehat{R}_1$ y $\widehat{R}_2$ representan las regiones de clasificación determinadas por las muestras de tamaño n_1 y n_2, respectivamente. Por ejemplo, si la regla de clasificación 2.23 es empleada, las regiones $\widehat{R}_1$ y $\widehat{R}_2$ son definidas por los conjuntos del espacio de observaciones de X para las cuales se satisfacen las siguientes desigualdades:

$$\widehat{R}_1 : [(\overline{x}_1 - \overline{x}_2)' S_p^{-1} x_0 - \frac{1}{2}(\overline{x}_1 - \overline{x}_2)' S_p^{-1}(\overline{x}_1 - \overline{x}_2)] \geq ln(\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.28}$$

$$\widehat{R}_2 : [(\overline{x}_1 - \overline{x}_2)' S_p^{-1} x_0 - \frac{1}{2}(\overline{x}_1 - \overline{x}_2)' S_p^{-1}(\overline{x}_1 - \overline{x}_2)] < ln(\frac{c(1|2)}{c(2|1)})(\frac{P_2}{P_1}) \tag{2.29}$$

El AER indica como la regla de clasificación se desempeñará en futuras muestras. Generalmente al igual que la tasa de error óptimo no puede ser calculado porque depende de las funciones de densidad $f_1(x)$ y $f_2(x)$ desconocidas. Sharma (1998). Sin embargo, un estimador de la cantidad relacionada con la tasa del error actual puede ser calculado.

La medida del desempeño que no depende de las poblaciones y que puede ser calculada para cualquier proceso de clasificación, es la llamada tasa de error aparente ($APER$), la cual está definida como la fracción de las observaciones en la muestra de entrenamiento que están clasificadas incorrectamente por la regla de clasificación.

La tasa de error aparente puede ser fácilmente calculada por medio de la matriz de confusión. Para n_1 observaciones de la población π_1 y n_2 observaciones de la población π_2, la matriz de confusión tiene la siguiente forma:

		Pertinencia Actual		
		π_1	π_2	
Pertinencia Real	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

n_{1C} : numero de ítems de π_1 correctamente clasificados en π_1,

n_{1M} : numero de ítems de π_1 incorrectamente clasificados en π_2,

n_{2C} : numero de ítems de π_2 correctamente clasificados en π_2,

n_{2M} : numero de ítems de π_2 incorrectamente clasificados en π_1. La tasa de error aparente es:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}, \tag{2.30}$$

que es reconocida como la proporción de ítems en la muestra de entrenamiento incorrectamente clasificados Johnson y Wichern (1999).

El $APER$ se aplica intuitivamente y es fácil de calcular. Desafortunadamente tiende a desestimar el AER, esto ocurre debido a que los datos usados para construir la función de clasificación son también usados para evaluarla y el problema no desaparece al menos que el tamaño de las muestras n_1 y n_1 sean muy grandes.

La tasa de error estimada puede ser construida mejor que la tasa de error aparente $APER$, y no necesita de suposiciones sobre distribución. El procedimiento requiere de dividir la muestra en muestras de entrenamiento, las cuales se usan para construir la regla de clasificación. A pesar que este método supera el problema del sesgo por no usar los mismos datos para construir y juzgar la función de clasificación, posee dos grandes defectos:

En primer lugar requiere de muestras grandes, y en segundo lugar la función evaluada no es la función de interés y finalmente todos los datos deben de ser utilizados para construir la función de clasificación para no perder información valiosa.

El procedimiento desarrollado por Lachenbruch llamado *"holdout"* a mostrado funcionar de mejor manera y consiste en los siguientes pasos:

1) Comienza con un grupo de observaciones de la población π_1. Omite una observación de esa población y desarrolla una función de clasificación con las observaciones restantes $n_1 - 1$, y n_2. 2) Clasifica la observación *"holdout"*, usando la función construida en el primer paso. 3) Repite los pasos 1 y 2 hasta lograr clasificar todas las observaciones de la población. Sea n_{1M}^H el número de las observaciones clasificadas incorrectamente en este grupo. 4) Repetir los pasos del 1 al 3 para la población π_2. Sea n_{2M}^H el número de las observaciones clasificadas incorrectamente en este grupo. Las estimaciones de las probabilidades condicionadas de clasificación incorrecta son:

$$\widehat{P}(2|1) = \frac{n_{1M}^H}{n_1} \qquad \widehat{P}(1|2) = \frac{n_{2M}^H}{n_2} \tag{2.31}$$

El valor esperado estimado para la tasa del error actual es:

$$\widehat{E}(AER) = \frac{n_{1M}^H + n_{2M}^H}{n_1 + n_2} \tag{2.32}$$

Capítulo 3

Propuesta de una nueva regla de clasificación

3.1. Propuesta inicial

En términos generales la propuesta consiste en elaborar una nueva regla de clasificación, la cual esta inspirada en la atracción que ejercen cuerpos fisicos, descrita en la ley de gravitación universal de Newton. Para formalizar la analogía entre la clasificación estadística y la ley de Newton, es preciso establecer un conjunto de condiciones requeridas, a manera de contexto.

Teniendo en cuenta que la ley de gravitación universal de Newton se refiere a la existencia de una fuerza de atracción entre dos cuerpos o partículas de materia, que se ejerce en dirección de la recta que los une y que es directamente proporcional a cada una de las masas e inversamente proporcional al cuadrado de la distancia que los separa, fuerza expresada por medio de:

$$F = \frac{m_1 \times m_2}{d^2}.$$

Recientemente algunos investigadores han realizado trabajos buscando semejanza con las fuerzas de atracción gravitacionales, utilizando otras tecnicas de clasificación Peng Yang Chen Abraham (2009).

La idea originaria de esta tesis, consiste en concebir a la ubicación de una unidad estadística como consecuencia de una mayor *atracción* hacia un cuerpo que hacia otro u otros. En este sentido, la regla de clasificación tendrá una estructura que de manera concreta responderá a la tarea clasificatoria como:

"Ubicar a la nueva unidad estadística en la población cuyo cuerpo representativo ejerza mayor fuerza de atracción sobre el cuerpo representativo de la citada unidad estadística"

Esta tesis experimentó con varias ideas a partir de la idea inicial que motivó la realización de ella. A continuación se expone lo pertinente a este proceso.

3.2. Estructura de la propuesta inicial

Teniendo en cuenta la idea escencial como eje, dentro del contexto y los elementos de la analogía, se requiere de la definición de un cuerpo, un cuerpo representativo de una población. Se propone que sea una elipse, un elipsoide o en su forma más general un hiperelipsoide, que corresponde al objeto geométrico asociado a una interpretación, de la misma índole, de la varianza generalizada. En efecto, de acuerdo a lo expresado por Johnson y Wichern (1998), la varianza generalizada tiene una interpretación en el espacio p- dimensional, y esta interpretación tiene que ver con el hiperelipsoide definido por la ecuación

$$(X-\overline{X})'S^{-1}(X-\overline{X})=c^2 \tag{3.1}$$

El volumen del hiperelipsoide está relacionado con $|S|$ de tal manera que:

$$Volumen\ de\ \{X:(X-\overline{X})'S^{-1}(X-\overline{X})\leq c^2\}=k_p\ |S|^{\frac{1}{2}}\ c^p.$$

Siendo $k_p=\frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})}$, donde: $\Gamma(z)$ es la función gamma evaluada en z.

Que dicho en palabras corrientes: "*El cuadrado del volumen del elipsoide es el producto de una constante y de la varianza generalizada de la muestra* " Este volumen igualmente puede formularse en términos de matrices de covarianzas poblacionales sin condición alguna.

¿Cuál cuerpo representa a la nueva unidad?. A diferencia de la población, la nueva unidad estadística es una entidad singular que merece tratarse como una "*partícula*"de tal manera que su masa puede asumirse igual a la unidad, con el fin de simplificar la analogía. Definidos los cuerpos dentro de la analogía, resta establecer un "*espacio*" donde se generan las fuerzas de atracción. Podría pensarse en un espacio definido por la dimensionalidad de la información, esto es, el número de variables utilizadas para la clasificación. Sin embargo en la práctica podrían presentarse dificultades colaterales que perturbarían el proceso de clasificación y podría tornarse engorrosa su utilización. Por ello se propone manejar un espacio que sea de menor dimensión del original, por medio de los componentes principales. El plano o espacio definidos por los dos o tres componentes principales ofrecen ambientes de fácil representación y de aprehensión de rasgos propios de una clasificación particular.

Las distancias, entonces serán determinadas entre la información de la nueva unidad entendida como un punto en el plano o en el espacio definido por las dos o tres, primeras componentes principales que se denominará núcleo de la unidad estadística, y el centro o centroide de la elipse o del elipsoide en consideración, que se denominará núcleo de la población en el nuevo sistema de coordenadas.

Sea $\pi_1, \pi_2, ..., \pi_g$ la colección de g grupos o poblaciones que definen la finalidad de discriminación. La i-ésima población π_i $i = 1, 2, ..., g$ se caracteriza por tener como centroide a μ_i y matriz de covarianzas a $\sum_i$. Como el determinante de $|\sum_i|$ es equivalente a $\prod_j^p \lambda_{ij}$ para $j = 1, 2, ..., p$. Siendo p el número de variables, λ_{ij} el j-ésimo valor propio de la matriz $\sum_i$, la masa del cuerpo que se circunscribe a la población i es asociada a la probabilidad a priori P_i, el producto del volumen y la probabilidad a priori da la idea de densidad física, y por tanto la masa del cuerpo en mención puede admitirse como:

$$m_i = k_p \sqrt{|\sum\nolimits_i|} P_i \tag{3.2}$$

$$m_i = \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} \sqrt{\prod_{j=1}^{p} \lambda_{ij}} P_i \tag{3.3}$$

.

Cuando se tiene una nueva unidad estadística para clasificar, se puede asumir que su masa es igual a uno como se había mencionado anteriormente, para no atiborrar la regla de clasificación final, con términos que no participan en el proceso clasificatorio. Prosiguiendo con la analogía, la constante G o constante de proporcionalidad no tiene relevancia alguna y se puede precindir de ella. Así entonces la fuerza de atracción entre la unidad estadística y la población i dentro del sistema de componentes principales para la población j que se denominará sistema j , $j = 1, 2, ..., g$ corresponde a:

$$F_{ij} = \frac{\frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} \sqrt{\prod_{j=1}^{p} \lambda_{ij}} P_i}{d_{ij}^2}, \tag{3.4}$$

siendo d_{ij}^2 el cuadrado de la distancia euclidiana entre el núcleo de la nueva unidad estadística y el núcleo de la i-esima población π_i, dentro del sistema j. Se prefiere la distancia euclidiana a cualquier otra distancia estadística con el objeto de que la analogía con la ley de atracción sea más precisa. La regla de clasificación finalmente será:

“Ubíquese a la nueva unidad estadística en la población k si: $F_k = \max\{F_1,\ F_2, ..., F_g\}$”

En el caso de homoscedasticidad sólo es necesario considerar un único sistema de componentes principales; siendo así, las masas serán las mismas y la fuerza de atracción se simplifica a la siguiente razón:

$$F_i = \frac{P_i}{d_i^2} \tag{3.5}$$

En la práctica se pueden presentar valores propios de magnitudes muy pequeñas cercanas a cero, originando que la varianza generalizada alcance un valor demasiado pequeño, en cuyo caso el cálculo de la fuerza de atracción sería cercana a cero, por lo cual se hace necesario trabajar con los valores propios que tengan mayor magnitud de tal manera que las fuerzas sean manifiestas, por lo tanto es conveniente llevar a cabo la discriminación con un número de componentes sensiblemente menor que el número de variables. La propuesta entonces estudiará el comportamiento de la regla al tenor del titulo propuesto para este trabajo de tesis: dos o tres componentes sobresalientes.

3.2.1. Simulación de la nueva regla de clasificación para dos poblaciones

Empleando el programa IML[2] del paquete estadístico SAS,® con el objeto de comparar la eficacia en la clasificación entre la regla discriminante lineal con la nueva regla se realizó el trabajo de simulación bajo las siguientes condiciones:

1. La simulación inicia con 6 variables, probabilidades a priori 0.2 para la población 1 y 0.8 para la población 2. Se generó aleatoriamente el vector μ_1 con el objeto de generar el vector μ_2 controlando la distancia entre los dos vectores, teniendo en cuenta la relación con los cosenos direccionales y distancia entre dos vectores, de acuerdo con los teoremas expresados por Basilevsky (1983). Ver Apendice B.

2. Se genera la matriz de covarianzas $\sum$, para cumplir con el requisito de homoscedasticidad, vigilando que los dos primeros componentes principales retienen al menos el 75 por ciento de la variabilidad.

3. Se generan 1000 unidades estadísticas a partir de la población 1 y 1000 unidades estadísticas a partir de de la población 2, aplicando las dos reglas con las mismas unidades tanto para la población 1 como para la población 2.

4. Se repiten los pasos anteriores 1000 veces.

5. Se repiten los pasos anteriores para las distancias entre los centroides 4, 6, 8, 10.

6. Se repiten los pasos anteriores, para probabilidades a priori 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 población 1 y 0.7, 0.6, 0.5, 0.4, y 0.3 para la población 2.

7. Se repiten los pasos anteriores para 7, 8, 9, y, 10 variables.

8. Se repiten los pasos anteriores para 50, 25, porciento de variabilidad

La simulación muestra que el número de unidades estadísticas correctamente clasificadas por la regla inicial de clasificación es demasiado baja frente a la regla basada en el puntaje lineal discriminante. Esto sugiere varios interrogantes: Como lograr que la probabilidad a priori tenga nenos peso en las fuerzas de atracción? ya que con probabilidades a priori altas y cuando la diferencia entre las distancias de la nueva unidad y los centroides de cada población son pequeñas la fuerza de atracción es mayor. Las respuestas se buscaron en las modificaciones realizadas.

[2]SAS Institute Inc., SAS/IML Software: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc., 1989. 501 p.

3.2.2. Modificación de la regla inicial

Dado que el caso de homoscedasticidad es uno de los supuestos explícitamente asumido sólo se consideró un único sistema de componentes principales, como se había mencionado anteriormente y por ello para simplificar la fuerza de atracción en:

$$F_i = \frac{P_i}{d_i^2}$$

En consecuencia la probabilidad a priori se hace determinante, sobre todo cuando existe una gran diferencia de éstas en las dos poblaciones, por ejemplo: 0.2 para una población y 0.8 para la otra; la fuerza de atracción será mayor para la población que posea mayor probabilidad dejando casi sin peso a la distancia.

Por lo tanto para continuar explorando mejores posibilidades se realizo algunas transformaciones a las probabilidades a priori para utilizando las funciones seno, coseno, logaritmo, raíz cuadrada y exponencial. Los mejores resultados más no satisfactorios, se lograron cuando se utilizó únicamente el valor absoluto del logaritmo natural del cociente de las probabilidades a priori de las poblaciones.

Persistiendo en el propósito de búsqueda y formulación de una mejor regla de clasificación, se consideró pertinente examinar dentro de la simulación, el comportamiento de los escalares que representan el cuadrado de las distancias entre la nueva unidad y el centroide de cada una de las poblaciones, buscando argumentos que simplificaran al máximo la estructura de la regla, transformado y concentrando el valor de estos escalares en dos valores únicos, emulando de alguna manera la simplicidad de la expresión de la ley de gravitación universal de Newton, inspiradora de esta tesis.

Asignar un valor constante al escalar de menor magnitud y a su vez otro valor constante al escalar de mayor magnitud, fue la meta inmediata que se trazó en el mejoramiento de la regla y como resultado de la simulación, se pudo establecer que al escalar con menor valor se le debe hacer corresponder el valor de 0,5 y al escalar de mayor magnitud el valor 1; de esta manera se logró mejorar ostensiblemente la clasificación de unidades estadísticas.

Finalmente, anuando las dos últimas modificaciones comentadas, es decir integrando el valor absoluto del logaritmo natural del cociente de las probabilidades a priori de las dos poblaciones y la sustitución de los valores de d_1^2 y d_2^2, por los valores de 0.5 o 1 según sea el caso de acuerdo a lo expuesto anteriormente, se genera una nueva propuesta la cual incluye las siguientes fuerzas de atracción:

$$F_1^* = \frac{|ln(\frac{a\ \ priori_1}{a\ \ priori_2})|}{fr_1} \qquad y \qquad F_2^* = \frac{|ln(\frac{a\ \ priori_2}{a\ \ priori_1})|}{fr_2} \tag{3.6}$$

Siendo:

$$fr_1 = 0.5 \qquad y \qquad fr_2 = 1 \quad si : fr_1 = min\{d_1^2\ ,\ d_2^2\} \tag{3.7}$$

$$fr_1 = 1 \qquad y \qquad fr_2 = 0.5 \quad si : fr_1 = max\{d_1^2\ ,\ d_2^2\}. \tag{3.8}$$

La Regla correspondiente a la anterior propuesta es:

" Ubíquese a la nueva unidad estadística en la población k si: $F_k^* = \max\{F_1^*,\ F_2^*\} \quad k = 1, 2$ ".

Si las probabilidades a priori son iguales para las dos poblaciones, el logaritmo natural de las dos cantidades es cero, razón por la cual no se debe de tener en cuenta y las fuerzas a comparar reducen a:

$$F_1^* = \frac{\frac{a\ priori_1}{a\ priori_2}}{fr_1} \qquad y \qquad F_2^* = \frac{\frac{a\ priori_2}{a\ priori_1}}{fr_2} \tag{3.9}$$

$$F_1^* = \frac{1}{fr_1} \qquad y \qquad F_2^* = \frac{1}{fr_2} \tag{3.10}$$

Capítulo 4

Exploración del desempeño de la nueva regla de clasificación

Con la finalidad de identificar el desempeño de la nueva regla de clasificación frente a la regla de puntajes discriminantes, se generaron 1000 unidades estadísticas para cada una de las dos poblaciones y una matriz de covarianzas $\sum$, el centroide de la población 1, μ_1 y el centroide de la población 2, μ_2 , cada unidad fue clasificada por las dos reglas, considerando los factores de simulación: distancia entre centroides, probabilidad a priori, y número de variables.

Cada combinación de factores se repitió 1000 veces, tanto para dos y tres componentes admitidos, es decir verificando que los dos primeros componentes principales retuvieran al menos el 75 por ciento de la variabilidad y de igual manera las reglas se cotejaron prescindiendo de la utilización de otro sistema de coordenadas, los componentes principales. Cada repetición implicó además de la generación de nuevas matrices de covarianzas y de nuevos centroides, el control de cada uno de los niveles de los factores de simulación así: 6, 7, 8, 9, 10 variables, probabilidades a priori 0.2, 0.4, 0.5, 0.6, 0.8, al igual que para las distancias entre centroides 2, 4, 6, 8, 10.

Debido a que los componentes principales reducen el espacio p dimensional en detrimento de la información, se hace necesario indagar como afecta el desempeño de la regla propuesta en su función clasificadora cuando ésta emplea los dos o tres componentes principales. Por lo tanto se hace ineludible cotejar las dos reglas en escenarios donde el espacio sea igual para ellas: empleando un nuevo sistema de coordenadas, los componentes principales y sin reducir el espacio p dimensional

Con el propósito de cumplir con lo anterior, la comparación de las dos reglas de discriminación, se ejecuto teniendo en cuenta las siguientes pautas:

1. La regla propuesta basada en dos y luego con tres componentes principales admitidos, frente a la regla de puntajes lineales discriminantes.

2. Las dos reglas empleando los componentes principales: La regla propuesta basada en con

dos y tres componentes principales admitidos, frente a la regla de puntajes lineales discriminantes utilizando ésta última, los componentes principales.

3. Incrementar las posibilidades de comparación, y generando las matrices de covarianzas subordinadas a que los dos componentes principales retuviesen al menos 75, 50, y 25 por ciento de variabilidad, y adicionalmente teniendo en cuenta las probabilidades a priori 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, pero confrontando las dos reglas, sin reducir el espacio p-dimensional, es decir sin utilizar los componentes principales.

Este ejercicio de simulación generó un número de casos específicos, que se han denominado unidades estadísticas, que realmente deben entenderse como la simulación de la información de unidades estadísticas propiamente dichas. El gran total de estas unidades fue de 2050.000.000, discriminadas como a continuación se indica.

Para la pauta 1: Dos componentes principales admitidos, cinco distancias 2, 4, 6, 8 10, cinco número de variables 6, 7, 8, 9 10, cinco probabilidades a priori 0.2, 0.4, 0.5, 0.6, 0.8, y 1000 repeticiones, producen 125 tratamientos para un total de 125.000.000 de unidades generadas para la población 1 y de 125.000.000 unidades generadas para la población 2. Total de unidades generadas para la pauta 1 admitiendo dos componentes principales: 250.000.000.

Tres componentes principales admitidos, cinco número de variables 6, 7, 8, 9, 10 cinco distancias 2, 4, 6, 8, 10, cinco probabilidades a priori 0.2, 0.4, 0.5, 0.6, 0.8, y 1000 repeticiones, producen 125 tratamientos para un total de 125.000.000 de unidades generadas para la población 1 y de 125.000.000 unidades generadas para la población 2. Total de unidades generadas para la pauta 1 admitiendo tres componentes principales: 250.000.000. Total de unidades generadas para la pauta 1 : 500.000.000.

Para la pauta 2: Se mantienen los mismos niveles y condiciones de la pauta 1, por tanto se generó el mismo número de unidades que para la pauta 1. Total de unidades generadas para la pauta 2 : 500.000.000.

Para la pauta 3: El factor adicional de mínima retención de variabilidad de los componentes principales, con niveles 75, 50 y 25 por ciento, los nuevos niveles de la probabilidad a priori 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, las cinco distancias 2, 4, 6, 8, 10, ya establecidas y los cinco números de variables 6, 7, 8, 9, 10, y 1000 repeticiones, producen 525 tratamientos para un total de 525.000.000 de unidades generadas para la población 1 y de 525.000.000 unidades generadas para la población 2. Total de unidades generadas para la pauta 3 : 1050.000.000.

Total de unidades estadísticas generadas por la simulación: 2050.000.000.

Los programas de simulación se presentan en el Apéndice A.

4.1. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la primera pauta.

Para realizar el análisis de los resultados obtenidos por la simulación se determina la diferencia de la proporción unidades correctamente clasificadas por la regla de puntajes lineales discriminantes y la regla propuesta, de tal manera que cuando la regla de puntajes lineales discriminantes tiene un mejor desempeño clasificatorio, el signo de la diferencia de la proporción de unidades correctamente clasificadas es positivo, en caso de presentarce un empate la diferencia de proporciones de unidades sera cero, de ésta manera se podrá también apreciar cuando el desempeño de las dos reglas que se comparan tiende a ser igual.

Conforme a la figura 4.1, la cual muestra el desempeño de la regla propuesta basada en dos componentes frente a la regla de puntajes lineales discriminantes, considerando el efecto de los factores número de variables y probabilidad a priori en la clasificación correcta de unidades estadísticas, se puede distinguir la conformación de una elevación central, que se encuentra situada entre las probabilidades a priori 0.3 y 0.6, cuya cima se localiza para probabilidad a priori de 0.5 y para cualquier número de variables, donde se encuentra la mayor diferencia en la proporción de unidades correctamente clasificadas por las dos reglas, y que disminuye ligeramente su altitud a medida que disminuye el número de variables.

De la misma manera se puede apreciar una elevación que encuentra su máximo a una probabilidad a priori de 0.2 y a una distancia de diez, donde la diferencia de proporción de unidades correctamente clasificadas por las dos reglas es mayor, cayendo precipitadamente hacia probabilidad a priori de 0.2 y seis variables, encontrándose ahí la mínima diferencia de unidades bien clasificadas por las dos reglas. Otra elevación, pero de menor altura que las anteriores tiene lugar a partir de la probabilidad a priori de 0.6 hasta la probabilidad a priori de 0.8.

Debido a que la diferencia de las proporciones de unidades bien clasificadas es siempre positiva se puede afirmar que la regla de puntajes lineales discriminantes tiene categóricamente un mejor desempeño que la regla propuesta cuando se la compara con ésta basada en dos componentes principales.

Cotejando las dos reglas, advirtiendo el comportamiento de éstas bajo la interacción de los factores: distancia y número de variables, se descubre una superficie con una pendiente que desciende a medida que la distancia aumenta, es decir que la diferencia de las proporciones de unidades correctamente clasificadas se hace cada vez más pequeña a medida que aumenta la distancia a partir de siete hasta llegar a diez, se alcanza a observar la conformación de un plano, en donde la diferencia de proporciones de unidades clasificadas por las dos reglas alcanza su menor valor, haciendo en este caso equivalentes a las reglas en comparación.

Sin embargo, al igual que se advertía en el caso anterior la regla de puntajes lineales discriminante tiene mejor desempeño comparada con la regla propuesta. Ver figura 4.2.

Acorde con la figura 4.3, donde se representa la interacción de los factores distancia y probabilidad a priori sobre la diferencia de proporciones de unidades correctamente clasificadas, se puede

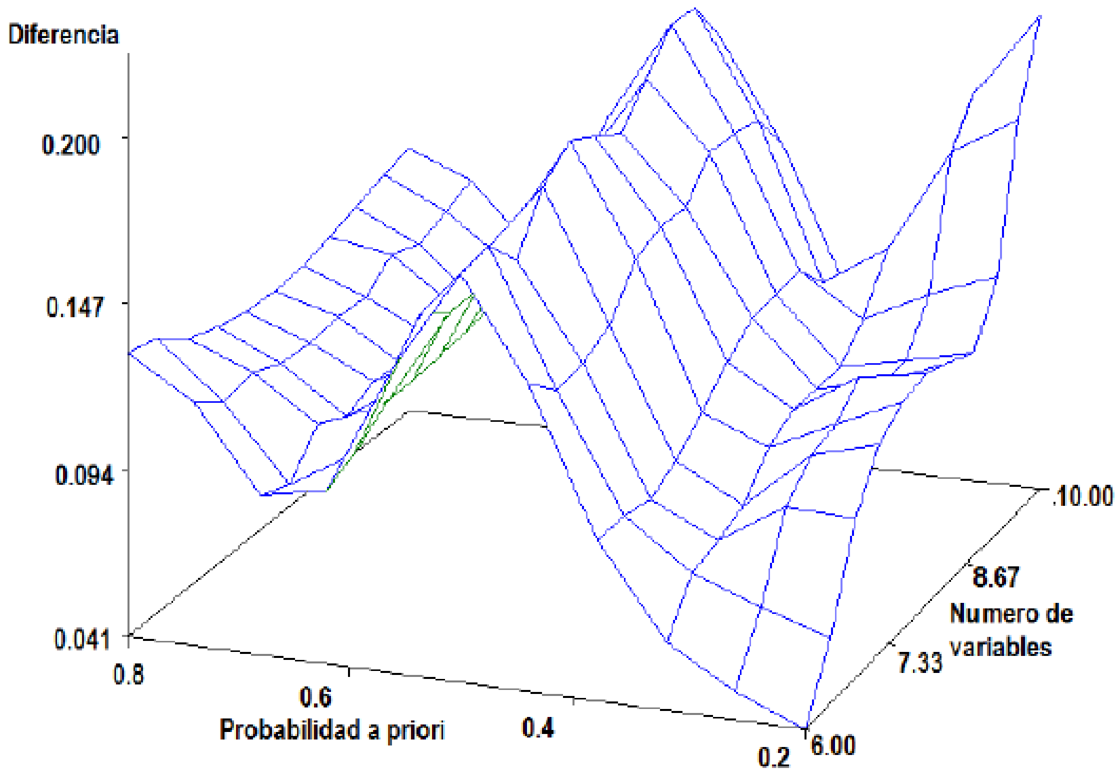

Figura 4.1. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la probabilidad a priori

apreciar una superficie irregular, que presenta una elevación hacia el frente, que coincide con distancias pequeñas, presentando puntos altos en probabilidades a priori de 0.5 y 0.8, siendo el de 0.5 el superior, que corresponde a la mayor diferencia en la proporción de unidades estadísticas correctamente clasificadas por las dos reglas.

Además se presentan algunos puntos más bajos si se le compara con los anteriores, que se encuentran en probabilidades a priori de 0.2 siendo el menor de ellos, y el otro que corresponde a la probabilidad a priori de 0.6.

Conforme con el aumento de la distancia hay una disminución en la diferencia de las proporciones de las unidades clasificadas correctamente; en esta región se encuentra la mínima diferencia entre las dos reglas, teniendo las dos reglas a ser iguales en su desempeño, no obstante éste siempre es favorable a la regla de puntajes lineales discriminantes.

La figura 4.4 representa la diferencia de la proporción de unidades correctamente clasificadas por la regla propuesta basada en tres componentes principales y la regla de puntajes lineales discriminantes, cuando se analiza tomando los factores número de variables y probabilidad a priori. Se puede apreciar una superficie que se puede dividir en dos partes: la primera con una concavidad definida entre las probabilidades a priori alrededor de 0.5 y 0.8 cuyo fondo se encuentra en la probabilidad a priori de 0,6, el cual representa la mínima diferencia en la

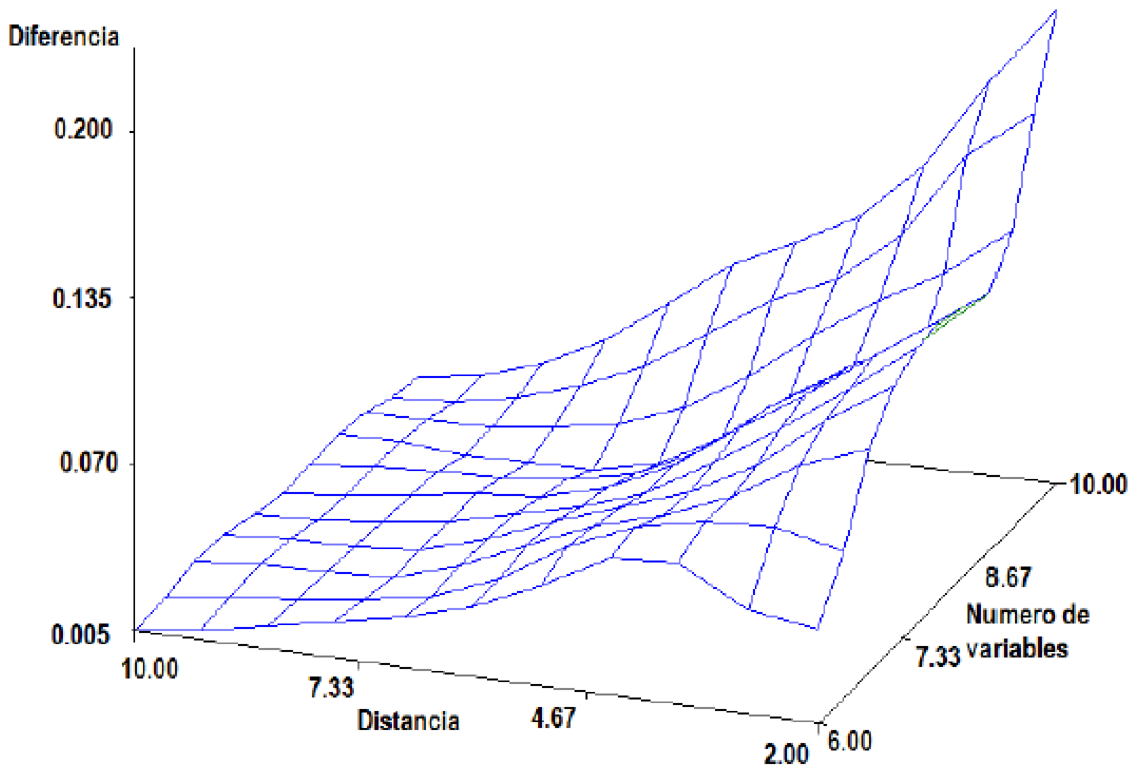

Figura 4.2. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides

proporción de unidades correctamente clasificadas por las dos reglas.

La segunda corresponde a una convexidad cuya parte alta se puede localizar alrededor de la probabilidad 0.4 y que aumenta ligeramente a medida que el número de variables crece, lo que significa que la diferencia máxima entre unidades correctamente clasificadas por las dos reglas se encuentra en esta región.

Al igual que en los casos anteriores la diferencia es positiva lo que significa que en toda esta región la regla de puntajes lineales discriminantes tiene un mejor desempeño frente a la regla propuesta .ver figura 4.4

Si se compara la figura 4.1 con la figura 4.4 se observa que para la regla propuesta basada en dos componentes principales la diferencia mínima se da en probabilidad a priori de 0.2 y seis variables, en tanto que para tres componentes se da en probabilidad a priori de 0.6, siendo menos la diferencia de proporciones de unidades en la regla propuesta basada en dos componentes.

Cuando la comparación se hace en referencia a la interacción de los factores número de variables y distancia, se presenta una situación afín a la encontrada con dos componentes principales; se tiene una pendiente cuya parte más elevada se encuentra cuando hay un mayor número de variables y a una probabilidad a priori de 0.2, encontrándose ahí la mayor diferencia en la

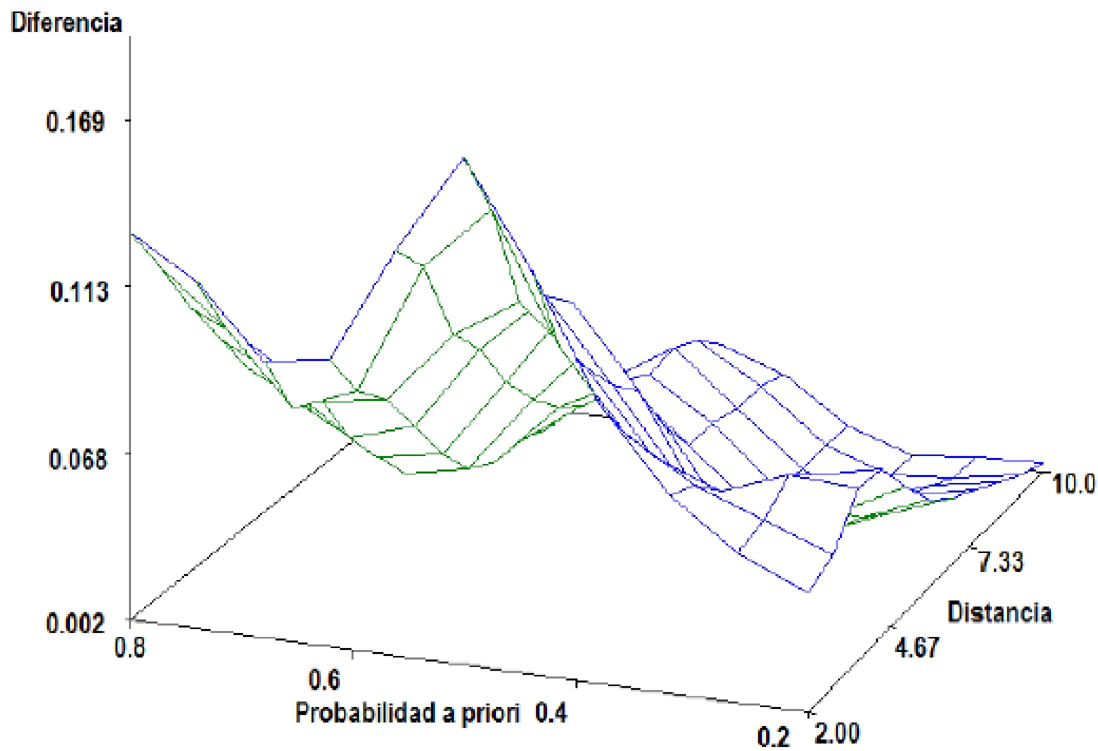

Figura 4.3. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides

proporción de unidades correctamente clasificada por las dos reglas.

Tal diferencia disminuye al aumentar la distancia y se hace ligeramente más rápida cuando el número de variables disminuye, hasta alcanzar la menor diferencia entre las dos reglas en la distancia máxima y en el mayor número de variables, permaneciendo siempre a favor de la regla de puntajes lineal discriminantes, tal como se puede observar en la figura 4.5.

Contrastando esta situación con la que se presenta cuando la regla propuesta emplea dos componentes, se tiene que el comportamiento es similar; sin embargo, cuando la regla propuesta utiliza dos componentes la diferencia mínima de proporciones de unidades correctamente clasificadas es menor que cuando ésta utiliza tres componentes, y se da para ambos a una distancia de diez con seis variables; en otras palabras el descenso de la pendiente para dos componentes se hace más rápido que para tres. Ver figuras 4.2 y figura 4.5

Examinando el comportamiento de las dos reglas, tomando en observación los factores distancia y probabilidad a priori, cuando la regla propuesta emplea tres componentes principales, como lo describe la figura 4.6, se revela una joroba, que se encuentra demarcada entre las probabilidades a priori de 0.2 y 0.5, entre las distancias dos y diez, la cual tiende a disminuir a medida que la distancia aumenta, por lo tanto la diferencia máxima en la proporción de unidades correctamente clasificadas por las reglas que se comparan se encuentra a una probabilidad de 0.3 y 0.4, a una

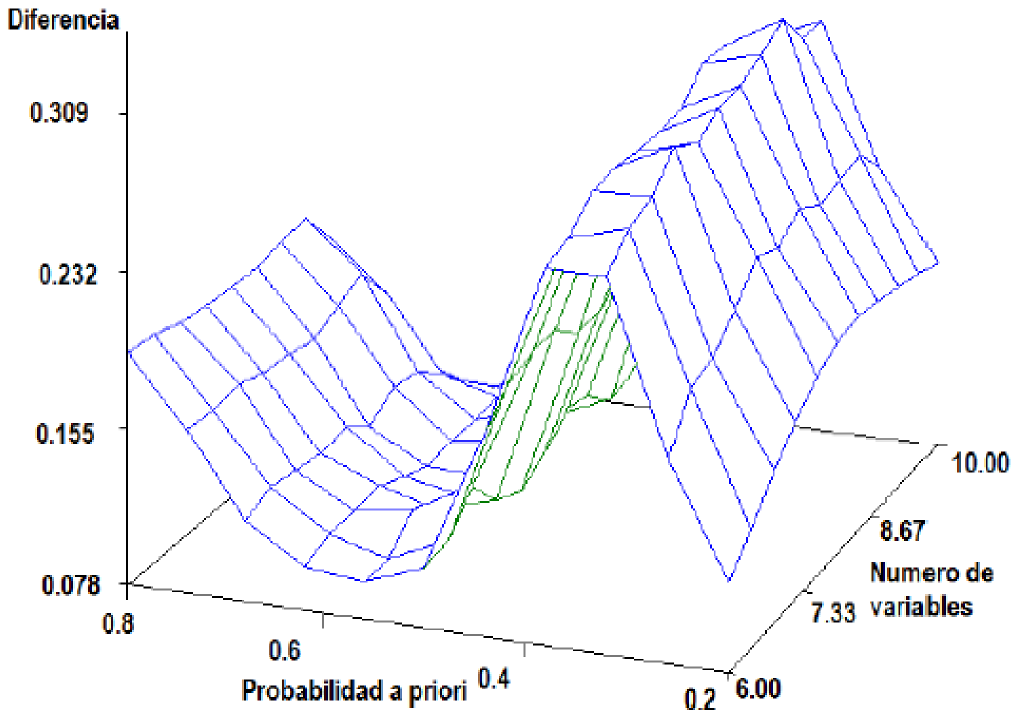

Figura 4.4. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y número de variables

distancia de dos.

También se observa una ligera pendiente que aminora cuando la distancia es de diez, encontrándose ahí la mínima diferencia entre la proporción de unidades clasificadas bien por las dos reglas. La diferencia de proporciones de unidades, siempre es favorable a la regla de puntajes lineales discriminante. Ver figura 4.6

Cotejando estos resultados con los obtenidos para la regla propuesta cuando ésta utilizaba dos componentes principales, se verifica que la diferencia mínima se da cuando la regla propuesta emplea dos componentes principales, y en general si se comparan las dos figuras se aprecia un mejor desempeño de la regla propuesta basada en dos componentes principales. Ver figuras 4.3 y 4.6

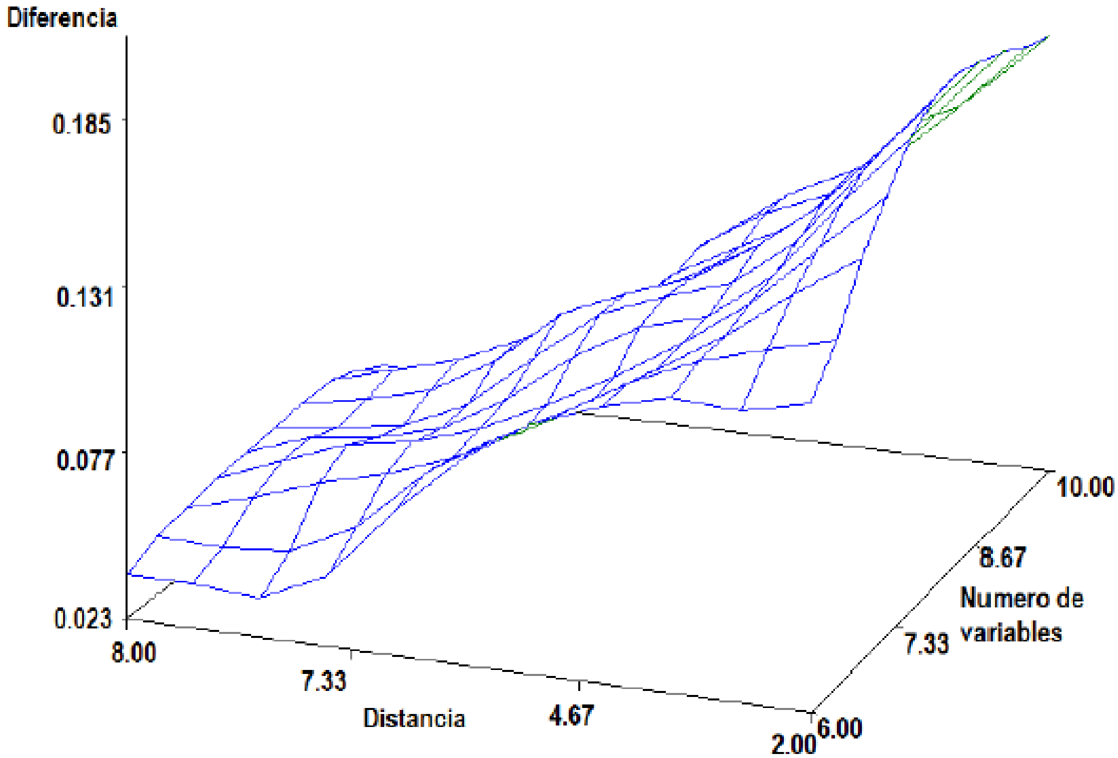

Figura 4.5. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides

4.2. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la segunda pauta.

La segunda pauta orienta la comparación del desempeño de la regla propuesta basada en dos y tres componentes principales admitidos, frente a la regla de puntajes lineales discriminantes utilizando un nuevo sistema de coordenadas, los componentes principales. Primero se analizara los resultados para las dos componentes principales.

De conformidad a lo encontrado en la figura 4.7 donde se representa la proporción de unidades clasificadas correctamente por las dos reglas bajo los factores número de variables y probabilidad a priori, se encuentra una región irregular que se la podría dividir en dos subregiones: la primera delimitada entre las probabilidades a priori 0.2 y 0.6 y entre seis y diez variables, que por su forma se asemeja a una silla de montar, la parte más baja de esa silla está ubicada entre el número de variables siete y ocho, los puntos más altos de la ella se encuentran para seis variables y para diez , lugar donde se ubica la mínima diferencia de la proporción de unidades clasificadas correctamente por las dos reglas, siendo favorable ésta a la regla propuesta.

De igual manera encontramos una pendiente cuya parte alta se encuentra en la probabilidad a priori de 0.8 y el pie de ésta se encuentra a una probabilidad a priori de 0.6, lugar que corresponde

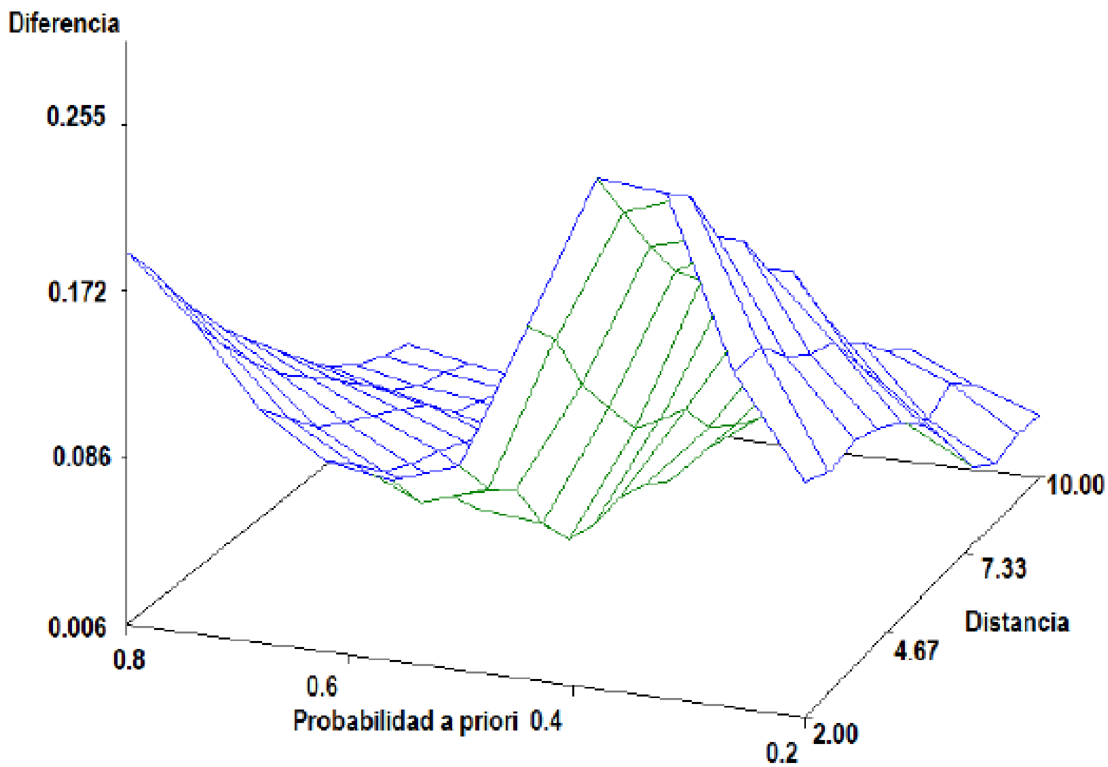

Figura 4.6. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides

a la diferencia máxima de la proporción de unidades bien clasificadas por las dos reglas.

Debido a que la diferencia entre la proporción de unidades clasificada por las dos reglas es negativa la regla propuesta tiene un mejor desempeño frente a la regla de puntajes lineales discriminantes bajo la pauta dos. Ver Figura 4.7

La figura 4.8 muestra la proporción de unidades estadísticas correctamente clasificadas por las dos reglas, teniendo en cuenta los factores distancia y el numero de variables. Se puede notar una superficie irregular ascendente que comienza con dos cuestas hacia distancias mayores, la primera parte de una distancia de dos, con un número de variables igual a seis, la segunda tiene origen en una distancia de tres con un número de variables igual a nueve, es aquí donde se presenta la máxima diferencia en la proporción de unidades clasificada por las dos reglas.

Hacia el centro de ésta se presenta una pequeña hondonada formada por dos pequeñas colinas, la cual se encuentra entre siete y ocho variables a distancias de seis, a partir de allí el ascenso se hace más pronunciado hacia distancias mayores donde la diferencia entre las dos reglas se hace menor, ratificando lo que se ha encontrado en otras oportunidades: a mayor distancia la diferencia entre las dos reglas se hace menor.

En toda esta región la proporción de unidades correctamente clasificadas es favorable a la regla propuesta. Ver figura 4.8.

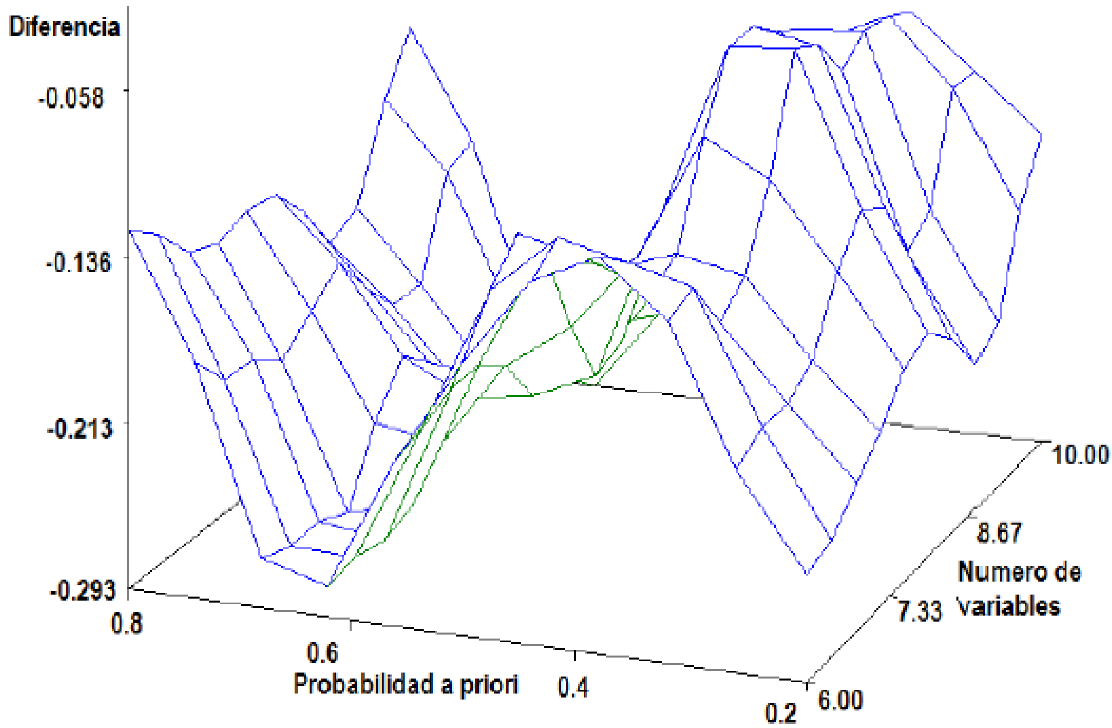

Figura 4.7. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la probabilidad a priori

Para indagar el comportamiento de las dos reglas incluyendo la interacción de los factores distancia y probabilidad a priori, se recurre a la figura 4.9 donde se puede observar una superficie irregular parecida a la que se presenta cuando se toman en cuenta los factores la probabilidad a priori y el número de variables: una fuerte depresión ubicada en la parte izquierda de la figura cuya sima corresponde a la probabilidad a priori de 0.6, punto en el cual se presenta la mayor diferencia en la proporción de unidades correctamente clasificadas por las dos reglas, siendo éstas favorables a la regla propuesta. La segunda corresponde a una superficie parecida a una silla de montar, que presenta su parte más baja entre las distancias cuatro y siete, las partes más altas de la silla corresponde a la probabilidad a priori de 0.4, distancia de dos y tres y distancias de siete a diez, donde la diferencia entre las dos reglas se hace más pequeña, pero sigue siendo favorable a la regla propuesta. Ver figura 4.9

Cuando la regla propuesta se vale de tres componentes principales admitidos y es cotejada con la regla de puntajes lineales discriminantes utilizando un nuevo sistema de coordenadas, los componentes principales, sigue teniendo un buen desempeño.

De acuerdo con la figura 4.10 que interpreta la interacción de los factores número de variables y la probabilidad a priori se percibe una superficie irregular con dos subregiones: una colina limitada por las probabilidades a priori de 0.2 a 0.6, cuya cima se encuentra a una probabilidad a priori de 0.4 y con un número de variables igual a ocho, en dicha cima se encuentra la diferencia mínima en la proporción de unidades correctamente clasificadas por las dos reglas, siendo esta

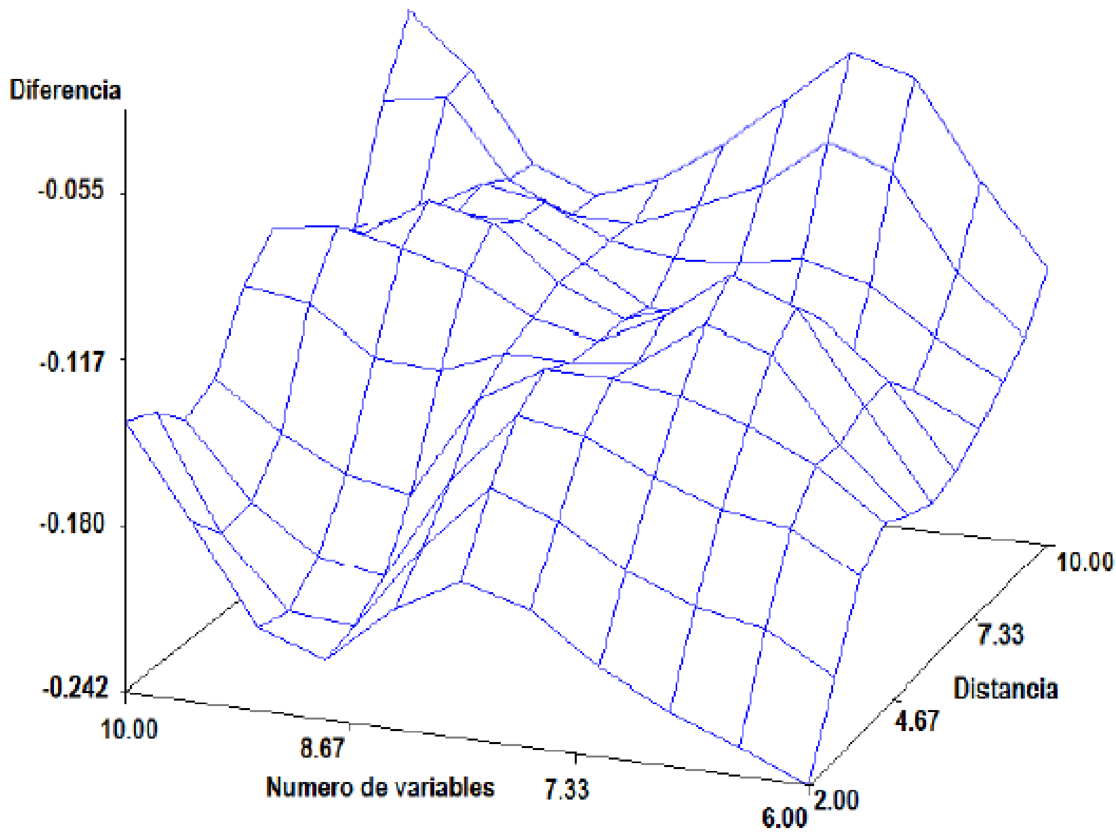

Figura 4.8. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides

favorable a la regla propuesta.

La otra subregión se encuentra circunscrita entre las probabilidades a priori de 0.6 y 0.8, entre seis y diez variables, formando con la falda de la colina una concavidad y en la parte más baja se encuentra la máxima diferencia en la proporción de unidades correctamente clasificadas, permaneciendo ésta a favor de la regla propuesta. Ver figura 4.10.

La figura 4.11 representa la diferencia de la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta los factores distancia y número de variables. La región de estudio presenta una superficie con una pendiente irregular, comienza cuesta empinada sobre todo entre el número de variables de ocho hasta diez, donde se encuentra la diferencia máxima entre las dos reglas permaneciendo ésta favorable a la regla propuesta; luego encuentra un pequeño repecho hacia la distancia seis, mientras que hacia la parte que es menor a ocho variables y a una distancia de seis se presenta una colina cuya cima se encuentra en la variable siete y la distancia seis donde se encuentra la mínima diferencia entre las dos reglas manteniéndose favorable a la regla propuesta . Ver figura 4.11.

Para estudiar la actuación de las dos reglas cuando se mira la interacción de los factores distancia y probabilidad a priori, se observa en la figura 4.12 una elevación similar a la presentada en el caso

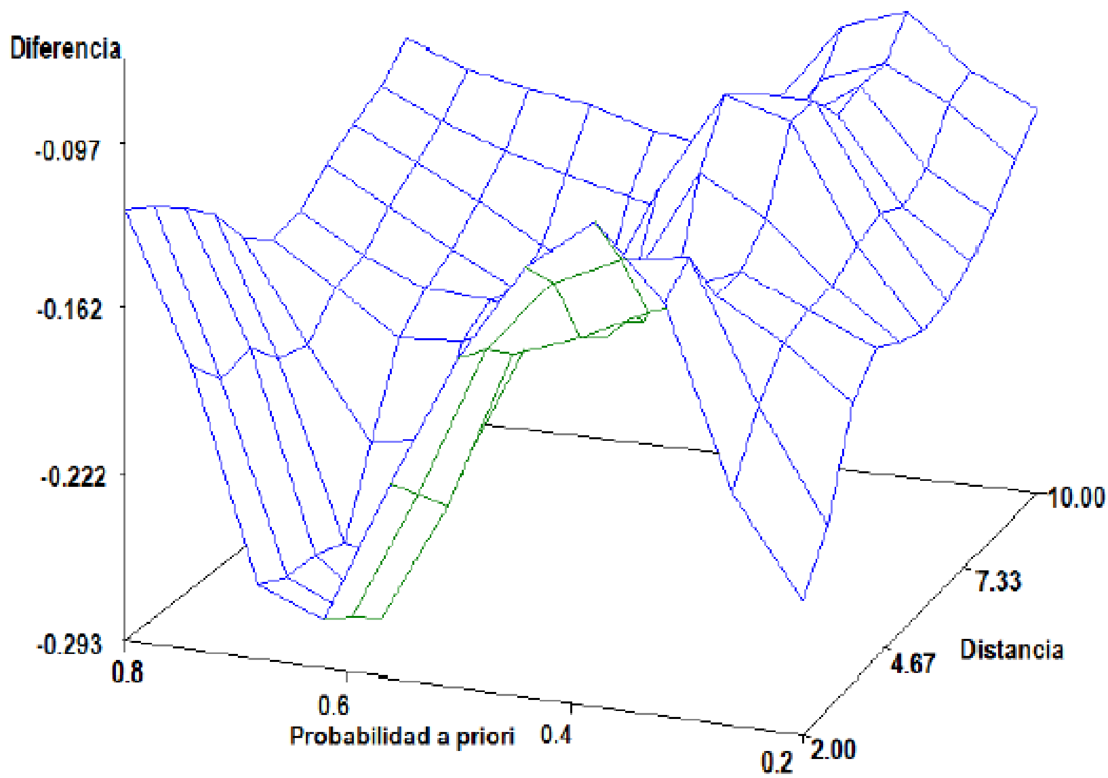

Figura 4.9. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides

anterior; una cuesta empinada que parte desde la probabilidad a priori de 0.6, donde se obtiene la máxima diferencia entre las dos reglas y que abarca la probabilidad a priori de 0.8 hasta la distancia de diez. La otra parte comienza con un picacho delimitado por las probabilidades a priori de 0.6 y 0.2 y una distancia de seis, en donde las diferencias entre las dos reglas se reducen, a medida que aumenta la distancia la diferencia entre las dos reglas se aminora, y continúa siendo favorable a la regla propuesta tal como se presentó en los casos anteriores. Ver figura 4.12.

Si se comparan las figuras 4.9 y 4.12 se distingue que la regla propuesta cuando emplea dos componentes principales tiene mejor desempeño que cuando lo hace con tres componentes admisibles.

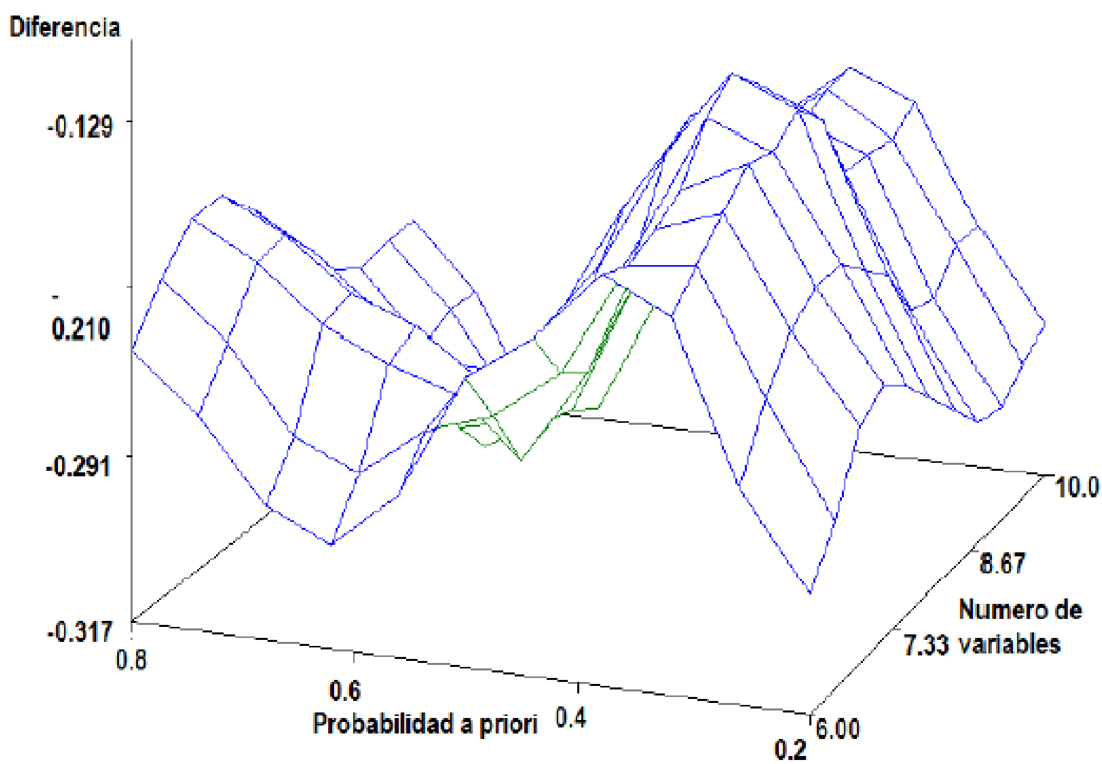

Figura 4.10. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y número de variables

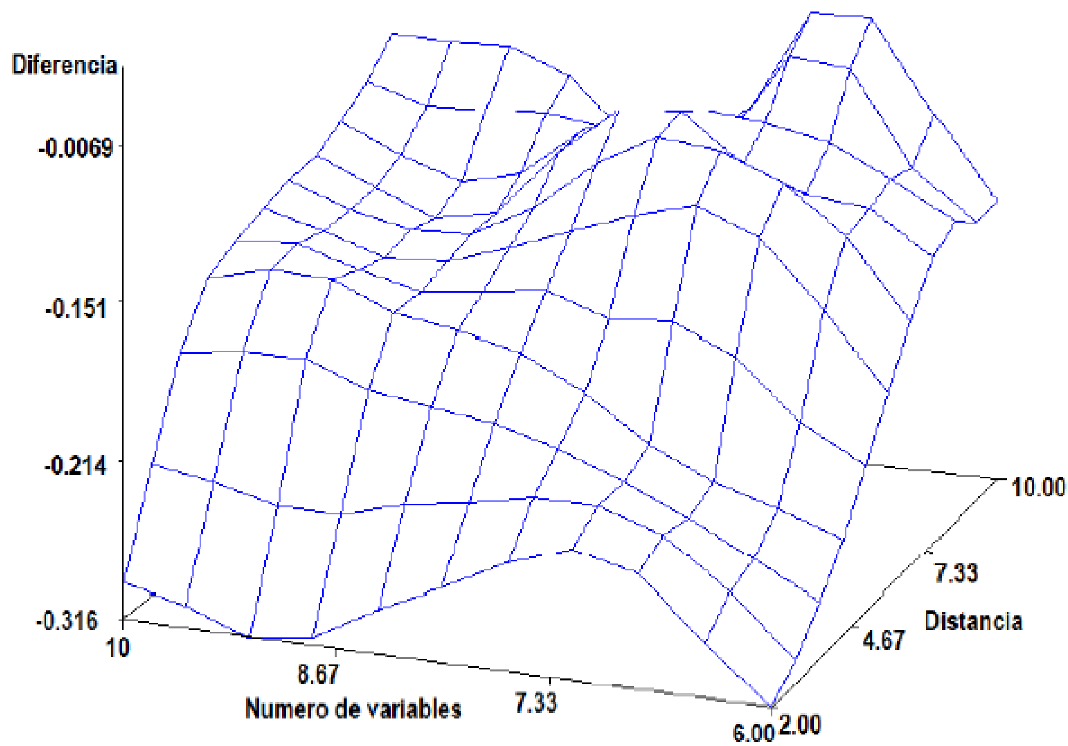

Figura 4.11. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el número de variables y la distancia entre centroides

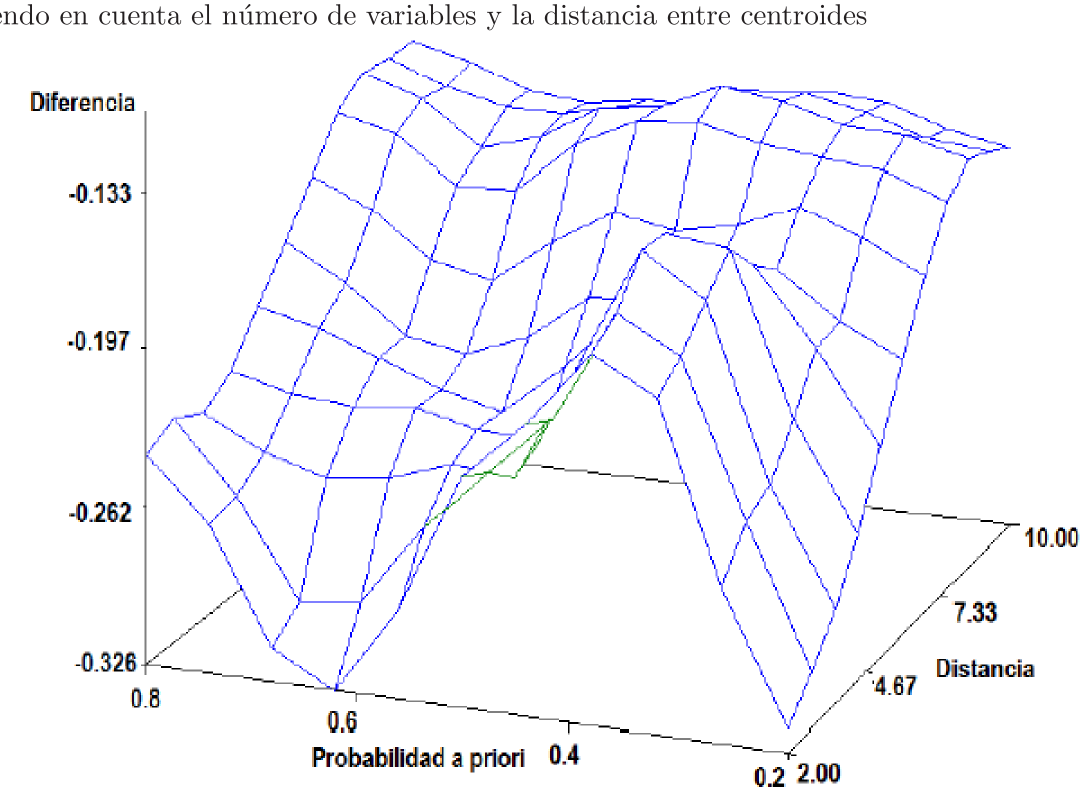

Figura 4.12. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides

4.3. Desempeño de la regla propuesta frente a la regla de puntajes lineales de acuerdo a la tercera pauta.

Considerando el desempeño de las dos reglas por medio de la diferencia de la proporción de número de unidades correctamente clasificadas, teniendo en cuenta los factores, distancia entre centroides y la probabilidad a priori de las poblaciones, se puede afirmar de acuerdo con la figura 4.13 la superficie que representa la diferencia de la proporción de unidades correctamente clasificadas para la región definida por los niveles de los factores en consideración, muestra claramente dos subregiones: Una de ellas está demarcada entre la distancia dos y alrededor de cinco, en dicha subregión la superficie es irregular cuya característica principal es la de presentar dos depresiones y dos elevaciones; las dos primeras tienen aproximadamente su punto más bajo en las probabilidades a priori de 0.3 y 0.6 con distancias de dos.

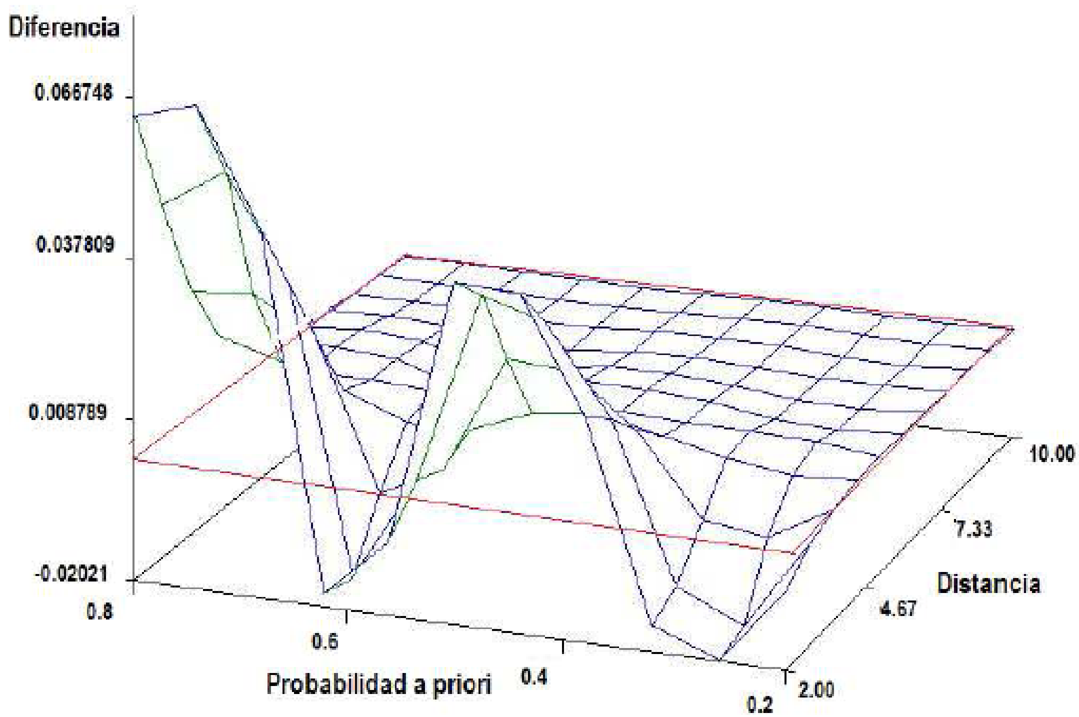

Figura 4.13. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y la distancia entre centroides

Esto indica un mejor desempeño para la regla propuesta frente a la regla de puntajes lineales discriminante, para las anteriores probabilidades a priori mencionadas y para distancias delimitadas por dos y cinco. Las dos elevaciones tienen sus cimas para probabilidades a priori alrededor de 0.5 y alrededor de 0.8. Lo cual indica que la regla de puntajes lineales discriminantes tiene un mejor desempeño frente al regla propuesta, para las anteriores probabilidades a priori y para las distancias anteriormente mencionadas; a partir de la distancia dos se pone en evidencia que la diferencia de la proporción de unidades correctamente clasificadas comienza a disminuir.

La segunda subregión que se encuentra localizada entre las distancias mayores a 4.67 y la distancia diez, presenta una planicie que descansa sobre el plano que representa al origen, hecho que revela que en esta subregión la cantidad de unidades correctamente clasificadas por las dos reglas es prácticamente igual.Ver figura 4.13.

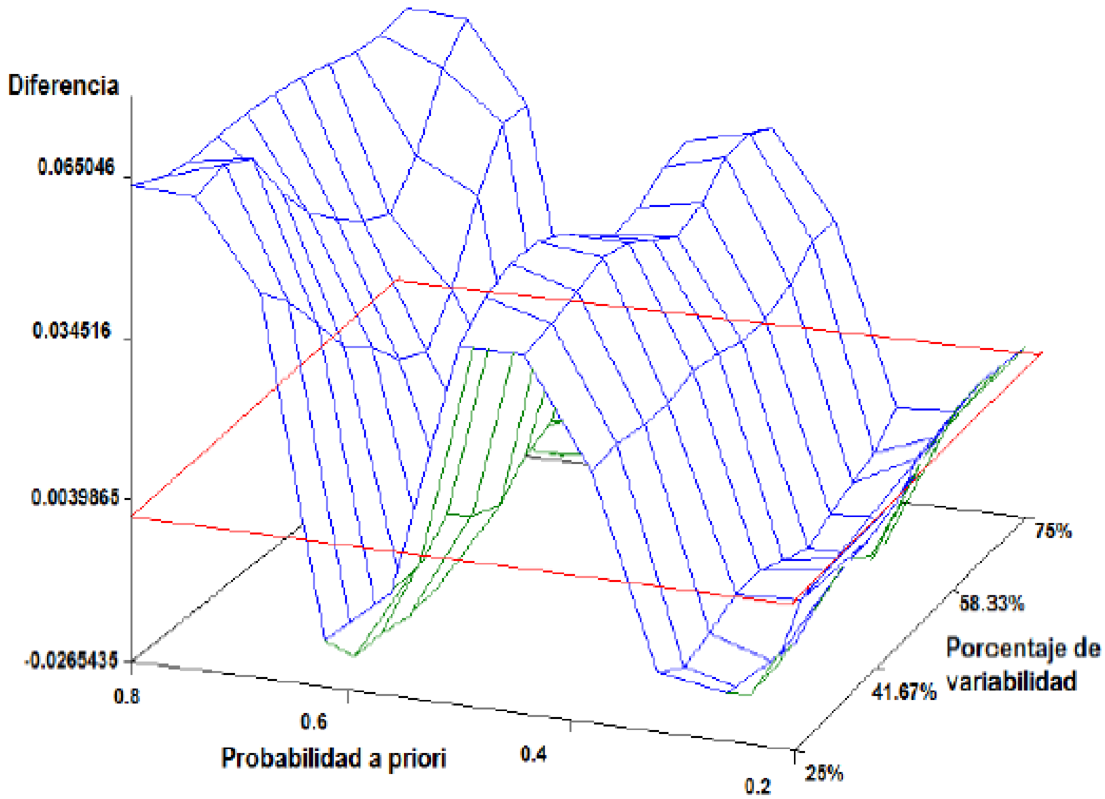

Figura 4.14. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la probabilidad a priori y el porcentaje de variabilidad

La superficie trazada en la figura 4.14, que representa la diferencia de unidades correctamente clasificadas, considerada ahora desde el punto de vista de los factores correspondientes al porcentaje de variabilidad mínima retenida por los componentes principales y a las probabilidades a priori, presenta un aspecto sinuoso con dos concavidades las cuales se encuentran definidas para probabilidades a priori entre 0.2 y 0.4 y de 0.5 hasta 0.8. Para la probabilidad a priori de 0.2 a cualquier porcentaje de variabilidad, la diferencia de unidades correctamente clasificadas por las dos reglas tiende a ser igual, siendo ligeramente mayor para la regla de puntajes lineales discriminantes cuando el porcentaje de variabilidad es de 75 por ciento.

La sima de la primera concavidad, tiende a inclinarse desde el 25 hasta el 75 por ciento de variabilidad y se extiende a través de la probabilidad a priori de 0.3. La sima de la segunda concavidad de comportamiento similar al primero corresponde a una probabilidad a priori de 0.6. Por estar ambos localizados debajo del plano que representa el origen, corresponden al mejor desempeño de la regla propuesta, hecho que es valido para cualquier porcentaje de variabilidad.

De manera similar las cimas corresponden a probabilidades a priori de 0.5 y de 0.8, lo que manifiesta que la regla de puntajes lineales discriminantes desempeña un mejor trabajo en la clasificación de unidades, para cualquier porcentaje de variabilidad en estas dos probabilidades a priori. Ver figura 4.14.

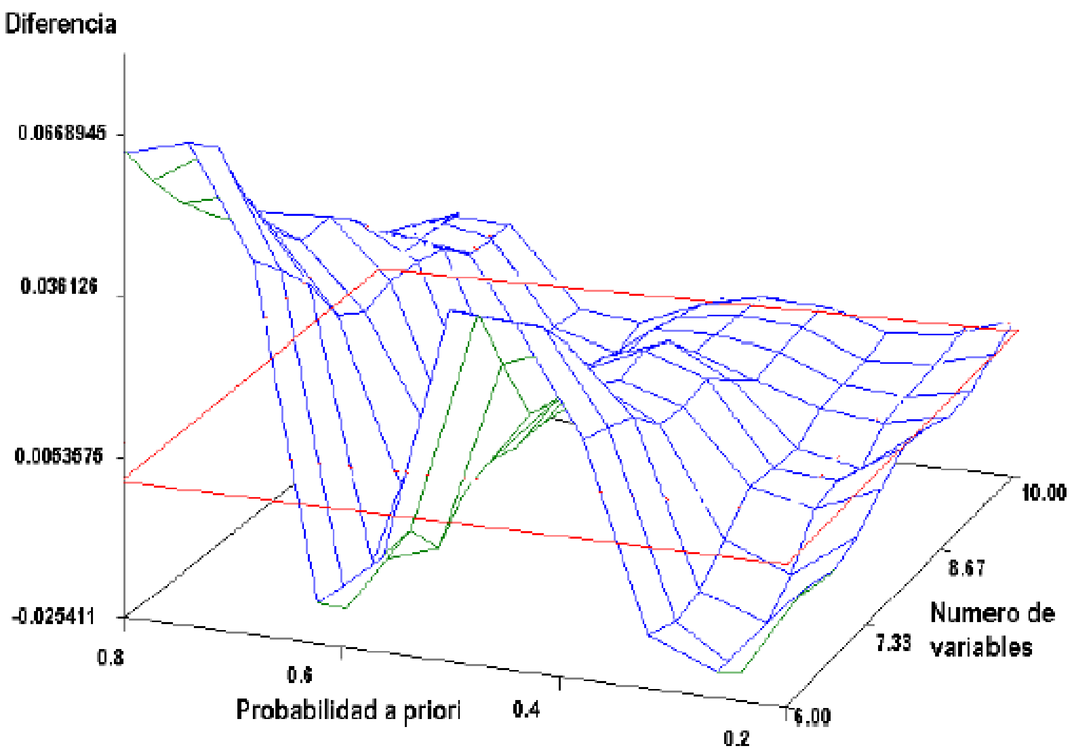

Figura 4.15. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el probabilidad a priori y el número de variables

Tomando en cuenta los factores probabilidad a priori y el número de variables, de acuerdo a lo representado en la figura 4.15, se puede advertir una superficie con ondulaciones, menos pronunciadas que el caso anterior pero que permiten observar la conformación de dos pequeñas elevaciones asentadas entre las probabilidades a priori 0.3 y 0.6 para la primera y 0.6 y 0.8 para la segunda; este hecho señala que la regla propuesta vuelve a tener un mejor desempeño para probabilidades a priori de 0.3 y de 0.6, en tanto que el mejor desempeño para la regla de puntajes lineales discriminantes ocurre para probabilidades a priori de 0.5 y 0.8, dado un número de variables pequeño.

A medida que aumenta el numero de variables, tanto para las probabilidades a priori de 0.6 y de 0.3 la diferencia en la proporción de unidades estadísticas correctamente clasificadas tiende a disminuir alcanzando diferencias mínimas para diez variables, este hecho se hace más notorio para la probabilidad a priori de 0.3.

Análogamente, la cima de las pequeñas elevaciones va declinando su altura a medida que aumenta el número de las variables, reduciendo la diferencia de proporción de unidades correctamente clasificadas por las dos reglas, por lo que se puede afirmar que para probabilidades de 0.5 y 0.8 con un número de variables de diez, la diferencia entre las dos reglas tiende a disminuir, y se hace más notoria cuando el numero de variables es pequeño. Ver figura 4.15.

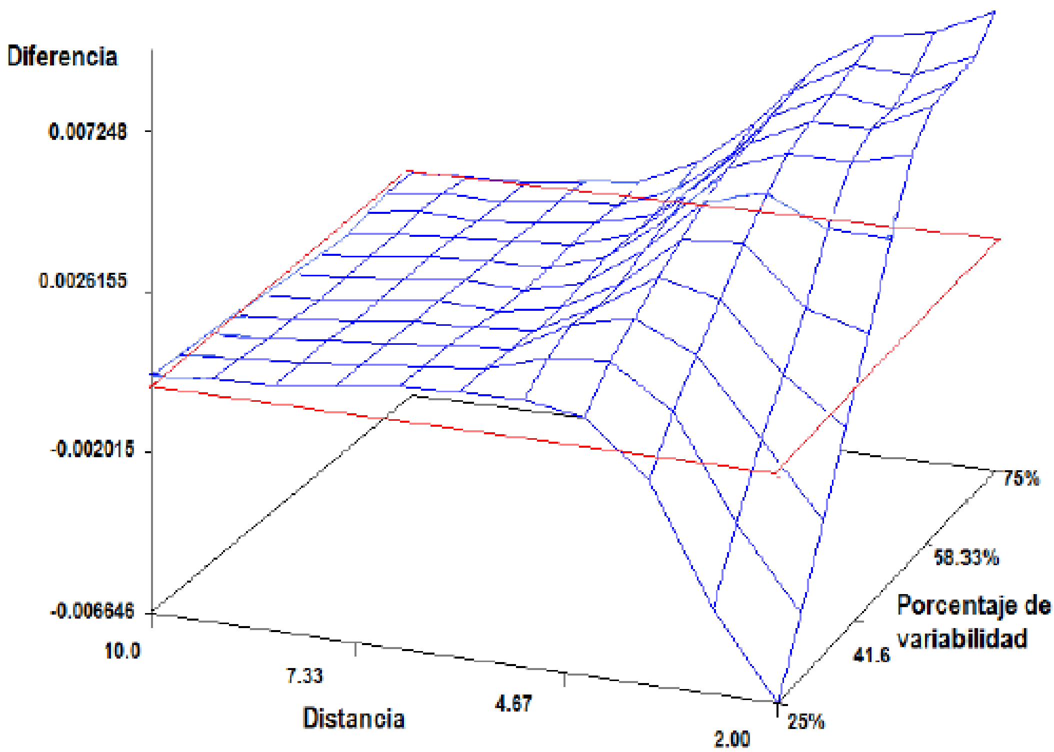

Figura 4.16. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la distancia y el porcentaje de variabilidad

Examinando la interacción de los factores distancia y porcentaje de variabilidad en el ejercicio de clasificación de unidades estadísticas realizada por las dos reglas, la figura 4.16 pone de manifiesto la configuración de dos subregiones: una planicie que está delimitada entre las distancias diez y alrededor de cinco y para todo porcentaje de variabilidad, y que parece reposar sobre el plano del origen, en cuyo caso la clasificación correcta de unidades estadísticas realizada por las dos reglas se podría colegir que es igual.

La segunda subregión se encuentra ubicada en distancias menores de cinco para todo porcentaje de variabilidad. Aquí se aprecia que para porcentajes de variabilidad de 25 y a distancias menores de cuatro, la diferencia en la proporción de unidades correctamente clasificadas es favorable a la regla propuesta; por otra parte la superficie manifiesta, en esta misma subregión, un desempeño con más ventaja de la regla de puntajes lineales discriminantes alrededor de un porcentaje de variabilidad del 45 por ciento hasta alcanzar un tope máximo a el 75 por ciento de variabilidad.Ver figura 4.16

Explorando la clasificación correcta de unidades estadísticas por las dos reglas bajo los factores: número de variables y probabilidad a priori de acuerdo a la figura 4.17, la región de compara-

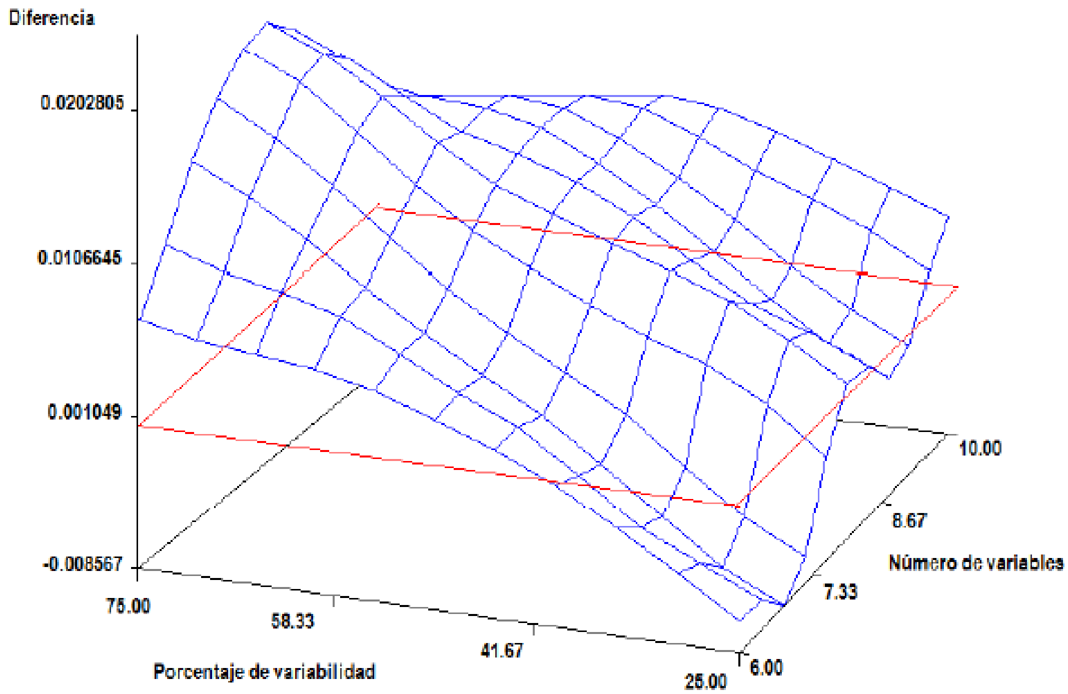

Figura 4.17. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta el porcentaje de variabilidad y el número de variables

ción de las reglas presenta una inclinación sostenida, que su parte más alta se encuentra para porcentajes de variabilidad de 75 por ciento y con un número de variables alrededor de siete y que cae a cada uno de los costados, pero que se mantiene por encima del plano que representa el origen, hecho que señala que la regla de puntajes lineales discriminantes tiene mejor desempeño que la regla propuesta.

De manera similar se observa un declive más pronunciado que concluye hacia el 25 por ciento de variabilidad y que a partir de un porcentaje de variabilidad del 40 por ciento se torna favorable a la regla propuesta, este hecho se mantiene hasta el 25 por ciento de variabilidad y un número de variables alrededor de ocho, donde la diferencia de la proporción de unidades clasificadas correctamente por las dos reglas tiende a disminuir, sin embargo para un número de variables de diez la clasificación correcta de unidades es favorable a la regla de puntajes lineales discriminantes sin importar el porcentaje de variabilidad. Ver figura 4,17

Al igual que en la mayoría de los casos anteriores cuando los factores estudiados son: el número de variables y la distancia, se presentan dos subregiones claramente definidas dentro de la región en la cual se realizó la simulación. La primera está conformada por una llanura, delimitada por distancias alrededor de cinco hasta la distancia diez para cualquier cantidad de variables y que de igual manera que en los anteriores casos ésta parece confundirse con el plano del origen. Esta subregión se caracteriza porque la diferencia de la proporción del número de unidades correctamente clasificadas por las dos reglas tiende a ser igual, hecho que se pone de manifiesto

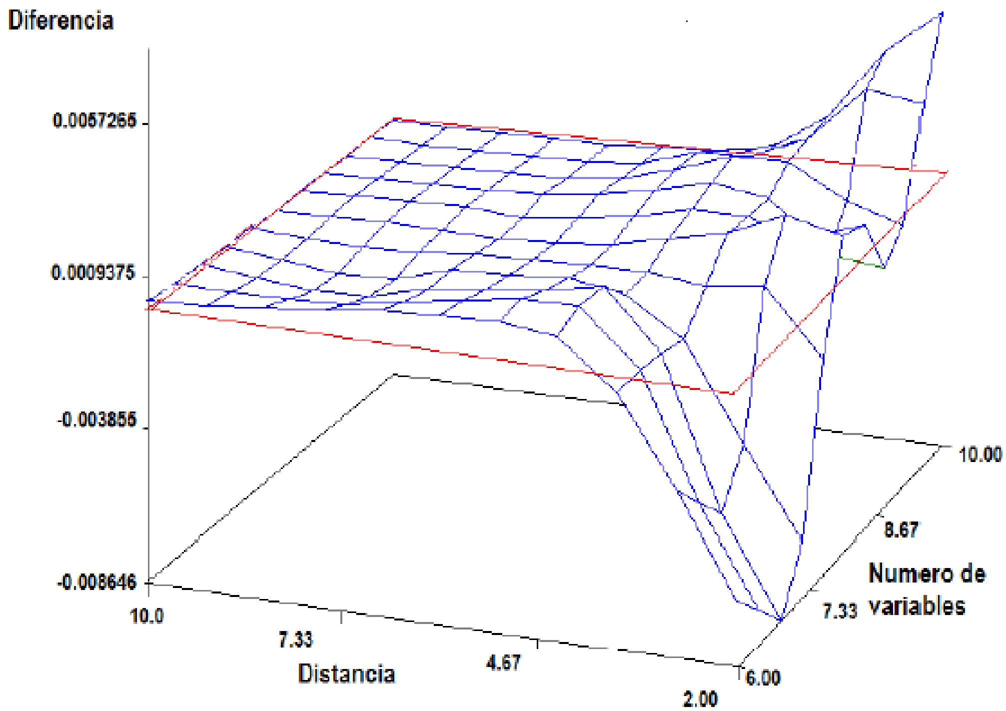

Figura 4.18. Diferencia en la proporción de unidades correctamente clasificadas por las dos reglas teniendo en cuenta la distancia y el número de variables

a medida que aumenta la distancia.

Las mayores diferencias en la proporción de unidades correctamente clasificadas por las dos reglas se presentan en la otra subregión que se encuentra delimitada para distancias menores de cinco hasta la distancia de dos y en la que se aprecia una fuerte caída para un número de variables de siete y una distancia de dos siendo favorable a la regla propuesta, a distancias pequeñas y con un número de variables igual a diez, el mejor desempeño corresponde a la regla de puntajes lineales discriminantes. ver figura 4.18.

Capítulo 5

Conclusiones

La simulación fue el recurso medular de este trabajo, planeada y llevada a cabo para reproducir situaciones factibles y controlada para los niveles escogidos de los factores definidos, con el propósito de evaluar inicialmente de una manera no analítica el desempeño de la regla de clasificación que esta tesis propuso y examinó. Por tal razón las conclusiones están sujetas a estas condiciones experimentales y sus alcances no trascienden más allá del ejercicio simulador. Aunque puntualmente en el capítulo anterior se hicieron afirmaciones de carácter concluyente, en este capítulo se hacen aseveraciones generales que de manera sintética ultiman el trabajo y destacan lo más relevante del comportamiento de la regla propuesta.

- En términos generales se puede afirmar que la regla propuesta presenta un buen desempeño en su función discriminadora, cuando no se reduce el espacio p-dimensional, al igual que cuando la clasificación debe de hacerse necesariamente usando los componentes principales.
- La regla de puntajes lineales discriminantes, si no utiliza los componentes principales, muestra indiscutiblemente una superioridad frente a la regla propuesta cuando ésta emplea dos o tres componentes principales admisibles.
- El desempeño de la regla propuesta con dos componentes principales es mejor que cuando ésta emplea tres componentes admisibles.
- En la región de estudio, la diferencia entre las dos reglas se reduce definitivamente cuando la distancia entre los centroides aumenta, de igual manera, pero menos evidente cuando el número de variables es mayor.
- La regla propuesta se desempeña bien en algunas partes de la región analizada, y donde el desempeño es favorable a la regla de puntajes lineales discriminantes las diferencias presentadas con la regla propuesta, no son demasiado grandes.

Con una primera evaluación no analítica sobre el desempeño de la regla propuesta, surgen algunas inquietudes cuyas soluciones se pueden buscar en trabajos posteriores, como por ejemplo en lugar de usar el valor de 0.5 usar el inverso de la norma del centroide, cuya distancia con la unidad estadística es menor. Se hace inevitable sondear el desempeño de la regla propuesta bajo simulación en condiciones de no normalidad, de igual manera indagar la eficacia de la regla para el caso de matriz de covarianzas diferentes.

Bibliografía

AFIFI, A.A.and CLARK, V. (1996) Computer-aided Multivariate Analysis. Third Edition. Texts in statistical Science. Chapman and Hall.

BASILEVSKY ALEXANDER Applied Matrix Algebra in the Statistical sciences. Elsevier Science Publishing (1983)

BROW P.J. FEARN T. and HAQUE M.S. Journal of American Statistical Association Vol. 94 No.448 (Dec. 1999) http://links.jstor.org/sici.

CRITCHELEY F. And FORD I.Interval Estimation in Discrimination: The Multivariate Normal Equal Covariance case Biometrika, vol 72, No.1(April, 1985)http:/links.jstor.org

HABBEMA, J. and HERMANS, J.(1979) Selection of variables in discriminant analysis by F-Statistic and error rate Techometrics, No.19 487-493 http:/links.jstor.org

HUBERTY,C.J. (1994). Discriminant Analysis. Review of Educational Research, Vol.45.No.4 (Autumn. 1994), 543-598 http:/links.jstor.org

JOHSON R. and WICHERN D. (1999) Applied Multivariate Statistical Analysis Pretince Hall

KING LESLIE J. Discriminant Analysis: Economic Geography, Vol 46, Suplement: proceedings. International Geographical Union Commission on quantitative Methods (jun. 1970), 367-378. http:/links.jstor.org

Peng Lizhi, Yang Bo, Chen Yuehui, Abraham Ajith. Information Sciences 179 (2009) 809-819. Journal homepage:www.elsiever.com /locate/ins.

O'GORMAN T. and WOOLSON R. (1991) Variable selection to discriminate between two groups,http:/links.jstor org/about7terms.html

MORRISON G. DONALD On the Interpretation of Discriminant Analysis, Jornal of Marketing Research, Vol. 6, No.2 (May, 1969), 156-163.http:/links.jstor org/sici

PEÑA D. (2002) Análisis de datos multivariantes M_c Graw Hill editora Concepcion Fernandez Madrid

SEBER G.A. (1984). Observations Multivariate John Wiley and sons

SHARMA, S. (1998). Applied Multivariate Techiques. John Wiley and sons.

Apéndice A

Programas de simulación para la clasificación de unidades estadísticas usando la regla de puntajes lineales discriminantes y la regla propuesta para las diferentes pautas

```
proc iml; /*Programa de clasificación de unidades estadísticas usando la regla de puntajes li-
neales discriminante (puede ser utilizado con componentes principales o sin ellos, según sea el
caso se activan o desactivan los comandos correspondientes)y la regla propuesta para dos y tres
componentes principales*/ k=6;
n=1000;
apriori1=0;
apriori2=0;
distac=0;
do k=6 to 10 by 1;
print ”DOS COMPONENTES PRINCIPALES”; print ”Numero de variables”; print k; do w=0.2
to 0.8 by 0.2;
apriori1=w; apriori2=1-apriori1; print ”probabilidades apriori grupo 1 y grupo 2”; print apriori1
apriori2; l2=1;
clasificacion=j(4,5);
do z1=2 to 10 by 2;
distac=z1**2;
suma1=0; suma2=0; suma3=0; suma4=0; x=1;
do while(x¡1001);
l=k; b=1; a=j(k); sigmainv=j(k); vp=j(k,1); vecp=j(k,k);
combinacion=j(k,k); mu1=j(k,1,1); mu2=j(k,1,1);
vecdif1=j(k,1,1); vecsuma=j(k,1,1);
unos=j(1,k); unos1=j(1,2); cosalfha=j(k,1); cosbetha=j(k,1); cosdirec=j(k,1);
distancia11=j(2,1); distancia12=j(2,1); distancia21=j(2,1); distancia22=j(2,1); nucleopobla1=j(2,1);
nucleopobla2=j(2,1); pobla1=j(2,1); pobla2=j(2,1);
nuevaunidad1=j(k, n); nuevaunidad2=j(k, n); unidad1=j(k, n);
unidad2=j(k, n); nucleounidad1=j(2,n); nucleounidad2=j(2,n); skip:
a=j(k);
/*Generación del centroide de la población uno*/
do i=1 to k by 1;
mu1[i]=abs(rannor(0));
/* Generación de la matriz de covarianzas*/
do j=1 to k by 1;
h=8;
yk=abs(rannor(0));
sk:
rol=(rannor(5));
if rol¡(-1)then goto sk;
else if rol¿1 then goto sk;
else if i=j then a[i, j]=yK;
else a[i, j]=sqrt(abs(a[i]))*sqrt(abs(a[h]))*rol;
a[j, i]=a[i, j];
end;
h=h+j; end; call eigen(vp,vecp,a);
/* condición para que las dos o tres componententes principales retengan una variabilidad ad-
misible*/
```

if any(vp¡0) then goto skip;
else varia=trace(a);
compo12=(vp[1]+vp[2])/varia;
compo23=compo12+vp[3]/varia;v if compo23¡0.95 then goto skip;
else norma1=sqrt(t(mu1)*mu1);
combinacion=t(vecp); skz:
/*generación del centroide de la población dos a una distancia determinada del centroide de la población uno*/ do i1=1 to k by 1;
cosalfha[i1]=mu1[i1]/norma1; cosbetha[i1]=rannor(0);
if cosbetha[i1]¡(0) then goto skz;
else if cosbetha[i1]¿1 then goto skz;
else cosdirec[i1]=cosalfha[i1]*cosbetha[i1]; end; costetha=unos*cosdirec; if costetha¡(-1) then goto skz;
else if costetha¿1 then goto skz; else b=2*costetha*norma1;
ac=(norma1**2)-distac; radical=(b**2)-4*ac;
if radical¡0 then goto skz; else norma2=abs((b+sqrt(radical))/2);
do j1=1 to k by 1;
mu2[j1]=cosbetha[j1]*norma2; end; do i2=1 to k by 1;
vecdif1[i2]=mu1[i2]-mu2[i2]; vecsuma[i2]=mu1[i2]+mu2[i2]; end;
/*generación de nuevas unidades estadísticas para la población uno y para la población dos esta generación de nuevas unidades se hace con semillas al azar, de tal manera que cada semilla va a ser diferente a medida que se genera unas nuevas condiciones de simulación*/ d1=ranuni(0);
da=ranuni(3); daa=d1*da*100; d2=ranuni(23); de=ranuni(5);
dee=d2*de*101; seed=daa; p=nrow(a);
m1=repeat(t(mu1),n,1); g=root(a);
z=normal(repeat(seed,n,p));
y1=z*g+m1; seed=dee; g=root(a);
z=normal(repeat(seed,n,p));
m2=repeat(t(mu2),n,1); y2=z*g+m2;
nuevaunidad1=t(y1); nuevaunidad2=t(y2); sigmainv=inv(a);x1o=0; x2o=0;y1o=0;xf1=0;xf2=0;
yf1=0;
yf2=0;tacig1=0;tacig2=0;grupo1=0; grupo2=0;yot=0;
fuerzg1=0; fuerzg2=0; propuesta1=0; propuesta2=0;yft=0; q=1;q0=1;q1=1;
loapri1=log(apriori2/apriori1); loapri2=log(apriori1/apriori2);
/* se encuentra el núcleo de cada una de las poblaciones de acuerdo a
los componentes principales admisibles*/ pobla1=combinacion*mu1; pobla2=combinacion*mu2;
do l=1 to 2 by 1; nucleopobla1[l,1]=pobla1[l,1]; nucleopobla2[l,1]=pobla2[l,1]; end;
/* Se encuentra el núcleo de las nuevas unidades estadísticas*/ vecdif=pobla1-pobla2; vecsuma= pobla1+pobla2; do while (q¡n+1);
do f=1 to 2 by 1;
nucleounidad1[f, q]=combinacion[f,]*nuevaunidad1[,q]; nucleounidad2[f, q]=combinacion[f,]*nuevaunidad2[,q];
end; q=q+1; end; do while (q0¡n+1);
do f=1 to n by 1;
/*Para utilizar la regla de puntajes lineales discriminantes

```
cuando emplea los componentes principales*/ unidad1[f, q0]=combinacion[f, ]*nuevaunidad1[
,q0]; unidad2[f, q0]=combinacion[f, ]*nuevaunidad2[ ,q0]; end; q0=q0+1; end;
do while (q1¡n+1); do f1=1 to 2 by 1;
distancia11[f1,1]=(nucleopobla1[f1,1]-nucleounidad1[f1,q1])**2;
distancia12[f1,1]=(nucleopobla2[f1,1]-nucleounidad1[f1,q1])**2;
distancia21[f1,1]=(nucleopobla1[f1,1]-nucleounidad2[f1,q1])**2;
distancia22[f1,1]=(nucleopobla2[f1,1]-nucleounidad2[f1,q1])**2;
end;
fa11=sqrt(t(distancia11)*distancia11);
fa12=sqrt(t(distancia12)*distancia12);
fa21=sqrt(t(distancia21)*distancia21);
fa22=sqrt(t(distancia22)*distancia22);
/* la condición para asignar los valores de 1 o 0.5 según sea el caso en la regla propuesta*/ if
fa11¡=fa12 then fa11=0.5; else fa11=1;
if fa12¡=fa11 then fa12=0.5; else fa12=1;
if fa22¡=fa21 then fa22=0.5; else fa22=1;
if fa21¡=fa22 then fa21=0.5; else fa21=1;
rp11=t(vecdif1)*sigmainv*unidad1[, q1];
rp12=0.5*t(vecdif1)*sigmainv*vecsuma;
rp21=t(vecdif1)*sigmainv*unidad2[, q1];
reglag1=rp11-rp12; reglag2=rp21-rp12;
if reglag1¿=loapri1 then x1o=x1o+1; else y1o=y1o+1;
if reglag2¿=loapri1 then x2o=x2o+1; else y2o=y2o+1;
masa11=(unos1)*(distancia11);
masa12=(unos1)*(distancia12); masa21=(unos1)*(distancia21); masa22=(unos1)*(distancia22);
/* calculo de la fuerza de atracción en la regla propuesta, para el caso donde
la probabilidad a priori de las dos poblaciones sea igual, el cociente de las
probabilidades a priori se reemplaza por uno*/
fuerza11=abs(loapri1)/fa11;
fuerza12=abs(loapri2)/fa12; fuerza21=abs(loapri1)/fa21;
fuerza22=abs(loapri2)/fa22; if fuerza11¿=fuerza12 then xf1=xf1+1;
else yf1=yf1+1;
if fuerza22¿fuerza21 then xf2=xf2+1; else yf2=yf2+1;
q1=q1+1; end;
tacig1=tacig1+x1o; tacig2=tacig2+y2o; fuerzg1=fuerzg1+xf1; fuerzg2=fuerzg2+xf2;
suma1=suma1+tacig1; suma2=suma2+tacig2; suma3=suma3+fuerzg1;
suma4=suma4+fuerzg2; x=x+1; end;
grupo1=grupo1+suma1; grupo2=grupo2+suma2; propuesta1=propuesta1+suma3;
propuesta2=propuesta2+suma4; clasificacion[1,l2]=grupo1;
clasificacion[2,l2]=propuesta1;
clasificacion[3,l2]=grupo2;
clasificacion[4,l2]=propuesta2;
l2=l2+1; end;
reglas=Tradicional1, propuesta1, Tradicional2, propuesta2;
probabilidades=dist2, dist4, dist6, dist8, dist10;
```

```
print clasificacion[rowname=reglas colname=probabilidades];
end; end; run;
```

```
proc iml;
/*Programa de clasificación de unidades estadísticas usando la regla de puntajes lineales discri-
minante y la regla propuesta sin reducir el espacio p-dimensional*/
k=6;
n=10;
apriori1=0;
apriori2=0;
distac=0;
do k=6 to 10 by 1;
print "Numero de variables";
print k;
do w=0.2 to 0.8 by 0.2;
apriori1=w;
apriori2=1-apriori1;
print "probabilidades apriori grupo 1 y grupo 2";
print apriori1 apriori2;
l2=1;
clasificacion1=j(2,5);
clasificacion2=j(2,5);
do z1=2 to 10 by 2;
distac=z1**2;
suma1=0;
suma2=0;
suma11=0;suma12=0;suma21=0;suma22=0;
suma31=0;suma32=0;
x=1;
do while(x¡11);
l=k;
b=1;
a=j(k);
sigmainv=j(k);
vp=j(k,1);
vecp=j(k,1);
combinacion=j(k,k);
mu1=j(k,1,1);
mu2=j(k,1,1);
vecdif1=j(k,1,1);
vecsuma=j(k,1,1);
unos=j(1,k);
unos1=j(1,k);
unos2=j(k,1);
cosalfha=j(k,1);
cosbetha=j(k,1);
cosdirec=j(k,1);
distancia11=j(k,1);
```

```
distancia12=j(k,1);
distancia21=j(k,1);
distancia22=j(k,1);
desviacion11=j(k,1);
desviacion12=j(k,1);
desviacion21=j(k,1);
desviacion22=j(k,1);
sumapobla=j(k,1);
skip:
a=j(k);
/*Generación del centroide de la población uno*/
do i=1 to k by 1;
mu1[i]=abs(rannor(0));
/* Generación de la matriz de covarianzas*/
do j=1 to k by 1;
h=8;
yk=abs(rannor(0));
sk:
rol=(rannor(5));
if rol¡(-1)then goto sk;
else if rol¿1 then goto sk;
else if i=j then a[i, j]=yK;
else a[i, j]=sqrt(abs(a[i]))*sqrt(abs(a[h]))*rol;
a[j, i]=a[i, j];
end;
h=h+j;
end;
/* Comprobación de que la matriz de varianzas retiene el porcentaje de variación deseada*/
call eigen(vp,vecp,a);
if any(vp¡0) then goto skip;
else varia=trace(a);
compo12=(vp[1]+vp[2])/varia;
compo23=compo12+vp[3]/varia;
if compo12¡0.5 then goto skip;
else norma1=sqrt(t(mu1)*mu1);
combinacion=t(vecp);
skz:
/* Generación de el centroide de la población dos a una distancia deseada del centroide uno*/
do i1=1 to k by 1;
cosalfha[i1]=mu1[i1]/norma1;
cosbetha[i1]=rannor(0);
if cosbetha[i1]¡(0) then goto skz;
else if cosbetha[i1]¿1 then goto skz;
else cosdirec[i1]=cosalfha[i1]*cosbetha[i1];
end;
```

```
costetha=unos*cosdirec;
if costetha¡(-1) then goto skz;
else if costetha¿1 then goto skz;
else b=2*costetha*norma1;
ac=(norma1**2)-distac;
radical=(b**2)-4*ac;
if radical¡0 then goto skz;
else norma2=abs((b+sqrt(radical))/2);
do j1=1 to k by 1;
mu2[j1]=cosbetha[j1]*norma2;
end;
do i2=1 to k by 1;
vecdif1[i2]=mu1[i2]-mu2[i2];
vecsuma[i2]=mu1[i2]+mu2[i2];
end;
/* Generación de unidades estadísticas a partir de un centroide de la población y la matriz de
varianzas*/
d1=ranuni(0);
da=ranuni(3);
daa=d1*da*100;
d2=ranuni(23);
de=ranuni(5);
dee=d2*de*101;
seed=daa;
p=nrow(a);
m1=repeat(t(mu1),n,1);
g=root(a);
z=normal(repeat(seed,n,p));
y1=z*g+m1;
seed=dee;
g=root(a);
z=normal(repeat(seed,n,p));
m2=repeat(t(mu2),n,1);
y2=z*g+m2;
nuevaunidad1=t(y1);
nuevaunidad2=t(y2);
sigmainv=inv(a);
x1o=0; x2o=0;y1o=0;y2o=0;
xf11=0;xf12=0;yf1=0;yf2=0;
xf21=0;xf22=0;yf12=0;yf22=0;
xf31=0;xf32=0;yf13=0;yf23=0;
tacig1=0;
tacig2=0;
grupo1=0;
grupo2=0;
```

```
yot=0;
fuerzg11=0;fuerzg12=0;fuerzg21=0;fuerzg22=0;
fuerzg31=0;fuerzg32=0;
propuesta11=0;propuesta12=0;
propuesta21=0;propuesta22=0;
propuesta31=0;propuesta32=0;
yft=0;
q=1;
q1=1;
loapri1=log(apriori2/apriori1);
loapri2=log(apriori1/apriori2);
poblacion1=mu1;
poblacion2=mu2;
do while (q1¡n+1);
do f1=1 to k by 1;
distancia11[f1,1]=((poblacion1[f1,1]-nuevaunidad1[f1,q1])**2);
distancia12[f1,1]=((poblacion2[f1,1]-nuevaunidad1[f1,q1])**2);
distancia21[f1,1]=((poblacion1[f1,1]-nuevaunidad2[f1,q1])**2);
distancia22[f1,1]=((poblacion2[f1,1]-nuevaunidad2[f1,q1])**2);
desviacion11[f1,1]=(poblacion1[f1,1]-nuevaunidad1[f1,q1]);
desviacion12[f1,1]=(poblacion2[f1,1]-nuevaunidad1[f1,q1]);
desviacion21[f1,1]=(poblacion1[f1,1]-nuevaunidad2[f1,q1]);
desviacion22[f1,1]=(poblacion2[f1,1]-nuevaunidad2[f1,q1]);
end;
sumapobla=vecdif1;
fa11=sqrt(t(desviacion11)*(desviacion11));
fa12=sqrt(t(desviacion12)*(desviacion12));
fa21=sqrt(t(desviacion21)*(desviacion21));
fa22=sqrt(t(desviacion22)*(desviacion22));
if fa11¡=fa12 then fa11=0.5;
else fa11=1;
if fa12¡=fa11 then fa12=0.5;
else fa12= 1;
if fa22¡=fa21 then fa22=0.5;
else fa22=1;
if fa21¡=fa22 then fa21=0.5;
else fa21=1;
rp11=t(vecdif1)*sigmainv*nuevaunidad1[, q1];
rp12=0.5*t(vecdif1)*sigmainv*vecsuma;
rp21=t(vecdif1)*sigmainv*nuevaunidad2[, q1];
reglag1=rp11-rp12;
reglag2=rp21-rp12;
if reglag1¿=loapri1 then x1o=x1o+1;
else y1o=y1o+1;
if reglag2¿=loapri1 then x2o=x2o+1;
```

```
else y2o=y2o+1;
masa11=(unos1)*(distancia11);
masa12=(unos1)*(distancia12);
masa21=(unos1)*(distancia21);
masa22=(unos1)*(distancia22);
fuerza11=abs(loapri1)/fa11;
fuerza12=abs(loapri2)/fa12;
fuerza21=abs(loapri1)/fa21;
fuerza22=abs(loapri2)/fa22;
if fuerza11¿=fuerza12 then xf11=xf11+1;
else yf1=yf1+1;
if fuerza22¿=fuerza21 then xf12=xf12+1;
else yf2=yf2+1;
q1=q1+1;
end;
tacig1=tacig1+x1o;
tacig2=tacig2+y2o;
fuerzg11=fuerzg11+xf11;
fuerzg12=fuerzg12+xf12;
suma1=suma1+tacig1;
suma2=suma2+tacig2;
suma11=suma11+fuerzg11;
suma12=suma12+fuerzg12;
x=x+1;
end;
grupo1=grupo1+suma1;
grupo2=grupo2+suma2;
propuesta11=propuesta11+suma11;
propuesta12=propuesta12+suma12;
clasificacion1[1,l2]=grupo1;
clasificacion1[2,l2]=propuesta11;
clasificacion2[1,l2]=grupo2;
clasificacion2[2,l2]=propuesta12;
l2=l2+1;
end;
reglas=Tradicional, propuesta11, propuesta21, propuesta31;
probabilidades=dist2, dist4, dist6, dist8, dist10;
print RESULTADO DE LA CLASIFICACION USANDO LA REGLA TRADICIONAL Y LA”;
print REGLA PROPUESTAS PARA EL GRUPO 1”;
print clasificacion1[rowname=reglas colname=probabilidades];
reglas=Tradiciona2, propuesta12, propuesta22, propuesta32;
probabilidades=dist2, dist4, dist6, dist8, dist10;
print RESULTADO DE LA CLASIFICACION USANDO LA REGLA TRADICIONAL Y LA”;
print REGLA PROPUESTAS PARA EL GRUPO 2”;
print clasificacion2[rowname=reglas colname=probabilidades];
```

```
end;
end;
run;
```

Apéndice B

Procedimiento para encontrar el vector de la población dos a partir de una distancia determinada con relación a el vector de la población

Teorema Sean $X_1 = (x_{11}, x_{12}, ..., x_{1n})$ y $X_2 = (x_{21}, x_{22}, ..., x_{2n})$ dos vectores diferentes de cero en el espacio n-dimensional. Si θ es el angulo entre X_1 y X_2 y si $a_1, a_2, ..., a_n$, y $b_1, b_2, ..., b_n$, son cosenos direccionales de X_1 y X_2 respectivamente, entonces:

$cos\theta = a_1b_1, a_2b_2, ..., a_nb_n,$, donde $cos\alpha_i = a_i$ y $cos\beta_i = b_i$, siendo α_i y β_i ángulos formados por X_1 y X_2 además:

$$| X_1 \cdot X_2 | = \| X_1 \| \| X_2 \| cos\theta \tag{B.1}$$

Se genera el centroide del grupo de la población uno con números aleatorios, una vez obtenido este vector se encuentra el vector de los cosenos alpha, estos son los cosenos direccionales del vector del centroide del grupo uno y que se obtienen a partir de:

$$\mu_1 \diagup \| \mu_1 \| \tag{B.2}$$

Una vez obtenidos los cosenos direccionales del grupo uno, se obtiene en forma aleatoria el vector de cosenos betha entre cero y uno con el fin de obtener el centroide del grupo dos positivo. Para obtener el vector de coseno tetha se encuentra cada elemento del vector con el producto de cada uno de los elementos de los vectores alpha y betha, de acuerdo con el teorema anterior. La distancia entre dos vectores μ_1 y μ_2 esta dada por:

$$d(\mu_1, \mu_2) = (\mu_1 - \mu_2) \cdot (\mu_1 - \mu_2) \tag{B.3}$$

$$d(\mu_1, \mu_2) = \| \mu_1 \|^2 - 2(\mu_1 \cdot \mu_2) + \| \mu_2 \|^2 \tag{B.4}$$

De la ecuación $D.1$, reemplazando en $D.4$ tenemos:

$$d(\mu_1, \mu_2) = \| \mu_1 \|^2 - 2\cos\theta \| \mu_1 \| \| \mu_2 \| + \| \mu_2 \|^2 \tag{B.5}$$

Si queremos generar el centroide de la población dos a una distancia conocida entonces:

$$\| \mu_1 \|^2 - 2\cos\theta \| \mu_1 \| \| \mu_2 \| + \| \mu_2 \|^2 - d(\mu_1, \mu_2) = 0 \tag{B.6}$$

Resolviendo la ecuación anterior podemos encontrar la norma del centroide de la población dos y como se conoce los cosenos betha, podemos hallar cada uno de los elementos del centroide con la ecuación $D.2$, pero en este caso aplicado a la población dos.

$$\mu_2 = \cos\beta \times \| \mu_2 \| \tag{B.7}$$

Apéndice C

Resultados de la simulación de la clasificación de unidades estadísticas por la regla propuesta y la regla de puntajes lineales discriminante para dos y tres componentes principales

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,900444	0,859592	0.040852	0.2	6	2
0,9885605	0,9290515	0,059509	0.2	6	4
0,997279	0,965704	0,031575	0.2	6	6
0,9992375	0,982729	0,0165085	0.2	6	8
0,9999485	0,9947725	0,005176	0.2	6	10
0,9716295	0,8678445	0,103785	0.4	6	2
0,993741	0,9326755	0,0610655	0.4	6	4
0,9991955	0,975699	0,0234965	0.4	6	6
0,99764	0,981853	0,015787	0.4	6	8
0,999903	0,9855435	0,0143595	0.4	6	10
0,9834645	0,8773795	0,106085	0.6	6	2
0,9962205	0,928769	0,0674515	0.6	6	4
0,999504	0,9709335	0,0285705	0.6	6	6
0,99863	0,979228	0,019402	0.6	6	8
0,9999255	0,986034	0,0138915	0.6	6	10
0,988588	0,8575135	0,1310745	0.8	6	2
0,997948	0,9154685	0,0824795	0.8	6	4
0,9994915	0,9573455	0,042146	0.8	6	6
0,9998395	0,974114	0,0257255	0.8	6	8
0,9999935	0,991773	0,0082205	0.8	6	10
0,965927	0,866175	0,099752	0.2	7	2
0,9962515	0,935417	0,0608345	0.2	7	4
0,997981	0,960555	0,037426	0.2	7	6
0,999995	0,9829195	0,0170755	0.2	7	8
0,9999975	0,985203	0,0147945	0.2	7	10
0,9832225	0,881837	0,1013855	0.4	7	2
0,997034	0,9405405	0,0564935	0.4	7	4
0,9997205	0,9754585	0,024262	0.4	7	6
0,999988	0,9743905	0,0255975	0.4	7	8
0,9999985	0,992886	0,0071125	0.4	7	10
0,99015	0,8717065	0,1184435	0.6	7	2
0,9982835	0,93298	0,0653035	0.6	7	4
0,9998265	0,972804	0,0270225	0.6	7	6
0,9999955	0,9855355	0,01446	0.6	7	8
0,9999995	0,9875055	0,012494	0.6	7	10
0,995071	0,875157	0,119914	0.8	7	2
0,9993015	0,93912	0,0601815	0.8	7	4
0,9996295	0,977043	0,0225865	0.8	7	6
0,9999995	0,9779895	0,02201	0.8	7	8
0,999999	0,993596	0,006403	0.8	7	10
0,9837455	0,868759	0,1149865	0.2	8	2
0,999417	0,928925	0,070492	0.2	8	4
0,9999815	0,9676265	0,032355	0.2	8	6
0,999999	0,978655	0,021344	0.2	8	8
1	0,986946	0,013054	0.2	8	10
0,9940105	0,871595	0,1224155	0.4	8	2
0,9998535	0,930141	0,0697125	0.4	8	4
0,9999945	0,955102	0,0448925	0.4	8	6
0,999935	0,987395	0,01254	0.4	8	8
1	0,982935	0,017065	0.4	8	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,9964515	0,87526	0,1211915	0.6	8	2
0,9999235	0,91792	0,0820035	0.6	8	4
0,9999965	0,9667025	0,033294	0.6	8	6
0,999957	0,9856485	0,0143085	0.6	8	8
1	0,987548	0,012452	0.6	8	10
0,9971025	0,880631	0,1164715	0.8	8	2
0,9998885	0,9366145	0,063274	0.8	8	4
0,9999995	0,962745	0,0372545	0.8	8	6
1	0,97806	0,02194	0.8	8	8
1	0,987015	0,012985	0.8	8	10
0,992456	0,87804	0,114416	0.2	9	2
0,9999445	0,927279	0,0726655	0.2	9	4
0,999999	0,9704995	0,0294995	0.2	9	6
1	0,98101	0,01899	0.2	9	8
0,9999995	0,9803015	0,019698	0.2	9	10
0,994501	0,861688	0,132813	0.4	9	2
0,9999815	0,936728	0,0632535	0.4	9	4
0,9999595	0,9652435	0,034716	0.4	9	6
1	0,9864325	0,0135675	0.4	9	8
1	0,9917325	0,0082675	0.4	9	10
0,9967035	0,870069	0,1266345	0.6	9	2
0,999992	0,9201225	0,0798695	0.6	9	4
0,9999995	0,9655925	0,034407	0.6	9	6
1	0,9871975	0,0128025	0.6	9	8
1	0,982468	0,017532	0.6	9	10
0,998722	0,875999	0,122723	0.8	9	2
0,999991	0,9238075	0,0761835	0.8	9	4
0,9999995	0,971111	0,0288885	0.8	9	6
1	0,9756515	0,0243485	0.8	9	8
1	0,991701	0,008299	0.8	9	10
0,9993065	0,7995545	0,199752	0.2	10	2
0,9995995	0,883078	0,1165215	0.2	10	4
0,999949	0,920412	0,079537	0.2	10	6
0,9999645	0,9647485	0,035216	0.2	10	8
0,9999995	0,9831255	0,016874	0.2	10	10
0,9994985	0,8864725	0,113026	0.4	10	2
0,9997495	0,9224115	0,077338	0.4	10	4
0,99992	0,9648035	0,0351165	0.4	10	6
0,9999945	0,981943	0,0180515	0.4	10	8
0,9999995	0,9875305	0,012469	0.4	10	10
0,9996545	0,8770075	0,122647	0.6	10	2
0,999828	0,9355615	0,0642665	0.6	10	4
0,9999395	0,966955	0,0329845	0.6	10	6
0,99998	0,978017	0,021963	0.6	10	8
0,9999995	0,9903145	0,009685	0.6	10	10
0,999833	0,868228	0,131605	0.8	10	2
0,9999295	0,934753	0,0651765	0.8	10	4
0,999995	0,9547835	0,0452115	0.8	10	6
0,9999995	0,982501	0,0174985	0.8	10	8
1	0,990036	0,009964	0.8	10	10

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,973163	0,804363	0,1688	0.5	6	2
0,9952555	0,8921065	0,103149	0.5	6	4
0,9982685	0,9262605	0,072008	0.5	6	6
0,999947	0,9535505	0,0463965	0.5	6	8
0,999923	0,9665185	0,0334045	0.5	6	10
0,985998	0,81866	0,167338	0.5	7	2
0,998278	0,8914085	0,1068695	0.5	7	4
0,9998875	0,9096295	0,090258	0.5	7	6
0,999986	0,9343255	0,0656605	0.5	7	8
0,9999995	0,94916	0,0508395	0.5	7	10
0,990915	0,8112705	0,1796445	0.5	8	2
0,997672	0,894279	0,103393	0.5	8	4
0,999942	0,896209	0,103733	0.5	8	6
0,999999	0,9355115	0,0644875	0.5	8	8
0,9999995	0,9468465	0,053153	0.5	8	10
0,998742	0,8058305	0,1929115	0.5	9	2
0,9998555	0,8674105	0,132445	0.5	9	4
0,9999945	0,924884	0,0751105	0.5	9	6
0,9999995	0,9280705	0,071929	0.5	9	8
0,9999995	0,958774	0,0412255	0.5	9	10
0,999033	0,817136	0,181897	0.5	10	2
0,9999945	0,8888995	0,111095	0.5	10	4
0,999999	0,8834905	0,1165085	0.5	10	6
1	0,9472625	0,0527375	0.5	10	8
1	0,9401135	0,0598865	0.5	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro C.1. Porcentaje de unidades estadísticas correctamente clasificadas por la regla de puntajes lineales y la regla propuesta con dos componentes principales admisibles

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,5668105	0,809246	-0,2424355	0.2	6	2
0,679183	0,856352	-0,177169	0.2	6	4
0,7201915	0,903253	-0,1830615	0.2	6	6
0,7786705	0,9452605	-0,16659	0.2	6	8
0,822326	0,9562765	-0,1339505	0.2	6	10
0,591214	0,703606	-0,112392	0.4	6	2
0,6502455	0,8213075	-0,171062	0.4	6	4
0,730399	0,8364405	-0,1060415	0.4	6	6
0,758936	0,90558	-0,146644	0.4	6	8
0,827151	0,9459445	-0,1187935	0.4	6	10
0,6148395	0,872353	-0,2575135	0.6	6	2
0,655755	0,934168	-0,278413	0.6	6	4
0,733612	0,936491	-0,202879	0.6	6	6
0,76144	0,966373	-0,204933	0.6	6	8
0,828609	0,981914	-0,153305	0.6	6	10
0,6701875	0,7934755	-0,123288	0.8	6	2
0,719186	0,8632335	-0,1440475	0.8	6	4
0,7381535	0,914053	-0,1758995	0.8	6	6
0,787722	0,9446895	-0,1569675	0.8	6	8
0,830793	0,963593	-0,1328	0.8	6	10
0,5710785	0,7879705	-0,216892	0.2	7	2
0,7105045	0,859515	-0,1490105	0.2	7	4
0,819175	0,927542	-0,108367	0.2	7	6
0,8131395	0,955814	-0,1426745	0.2	7	8
0,89736	0,953489	-0,056129	0.2	7	10
0,605452	0,75273	-0,147278	0.4	7	2
0,77089	0,820521	-0,049631	0.4	7	4
0,77041	0,874561	-0,104151	0.4	7	6
0,875006	0,9272295	-0,0522235	0.4	7	8
0,891026	0,946251	-0,055225	0.4	7	10
0,6275835	0,882893	-0,2553095	0.6	7	2
0,7772775	0,9322355	-0,154958	0.6	7	4
0,7736625	0,948461	-0,1747985	0.6	7	6
0,876932	0,975656	-0,098724	0.6	7	8
0,891996	0,982849	-0,090853	0.6	7	10
0,63445	0,7917545	-0,1573045	0.8	7	2
0,738394	0,865123	-0,126729	0.8	7	4
0,8285425	0,9194485	-0,090906	0.8	7	6
0,8199135	0,947817	-0,1279035	0.8	7	8
0,8997795	0,9575935	-0,057814	0.8	7	10
0,617371	0,802475	-0,185104	0.2	8	2
0,747449	0,8814515	-0,1340025	0.2	8	4
0,7817705	0,9183385	-0,136568	0.2	8	6
0,8086555	0,9436115	-0,134956	0.2	8	8
0,857757	0,9559465	-0,0981895	0.2	8	10
0,6474755	0,729883	-0,0824075	0.4	8	2
0,7262195	0,819062	-0,0928425	0.4	8	4
0,764556	0,862737	-0,098181	0.4	8	6
0,8366245	0,9015445	-0,06492	0.4	8	8
0,813203	0,920521	-0,107318	0.4	8	10

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,6565445	0,8959675	-0,239423	0.6	8	2
0,730263	0,9354325	-0,2051695	0.6	8	4
0,766852	0,955203	-0,188351	0.6	8	6
0,837801	0,969876	-0,132075	0.6	8	8
0,813932	0,9749785	-0,1610465	0.6	8	10
0,6563275	0,811756	-0,1554285	0.8	8	2
0,763202	0,8827605	-0,1195585	0.8	8	4
0,7895635	0,9097555	-0,120192	0.8	8	6
0,8128	0,945606	-0,132806	0.8	8	8
0,860787	0,9549245	-0,0941375	0.8	8	10
0,586378	0,8053575	-0,2189795	0.2	9	2
0,6714225	0,869089	-0,1976665	0.2	9	4
0,789587	0,9019685	-0,1123815	0.2	9	6
0,8416525	0,9507285	-0,109076	0.2	9	8
0,827629	0,948716	-0,121087	0.2	9	10
0,653942	0,7454875	-0,0915455	0.4	9	2
0,7014215	0,7987375	-0,097316	0.4	9	4
0,871578	0,8903445	-0,0187665	0.4	9	6
0,804677	0,909891	-0,105214	0.4	9	8
0,862708	0,920488	-0,05778	0.4	9	10
0,6598845	0,907044	-0,2471595	0.6	9	2
0,706187	0,9339555	-0,2277685	0.6	9	4
0,872654	0,9666925	-0,0940385	0.6	9	6
0,8049475	0,974778	-0,1698305	0.6	9	8
0,8634295	0,975801	-0,1123715	0.6	9	10
0,6110335	0,810162	-0,1991285	0.8	9	2
0,685296	0,862119	-0,176823	0.8	9	4
0,7950135	0,914552	-0,1195385	0.8	9	6
0,845036	0,9532345	-0,1081985	0.8	9	8
0,8297575	0,9424055	-0,112648	0.8	9	10
0,6814155	0,821564	-0,1401485	0.2	10	2
0,715121	0,8698095	-0,1546885	0.2	10	4
0,8007885	0,9074375	-0,106649	0.2	10	6
0,795365	0,933633	-0,138268	0.2	10	8
0,8912505	0,947994	-0,0567435	0.2	10	10
0,6176725	0,7391255	-0,121453	0.4	10	2
0,726295	0,8083705	-0,0820755	0.4	10	4
0,794296	0,8596775	-0,0653815	0.4	10	6
0,8972315	0,900816	-0,0035845	0.4	10	8
0,8685955	0,932609	-0,0640135	0.4	10	10
0,6242785	0,903217	-0,2789385	0.6	10	2
0,7289245	0,9358085	-0,206884	0.6	10	4
0,795621	0,954595	-0,158974	0.6	10	6
0,898544	0,966436	-0,067892	0.6	10	8
0,869288	0,980103	-0,110815	0.6	10	10
0,698715	0,819734	-0,121019	0.8	10	2
0,7261155	0,8521085	-0,125993	0.8	10	4
0,804363	0,919646	-0,115283	0.8	10	6
0,7966805	0,929628	-0,1329475	0.8	10	8
0,8998505	0,9415675	-0,041717	0.8	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON DOS COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,648296	0,8143105	-0,1660145	0.5	6	2
0,699521	0,8793635	-0,1798425	0.5	6	4
0,7349895	0,927122	-0,1921325	0.5	6	6
0,7446755	0,954677	-0,2100015	0.5	6	8
0,8004015	0,9638505	-0,163449	0.5	6	10
0,6613845	0,794052	-0,1326675	0.5	7	2
0,811675	0,8705815	-0,0589065	0.5	7	4
0,8131755	0,904676	-0,0915005	0.5	7	6
0,843007	0,937844	-0,094837	0.5	7	8
0,886252	0,9748795	-0,0886275	0.5	7	10
0,59149	0,8163135	-0,2248235	0.5	8	2
0,7574855	0,8736715	-0,116186	0.5	8	4
0,754825	0,9331635	-0,1783385	0.5	8	6
0,866326	0,9512965	-0,0849705	0.5	8	8
0,8899765	0,973763	-0,0837865	0.5	8	10
0,6185325	0,8057655	-0,187233	0.5	9	2
0,666408	0,866432	-0,200024	0.5	9	4
0,8147385	0,907395	-0,0926565	0.5	9	6
0,7944825	0,935782	-0,1412995	0.5	9	8
0,8155905	0,946811	-0,1312205	0.5	9	10
0,713656	0,9175775	-0,2039215	0.5	10	2
0,744574	0,9347965	-0,1902225	0.5	10	4
0,7369305	0,951811	-0,2148805	0.5	10	6
0,855362	0,9655775	-0,1102155	0.5	10	8
0,8038365	0,967293	-0,1634565	0.5	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro C.2. Porcentaje de unidades estadísticas correctamente clasificadas por la regla de puntajes lineales utilizando los componentes principales y la regla propuesta con dos componentes principales admisibles

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON TRES COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,9220345	0,8019035	0,120131	0.2	6	2
0,983046	0,868451	0,114595	0.2	6	4
0,995921	0,906553	0,089368	0.2	6	6
0,9995165	0,9599755	0,039541	0.2	6	8
0,999742	0,961796	0,037946	0.2	6	10
0,964673	0,7311895	0,2334835	0.4	6	2
0,993486	0,827332	0,166154	0.4	6	4
0,9960485	0,8704575	0,125591	0.4	6	6
0,9998625	0,9264975	0,073365	0.4	6	8
0,999997	0,9548425	0,0451545	0.4	6	10
0,9805585	0,880436	0,1001225	0.6	6	2
0,996176	0,9314105	0,0647655	0.6	6	4
0,9979165	0,951394	0,0465225	0.6	6	6
0,9999155	0,9745955	0,02532	0.6	6	8
0,9999985	0,9853165	0,014682	0.6	6	10
0,989461	0,7978985	0,1915625	0.8	6	2
0,9975225	0,864823	0,1326995	0.8	6	4
0,9994375	0,911354	0,0880835	0.8	6	6
0,9999055	0,9492975	0,050608	0.8	6	8
0,9999465	0,9646175	0,035329	0.8	6	10
0,974772	0,81601	0,158762	0.2	7	2
0,9971025	0,881438	0,1156645	0.2	7	4
0,9990595	0,916429	0,0826305	0.2	7	6
0,999933	0,9371515	0,0627815	0.2	7	8
0,999999	0,954879	0,04512	0.2	7	10
0,981244	0,726753	0,254491	0.4	7	2
0,9963365	0,818795	0,1775415	0.4	7	4
0,9989925	0,901206	0,0977865	0.4	7	6
0,9999615	0,9287415	0,07122	0.4	7	8
0,9999235	0,9531835	0,04674	0.4	7	10
0,9889815	0,8822615	0,10672	0.6	7	2
0,9977435	0,929363	0,0683805	0.6	7	4
0,999385	0,966496	0,032889	0.6	7	6
0,9999695	0,9748085	0,025161	0.6	7	8
0,999959	0,984486	0,015473	0.6	7	10
0,9960145	0,8108815	0,185133	0.8	7	2
0,999422	0,8776325	0,1217895	0.8	7	4
0,999834	0,9184965	0,0813375	0.8	7	6
0,9999875	0,9375235	0,062464	0.8	7	8
1	0,9519615	0,0480385	0.8	7	10
0,9907975	0,807272	0,1835255	0.2	8	2
0,99676	0,87375	0,12301	0.2	8	4
0,999923	0,9005265	0,0993965	0.2	8	6
0,999996	0,9411655	0,0588305	0.2	8	8
1	0,95169	0,04831	0.2	8	10
0,995956	0,7469175	0,2490385	0.4	8	2
0,9986225	0,81448	0,1841425	0.4	8	4
0,9996305	0,883284	0,1163465	0.4	8	6
1	0,908524	0,091476	0.4	8	8
0,9999995	0,9355825	0,064417	0.4	8	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON TRES COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,997555	0,903322	0,094233	0.6	8	2
0,9991885	0,930842	0,0683465	0.6	8	4
0,999781	0,9576735	0,0421075	0.6	8	6
0,9602455	0,970092	-0,0098465	0.6	8	8
0,972912	0,976716	-0,003804	0.6	8	10
0,9983355	0,815058	0,1832775	0.8	8	2
0,9994565	0,873041	0,1264155	0.8	8	4
0,9999815	0,909187	0,0907945	0.8	8	6
0,9999985	0,943792	0,0562065	0.8	8	8
1	0,956653	0,043347	0.8	8	10
0,9951565	0,8137145	0,181442	0.2	9	2
0,9998445	0,8752145	0,12463	0.2	9	4
0,999991	0,907791	0,0922	0.2	9	6
1	0,9397805	0,0602195	0.2	9	8
1	0,952176	0,047824	0.2	9	10
0,9979125	0,7320575	0,265855	0.4	9	2
0,99991	0,816936	0,182974	0.4	9	4
1	0,851959	0,148041	0.4	9	6
1	0,902941	0,097059	0.4	9	8
0,9999985	0,94009	0,0599085	0.4	9	10
0,9987135	0,8996975	0,099016	0.6	9	2
0,9999365	0,934624	0,0653125	0.6	9	4
1	0,94713	0,05287	0.6	9	6
1	0,967167	0,032833	0.6	9	8
0,999999	0,982425	0,017574	0.6	9	10
0,999201	0,8182885	0,1809125	0.8	9	2
0,999974	0,881289	0,118685	0.8	9	4
0,999997	0,90707	0,092927	0.8	9	6
1	0,9418695	0,0581305	0.8	9	8
1	0,945057	0,054943	0.8	9	10
0,9975365	0,825502	0,1720345	0.2	10	2
0,999998	0,8739165	0,1260815	0.2	10	4
1	0,9074855	0,0925145	0.2	10	6
1	0,9720305	0,0279695	0.2	10	8
1	0,969192	0,030808	0.2	10	10
0,9996655	0,7485065	0,251159	0.4	10	2
0,999983	0,805676	0,194307	0.4	10	4
1	0,884962	0,115038	0.4	10	6
1	0,901835	0,098165	0.4	10	8
1	0,9439635	0,0560365	0.4	10	10
0,9997765	0,9098875	0,089889	0.6	10	2
0,9999845	0,930166	0,0698185	0.6	10	4
1	0,9638835	0,0361165	0.6	10	6
1	0,9700325	0,0299675	0.6	10	8
1	0,9849385	0,0150615	0.6	10	10
0,999552	0,9081135	0,0914385	0.8	10	2
1	0,9356665	0,0643335	0.8	10	4
1	0,954359	0,045641	0.8	10	6
1	0,9720305	0,0279695	0.8	10	8
1	0,969192	0,030808	0.8	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 1 CON TRES COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,9694065	0,862867	0,1065395	0.5	6	2
0,991887	0,9344935	0,0573935	0.5	6	4
0,9993955	0,981184	0,0182115	0.5	6	6
0,9999755	0,9904885	0,009487	0.5	6	8
1	0,989276	0,010724	0.5	6	10
0,985851	0,870169	0,115682	0.5	7	2
0,999646	0,9384375	0,0612085	0.5	7	4
0,999998	0,9636985	0,0362995	0.5	7	6
0,999988	0,983267	0,016721	0.5	7	8
1	0,982838	0,017162	0.5	7	10
0,9935635	0,8865275	0,107036	0.5	8	2
0,999393	0,927468	0,071925	0.5	8	4
0,999998	0,9616595	0,0383385	0.5	8	6
1	0,9707995	0,0292005	0.5	8	8
1	0,986965	0,013035	0.5	8	10
0,995851	0,8869515	0,1088995	0.5	9	2
0,99998	0,9295155	0,0704645	0.5	9	4
1	0,9685945	0,0314055	0.5	9	6
1	0,985963	0,014037	0.5	9	8
1	0,986598	0,013402	0.5	9	10
0,9994865	0,877739	0,1217475	0.5	10	2
0,999808	0,95315	0,046658	0.5	10	4
0,9999975	0,962562	0,0374355	0.5	10	6
1	0,987402	0,012598	0.5	10	8
1	0,978665	0,021335	0.5	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro C.3. Clasificación correcta de unidades estadísticas por la regla de puntajes lineales y la regla propuesta con tres componentes principales admisibles

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON TRES COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,5478805	0,864341	-0,3164605	0.2	6	2
0,6447125	0,922742	-0,2780295	0.2	6	4
0,743416	0,956462	-0,213046	0.2	6	6
0,8330375	0,9897785	-0,156741	0.2	6	8
0,8257675	0,9899155	-0,164148	0.2	6	10
0,5926765	0,7823375	-0,189661	0.4	6	2
0,71601	0,8791645	-0,1631545	0.4	6	4
0,7516945	0,9454325	-0,193738	0.4	6	6
0,8296135	0,9676725	-0,138059	0.4	6	8
0,8320275	0,9811815	-0,149154	0.4	6	10
0,6179745	0,9372535	-0,319279	0.6	6	2
0,7301435	0,973724	-0,2435805	0.6	6	4
0,7610565	0,991739	-0,2306825	0.6	6	6
0,832794	0,9945895	-0,1617955	0.6	6	8
0,834248	0,9966215	-0,1623735	0.6	6	10
0,631776	0,871906	-0,24013	0.8	6	2
0,694187	0,931389	-0,237202	0.8	6	4
0,771879	0,964864	-0,192985	0.8	6	6
0,840333	0,9878645	-0,1475315	0.8	6	8
0,836279	0,980577	-0,144298	0.8	6	10
0,6116515	0,864062	-0,2524105	0.2	7	2
0,687599	0,930951	-0,243352	0.2	7	4
0,806981	0,9376405	-0,1306595	0.2	7	6
0,855217	0,978726	-0,123509	0.2	7	8
0,9236835	0,987713	-0,0640295	0.2	7	10
0,605476	0,798944	-0,193468	0.4	7	2
0,77496	0,9038265	-0,1288665	0.4	7	4
0,8337645	0,95225	-0,1184855	0.4	7	6
0,852729	0,961938	-0,109209	0.4	7	8
0,8920005	0,9868335	-0,094833	0.4	7	10
0,6215265	0,873717	-0,2521905	0.6	7	2
0,782808	0,9400895	-0,1572815	0.6	7	4
0,836019	0,973353	-0,137334	0.6	7	6
0,8548625	0,9705795	-0,115717	0.6	7	8
0,893732	0,9906575	-0,0969255	0.6	7	10
0,679294	0,8730135	-0,1937195	0.8	7	2
0,726205	0,9450875	-0,2188825	0.8	7	4
0,818844	0,9516305	-0,1327865	0.8	7	6
0,864767	0,9726745	-0,1079075	0.8	7	8
0,9276685	0,9899465	-0,062278	0.8	7	10
0,5991565	0,877368	-0,2782115	0.2	8	2
0,732752	0,934993	-0,202241	0.2	8	4
0,8204435	0,9678615	-0,147418	0.2	8	6
0,8699575	0,9748725	-0,104915	0.2	8	8
0,825253	0,9868285	-0,1615755	0.2	8	10
0,643997	0,7839805	-0,1399835	0.4	8	2
0,657582	0,9035005	-0,2459185	0.4	8	4
0,7594285	0,9516065	-0,192178	0.4	8	6
0,806737	0,9591855	-0,1524485	0.4	8	8
0,8465105	0,9554205	-0,10891	0.4	8	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON TRES COMPONENTES

Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,6566335	0,950939	-0,2943055	0.6	8	2
0,661255	0,9814985	-0,3202435	0.6	8	4
0,761167	0,992205	-0,231038	0.6	8	6
0,8085145	0,9914835	-0,182969	0.6	8	8
0,8470875	0,9929625	-0,145875	0.6	8	10
0,636057	0,868223	-0,232166	0.8	8	2
0,7478665	0,9282265	-0,18036	0.8	8	4
0,8264975	0,9636335	-0,137136	0.8	8	6
0,875527	0,9673575	-0,0918305	0.8	8	8
0,828544	0,981557	-0,153013	0.8	8	10
0,5683975	0,8846165	-0,316219	0.2	9	2
0,762318	0,945079	-0,182761	0.2	9	4
0,7979105	0,970047	-0,1721365	0.2	9	6
0,8109605	0,9777225	-0,166762	0.2	9	8
0,868779	0,978986	-0,110207	0.2	9	10
0,6213605	0,807197	-0,1858365	0.4	9	2
0,726226	0,8833835	-0,1571575	0.4	9	4
0,8673865	0,9575805	-0,090194	0.4	9	6
0,8344365	0,9578315	-0,123395	0.4	9	8
0,8473025	0,980722	-0,1334195	0.4	9	10
0,6316475	0,95741	-0,3257625	0.6	9	2
0,7309175	0,9788655	-0,247948	0.6	9	4
0,8688915	0,9947205	-0,125829	0.6	9	6
0,8350725	0,994435	-0,1593625	0.6	9	8
0,8476675	0,9976505	-0,149983	0.6	9	10
0,598057	0,87989	-0,281833	0.8	9	2
0,772404	0,9363745	-0,1639705	0.8	9	4
0,802271	0,9663705	-0,1640995	0.8	9	6
0,814037	0,974681	-0,160644	0.8	9	8
0,8703135	0,9899305	-0,119617	0.8	9	10
0,5796655	0,876873	-0,2972075	0.2	10	2
0,7563515	0,9393775	-0,183026	0.2	10	4
0,808772	0,962726	-0,153954	0.2	10	6
0,8454805	0,9862285	-0,140748	0.2	10	8
0,8807945	0,9858665	-0,105072	0.2	10	10
0,599257	0,78118	-0,181923	0.4	10	2
0,7315825	0,878719	-0,1471365	0.4	10	4
0,8189645	0,926763	-0,1077985	0.4	10	6
0,860805	0,9563165	-0,0955115	0.4	10	8
0,837058	0,962141	-0,125083	0.4	10	10
0,6045165	0,951763	-0,3472465	0.6	10	2
0,7354775	0,9782895	-0,242812	0.6	10	4
0,820365	0,987605	-0,16724	0.6	10	6
0,8611145	0,993593	-0,1324785	0.6	10	8
0,8372225	0,9940055	-0,156783	0.6	10	10
0,595953	0,8758115	-0,2798585	0.8	10	2
0,7638055	0,934619	-0,1708135	0.8	10	4
0,8143565	0,9537875	-0,139431	0.8	10	6
0,848029	0,983251	-0,135222	0.8	10	8
0,8824815	0,9854005	-0,102919	0.8	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 2 CON TRES COMPONENTES					
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia
0,62408	0,85764	-0,23356	0.5	6	2
0,6786865	0,928698	-0,2500115	0.5	6	4
0,770842	0,973345	-0,202503	0.5	6	6
0,836819	0,9792505	-0,1424315	0.5	6	8
0,827839	0,9883635	-0,1605245	0.5	6	10
0,563104	0,859743	-0,296639	0.5	7	2
0,695052	0,923385	-0,228333	0.5	7	4
0,787612	0,9846715	-0,1970595	0.5	7	6
0,8195115	0,977689	-0,1581775	0.5	7	8
0,866049	0,9908695	-0,1248205	0.5	7	10
0,6448835	0,8781265	-0,233243	0.5	8	2
0,739285	0,9542295	-0,2149445	0.5	8	4
0,7157295	0,959816	-0,2440865	0.5	8	6
0,8421885	0,9719625	-0,129774	0.5	8	8
0,814082	0,9817965	-0,1677145	0.5	8	10
0,65919	0,8675045	-0,2083145	0.5	9	2
0,669636	0,945602	-0,275966	0.5	9	4
0,785148	0,958612	-0,173464	0.5	9	6
0,840562	0,9761205	-0,1355585	0.5	9	8
0,847548	0,977841	-0,130293	0.5	9	10
0,646764	0,892052	-0,245288	0.5	10	2
0,742319	0,9249585	-0,1826395	0.5	10	4
0,8225265	0,9698385	-0,147312	0.5	10	6
0,849511	0,982644	-0,133133	0.5	10	8
0,9116335	0,9906665	-0,079033	0.5	10	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro C.4. Porcentaje de unidades estadísticas correctamente clasificadas por la regla de puntajes lineales utilizando los componentes principales y la regla propuesta con tres componentes principales admisibles

Apéndice D

Resultados de la simulación de la clasificación de unidades estadísticas por la regla propuesta y la regla de puntajes lineales discriminante

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9191565	0,9258025	-0,006646	0.2	6	2	25
0,983866	0,9838315	0,0000345	0.2	6	4	25
0,99688	0,996034	0,000846	0.2	6	6	25
0,9991735	0,9986895	0,000484	0.2	6	8	25
0,999943	0,999739	0,000204	0.2	6	10	25
0,9418225	0,947237	-0,0054145	0.3	6	2	25
0,9877295	0,985496	0,0022335	0.3	6	4	25
0,9982245	0,9976235	0,000601	0.3	6	6	25
0,999346	0,9992545	0,0000915	0.3	6	8	25
0,9997515	0,999852	-0,0001005	0.3	6	10	25
0,955906	0,9285825	0,0273235	0.4	6	2	25
0,991085	0,982152	0,008933	0.4	6	4	25
0,9982765	0,9958595	0,002417	0.4	6	6	25
0,9992875	0,998503	0,0007845	0.4	6	8	25
0,9996805	0,999315	0,0003655	0.4	6	10	25
0,96997	0,929305	0,040665	0.5	6	2	25
0,9935915	0,982008	0,0115835	0.5	6	4	25
0,9982245	0,995232	0,0029925	0.5	6	6	25
0,9994825	0,99904	0,0004425	0.5	6	8	25
0,9999155	0,999846	0,0000695	0.5	6	10	25
0,9755805	0,998333	-0,0227525	0.6	6	2	25
0,9831385	0,9258145	0,057324	0.7	6	2	25
0,996146	0,981243	0,014903	0.7	6	4	25
0,9993575	0,9959395	0,003418	0.7	6	6	25
0,999774	0,99858	0,001194	0.7	6	8	25
0,999924	0,9995755	0,0003485	0.7	6	10	25
0,989525	0,9260875	0,0634375	0.8	6	2	25
0,997478	0,980094	0,017384	0.8	6	4	25
0,9995085	0,9955545	0,003954	0.8	6	6	25
0,9998575	0,9987325	0,001125	0.8	6	8	25
0,9999855	0,999896	0,0000895	0.8	6	10	25
0,940585	0,9486625	-0,0080775	0.2	7	2	25
0,988545	0,9886525	-0,0001075	0.2	7	4	25
0,998466	0,9981475	0,0003185	0.2	7	6	25
0,999379	0,999758	-0,000379	0.2	7	8	25
0,999975	0,9999375	0,0000375	0.2	7	10	25
0,958221	0,950415	0,007806	0.3	7	2	25
0,993584	0,990923	0,002661	0.3	7	4	25
0,998614	0,999067	-0,000453	0.3	7	6	25
0,9994215	0,9996445	-0,000223	0.3	7	8	25
0,999883	0,9998995	-0,0000165	0.3	7	10	25
0,9669195	0,949452	0,0174675	0.4	7	2	25
0,994077	0,9895505	0,0045265	0.4	7	4	25
0,9986145	0,997725	0,0008895	0.4	7	6	25
0,999729	0,999382	0,000347	0.4	7	8	25
0,999827	0,9996995	0,0001275	0.4	7	10	25
0,952528	0,958048	-0,00552	0.5	7	2	25
0,9942245	0,988218	0,0060065	0.5	7	4	25
0,9994465	0,998644	0,0008025	0.5	7	6	25

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9997875	0,999436	0,0003515	0.5	7	8	25
0,999993	0,9999935	-0,0000005	0.5	7	10	25
0,9814645	0,9995625	-0,018098	0.6	7	2	25
0,996551	0,9999475	-0,0033965	0.6	7	4	25
0,999184	0,999996	-0,000812	0.6	7	6	25
0,9998385	0,9999995	-0,000161	0.6	7	8	25
0,99989	1	-0,00011	0.6	7	10	25
0,987226	0,950423	0,036803	0.7	7	2	25
0,9977975	0,990381	0,0074165	0.7	7	4	25
0,999514	0,997901	0,001613	0.7	7	6	25
0,9998105	0,999231	0,0005795	0.7	7	8	25
0,9999665	0,99983	0,0001365	0.7	7	10	25
0,991396	0,948485	0,042911	0.8	7	2	25
0,9980885	0,9886225	0,009466	0.8	7	4	25
0,999703	0,9981135	0,0015895	0.8	7	6	25
0,999892	0,9994485	0,0004435	0.8	7	8	25
0,999994	0,9999795	0,0000145	0.8	7	10	25
0,962715	0,9613355	0,0013795	0.2	8	2	25
0,994615	0,993805	0,00081	0.2	8	4	25
0,9961105	0,9963555	-0,000245	0.3	8	4	25
0,999031	0,9993985	-0,0003675	0.3	8	6	25
0,9999075	0,9999135	-0,000006	0.3	8	8	25
0,999988	0,9999765	0,0000115	0.3	8	10	25
0,980866	0,968167	0,012699	0.4	8	2	25
0,9970095	0,993137	0,0038725	0.4	8	4	25
0,999551	0,9988205	0,0007305	0.4	8	6	25
0,999956	0,999813	0,000143	0.4	8	8	25
0,9999775	0,9999715	0,000006	0.4	8	10	25
0,986106	0,9769	0,009206	0.5	8	2	25
0,997922	0,9932715	0,0046505	0.5	8	4	25
0,999851	0,9992395	0,0006115	0.5	8	6	25
0,9998225	0,999572	0,0002505	0.5	8	8	25
0,999995	0,999934	0,000061	0.5	8	10	25
0,988955	0,9999155	-0,0109605	0.6	8	2	25
0,998123	0,9999955	-0,0018725	0.6	8	4	25
0,9997525	0,999999	-0,0002465	0.6	8	6	25
0,999977	1	-0,000023	0.6	8	8	25
0,9999955	1	-0,0000045	0.6	8	10	25
0,9918825	0,9642365	0,027646	0.7	8	2	25
0,9987125	0,993006	0,0057065	0.7	8	4	25
0,9996705	0,998416	0,0012545	0.7	8	6	25
0,999959	0,999786	0,000173	0.7	8	8	25
0,999995	0,999955	0,00004	0.7	8	10	25
0,9941195	0,9613425	0,032777	0.8	8	2	25
0,999024	0,994618	0,004406	0.8	8	4	25
0,999785	0,998547	0,001238	0.8	8	6	25
0,999958	0,999777	0,000181	0.8	8	8	25
0,999991	0,9999515	0,0000395	0.8	8	10	25
0,980594	0,981344	-0,00075	0.2	9	2	25

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9974135	0,9960725	0,001341	0.2	9	4	25
0,9996655	0,999269	0,0003965	0.2	9	6	25
0,9999795	0,9999165	0,000063	0.2	9	8	25
0,999989	0,9999845	0,0000045	0.2	9	10	25
0,9837945	0,9811875	0,002607	0.3	9	2	25
0,9979535	0,993205	0,0047485	0.3	9	4	25
0,999656	0,999638	0,000018	0.3	9	6	25
0,999965	0,9999155	0,0000495	0.3	9	8	25
0,999997	0,9999975	-0,0000005	0.3	9	10	25
0,986962	0,9836105	0,0033515	0.4	9	2	25
0,9981855	0,9954645	0,002721	0.4	9	4	25
0,999792	0,9993785	0,0004135	0.4	9	6	25
0,999979	0,9998885	0,0000905	0.4	9	8	25
0,9999885	0,9999725	0,000016	0.4	9	10	25
0,9909515	0,9862375	0,004714	0.5	9	2	25
0,9924085	0,999988	-0,0075795	0.6	9	2	25
0,9989035	0,999999	-0,0010955	0.6	9	4	25
0,999865	0,9999995	-0,0001345	0.6	9	6	25
0,9999815	1	-0,0000185	0.6	9	8	25
1	1	0	0.6	9	10	25
0,9946805	0,9735015	0,021179	0.7	9	2	25
0,9992505	0,995513	0,0037375	0.7	9	4	25
0,999877	0,999328	0,000549	0.7	9	6	25
0,999984	0,9998735	0,0001105	0.7	9	8	25
0,999999	0,999996	0,000003	0.7	9	10	25
0,996838	0,986289	0,010549	0.8	9	2	25
0,9995475	0,9961535	0,003394	0.8	9	4	25
0,9999415	0,9993225	0,000619	0.8	9	6	25
0,99999	0,9999555	0,0000345	0.8	9	8	25
0,999999	0,999972	0,000027	0.8	9	10	25
0,9892945	0,983568	0,0057265	0.2	10	2	25
0,998867	0,9976925	0,0011745	0.2	10	4	25
0,999904	0,9997055	0,0001985	0.2	10	6	25
0,9999965	0,999979	0,0000175	0.2	10	8	25
0,9999985	0,999995	0,0000035	0.2	10	10	25
0,991801	0,9881835	0,0036175	0.3	10	2	25
0,9991925	0,998514	0,0006785	0.3	10	4	25
0,9999345	0,999872	0,0000625	0.3	10	6	25
0,999997	0,9999905	0,0000065	0.3	10	8	25
0,9999995	0,9999975	0,000002	0.3	10	10	25
0,9935365	0,98854	0,0049965	0.4	10	2	25
0,9993905	0,997657	0,0017335	0.4	10	4	25
0,999951	0,9998205	0,0001305	0.4	10	6	25
0,9999965	0,9999765	0,00002	0.4	10	8	25
1	0,9999975	0,0000025	0.4	10	10	25
0,9944685	0,9930245	0,001444	0.5	10	2	25
0,999044	0,99717	0,001874	0.5	10	4	25
0,9998775	0,999605	0,0002725	0.5	10	6	25
0,999999	0,999984	0,000015	0.5	10	8	25

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9999825	0,999995	-0,0000125	0.5	10	10	25
0,9960685	0,999996	-0,0039275	0.6	10	2	25
0,9996515	0,999999	-0,0003475	0.6	10	4	25
0,9999665	0,999999	-0,0000325	0.6	10	6	25
0,999995	1	-0,000005	0.6	10	8	25
1	1	0	0.6	10	10	25
0,9999825	0,999995	-0,0000125	0.5	10	10	25
0,9971925	0,9834755	0,013717	0.7	10	2	25
0,9997385	0,997746	0,0019925	0.7	10	4	25
0,9999715	0,9997185	0,000253	0.7	10	6	25
0,999999	0,9999835	0,0000155	0.7	10	8	25
0,999999	0,999985	0,000014	0.8	10	8	25
1	0,9999995	0,0000005	0.8	10	10	25
0,929797	0,926443	0,003354	0.2	6	2	50
0,9849125	0,9807455	0,004167	0.2	6	4	50
0,9964875	0,995548	0,0009395	0.2	6	6	50
0,99937	0,9988365	0,0005335	0.2	6	8	50
0,9997235	0,999631	0,0000925	0.2	6	10	50
0,941004	0,9620075	-0,0210035	0.3	6	2	50
0,9874805	0,9905105	-0,00303	0.3	6	4	50
0,997838	0,997921	-0,000083	0.3	6	6	50
0,9991525	0,9993845	-0,000232	0.3	6	8	50
0,999839	0,9997745	0,0000645	0.3	6	10	50
0,958045	0,9280785	0,0299665	0.4	6	2	50
0,991	0,981302	0,009698	0.4	6	4	50
0,997617	0,9947205	0,0028965	0.4	6	6	50
0,9992625	0,9983565	0,000906	0.4	6	8	50
0,999549	0,9992505	0,0002985	0.4	6	10	50
0,9672945	0,9252115	0,042083	0.5	6	2	50
0,993087	0,9807995	0,0122875	0.5	6	4	50
0,9988135	0,99537	0,0034435	0.5	6	6	50
0,9997175	0,9988095	0,000908	0.5	6	8	50
0,999944	0,9998425	0,0001015	0.5	6	10	50
0,976807	0,9984595	-0,0216525	0.6	6	2	50
0,9948145	0,9997655	-0,004951	0.6	6	4	50
0,9986345	0,9999425	-0,001308	0.6	6	6	50
0,999582	0,9999755	-0,0003935	0.6	6	8	50
0,999749	0,9999915	-0,0002425	0.6	6	10	50
0,9536155	0,9302215	0,023394	0.7	6	2	50
0,9882915	0,980797	0,0074945	0.7	6	4	50
0,9971755	0,995399	0,0017765	0.7	6	6	50
0,9990225	0,998561	0,0004615	0.7	6	8	50
0,9997945	0,9996115	0,000183	0.7	6	10	50
0,990237	0,927178	0,063059	0.8	6	2	50
0,997712	0,9806985	0,0170135	0.8	6	4	50
0,9994575	0,995485	0,0039725	0.8	6	6	50
0,999883	0,998804	0,001079	0.8	6	8	50
0,999959	0,9996295	0,0003295	0.8	6	10	50
0,9577595	0,953766	0,0039935	0.2	7	2	50
0,992362	0,989611	0,002751	0.2	7	4	50

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,99889	0,9982585	0,0006315	0.2	7	6	50
0,999481	0,999319	0,000162	0.2	7	8	50
0,999952	0,999892	0,00006	0.2	7	10	50
0,9670705	0,9709805	-0,00391	0.3	7	2	50
0,994267	0,994577	-0,00031	0.3	7	4	50
0,9988945	0,9988995	-0,000005	0.3	7	6	50
0,999836	0,999616	0,00022	0.3	7	8	50
0,999905	0,9998905	0,0000145	0.3	7	10	50
0,976417	0,9529665	0,0234505	0.4	7	2	50
0,995584	0,989642	0,005942	0.4	7	4	50
0,999239	0,998218	0,001021	0.4	7	6	50
0,9999325	0,999732	0,0002005	0.4	7	8	50
0,999908	0,999484	0,000424	0.5	7	8	50
0,9999635	0,9998355	0,000128	0.5	7	10	50
0,9863905	0,9996425	-0,013252	0.6	7	2	50
0,997412	0,999951	-0,002539	0.6	7	4	50
0,9995505	0,999992	-0,0004415	0.6	7	6	50
0,999957	0,999999	-0,000042	0.6	7	8	50
0,9999905	0,9999995	-0,000009	0.6	7	10	50
0,98979	0,9584735	0,0313165	0.7	7	2	50
0,998072	0,9901385	0,0079335	0.7	7	4	50
0,9996405	0,998014	0,0016265	0.7	7	6	50
0,9999345	0,999429	0,0005055	0.7	7	8	50
0,9999615	0,9998335	0,000128	0.7	7	10	50
0,9936315	0,953716	0,0399155	0.8	7	2	50
0,9987415	0,9896125	0,009129	0.8	7	4	50
0,999795	0,998269	0,001526	0.8	7	6	50
0,999936	0,999384	0,000552	0.8	7	8	50
0,9999895	0,999884	0,0001055	0.8	7	10	50
0,9747915	0,965006	0,0097855	0.2	8	2	50
0,9971805	0,993775	0,0034055	0.2	8	4	50
0,9997195	0,998995	0,0007245	0.2	8	6	50
0,9999195	0,9997395	0,00018	0.2	8	8	50
0,9999995	0,9999885	0,000011	0.2	8	10	50
0,983208	0,978599	0,004609	0.3	8	2	50
0,9973045	0,9965225	0,000782	0.3	8	4	50
0,99945	0,9992055	0,0002445	0.3	8	6	50
0,999887	0,999987	-0,0001	0.3	8	8	50
0,999998	0,9999985	-0,0000005	0.3	8	10	50
0,9881215	0,9679205	0,020201	0.4	8	2	50
0,9984215	0,993994	0,0044275	0.4	8	4	50
0,9997995	0,999027	0,0007725	0.4	8	6	50
0,9999625	0,999798	0,0001645	0.4	8	8	50
0,9999845	0,9999595	0,000025	0.4	8	10	50
0,989527	0,965841	0,023686	0.5	8	2	50
0,9986995	0,9936555	0,005044	0.5	8	4	50
0,9997045	0,998683	0,0010215	0.5	8	6	50
0,99998	0,99975	0,00023	0.5	8	8	50
0,999975	0,999949	0,000026	0.5	8	10	50

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9930785	0,9999195	-0,006841	0.6	8	2	50
0,999008	0,999991	-0,000983	0.6	8	4	50
0,999881	0,9999975	-0,0001165	0.6	8	6	50
0,9999805	0,9999995	-0,000019	0.6	8	8	50
0,9999925	1	-0,0000075	0.6	8	10	50
0,9945935	0,971895	0,0226985	0.7	8	2	50
0,9991165	0,993348	0,0057685	0.7	8	4	50
0,999818	0,998658	0,00116	0.7	8	6	50
0,9999605	0,9997525	0,000208	0.7	8	8	50
0,9999995	0,999964	0,0000355	0.7	8	10	50
0,996172	0,965278	0,030894	0.8	8	2	50
0,989566	0,97865	0,010916	0.2	9	2	50
0,999149	0,9969055	0,0022435	0.2	9	4	50
0,9997855	0,99944	0,0003455	0.2	9	6	50
0,9999955	0,9999545	0,000041	0.2	9	8	50
0,9999995	0,999999	0,0000005	0.2	9	10	50
0,991875	0,988788	0,003087	0.3	9	2	50
0,99907	0,9980405	0,0010295	0.3	9	4	50
0,9999325	0,9997135	0,000219	0.3	9	6	50
0,9999685	0,99993	0,0000385	0.3	9	8	50
0,999999	0,999998	0,000001	0.3	9	10	50
0,99519	0,9798965	0,0152935	0.4	9	2	50
0,999274	0,996342	0,002932	0.4	9	4	50
0,999906	0,999411	0,000495	0.4	9	6	50
0,999975	0,9999365	0,0000385	0.4	9	8	50
0,999998	0,9999765	0,0000215	0.4	9	10	50
0,994889	0,9782415	0,0166475	0.5	9	2	50
0,999603	0,996759	0,002844	0.5	9	4	50
0,9999735	0,9995575	0,000416	0.5	9	6	50
0,999954	0,999964	-0,00001	0.5	9	8	50
0,9999995	0,999999	0,0000005	0.5	9	10	50
0,997201	0,999982	-0,002781	0.6	9	2	50
0,9995775	0,999999	-0,0004215	0.6	9	4	50
0,99993	0,999999	-0,000069	0.6	9	6	50
0,999984	0,9999995	-0,0000155	0.6	9	8	50
0,9999985	1	-0,0000015	0.6	9	10	50
0,9973185	0,977527	0,0197915	0.7	9	2	50
0,9996465	0,996576	0,0030705	0.7	9	4	50
0,999982	0,999432	0,00055	0.7	9	6	50
0,999988	0,9999055	0,0000825	0.7	9	8	50
1	0,9999945	0,0000055	0.7	9	10	50
0,9982975	0,9785805	0,019717	0.8	9	2	50
0,9998475	0,9969425	0,002905	0.8	9	4	50
0,9999515	0,9994275	0,000524	0.8	9	6	50
0,9999995	0,999946	0,0000535	0.8	9	8	50
1	0,9999995	0,0000005	0.8	9	10	50
0,997206	0,985443	0,011763	0.2	10	2	50
0,9997645	0,9982995	0,001465	0.2	10	4	50
0,999987	0,999929	0,000058	0.2	10	6	50

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,999995	0,999993	0,000002	0.2	10	8	50
0,9999995	0,9999985	0,000001	0.2	10	10	50
0,9967535	0,9870025	0,009751	0.3	10	2	50
0,999632	0,9986995	0,0009325	0.3	10	4	50
0,999936	0,999824	0,000112	0.3	10	6	50
0,9999975	0,999984	0,0000135	0.3	10	8	50
0,9999995	0,9999955	0,000004	0.3	10	10	50
0,9976695	0,9860285	0,011641	0.4	10	2	50
0,999822	0,998305	0,001517	0.4	10	4	50
0,9999905	0,9998455	0,000145	0.4	10	6	50
0,999995	0,9999875	0,0000075	0.4	10	8	50
0,9999975	0,9999535	0,000044	0.5	10	8	50
1	0,9999995	0,0000005	0.5	10	10	50
0,998598	0,999997	-0,001399	0.6	10	2	50
0,999888	0,9999995	-0,0001115	0.6	10	4	50
0,9999905	1	-0,0000095	0.6	10	6	50
0,9999995	1	-0,0000005	0.6	10	8	50
1	1	0	0.6	10	10	50
0,998879	0,9870725	0,0118065	0.7	10	2	50
0,999877	0,998008	0,001869	0.7	10	4	50
0,9999775	0,9996705	0,000307	0.7	10	6	50
0,9999995	0,999962	0,0000375	0.7	10	8	50
0,999999	0,9999935	0,0000055	0.7	10	10	50
0,9978235	0,9854755	0,012348	0.8	10	2	50
0,999842	0,998217	0,001625	0.8	10	4	50
0,999992	0,999922	0,00007	0.8	10	6	50
0,999999	0,9999925	0,0000065	0.8	10	8	50
1	1	0	0.8	10	10	50
0,936006	0,9287595	0,0072465	0.2	6	2	75
0,987319	0,9816855	0,0056335	0.2	6	4	75
0,997566	0,995907	0,001659	0.2	6	6	75
0,9994455	0,998914	0,0005315	0.2	6	8	75
0,999805	0,9995235	0,0002815	0.2	6	10	75
0,9521555	0,9642265	-0,012071	0.3	6	2	75
0,990108	0,9907295	-0,0006215	0.3	6	4	75
0,997871	0,9978375	0,0000335	0.3	6	6	75
0,9995415	0,999376	0,0001655	0.3	6	8	75
0,999768	0,999708	0,00006	0.3	6	10	75
0,9606965	0,925263	0,0354335	0.4	6	2	75
0,992455	0,982457	0,009998	0.4	6	4	75
0,998687	0,9956475	0,0030395	0.4	6	6	75
0,999596	0,9987235	0,0008725	0.4	6	8	75
0,999762	0,9996305	0,0001315	0.4	6	10	75
0,973299	0,9311565	0,0421425	0.5	6	2	75
0,993219	0,979931	0,013288	0.5	6	4	75
0,9980965	0,9942855	0,003811	0.5	6	6	75
0,999481	0,9982315	0,0012495	0.5	6	8	75
0,9998805	0,9997585	0,000122	0.5	6	10	75
0,978235	0,9981625	-0,0199275	0.6	6	2	75

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,995554	0,9997885	-0,0042345	0.6	6	4	75
0,9992005	0,9999815	-0,000781	0.6	6	6	75
0,9997465	0,999977	-0,0002305	0.6	6	8	75
0,999862	1	-0,000138	0.6	6	10	75
0,985859	0,9293375	0,0565215	0.7	6	2	75
0,9968105	0,9818385	0,014972	0.7	6	4	75
0,999286	0,995106	0,00418	0.7	6	6	75
0,9998585	0,9986825	0,001176	0.7	6	8	75
0,999925	0,999415	0,00051	0.7	6	10	75
0,991104	0,928499	0,062605	0.8	6	2	75
0,9979385	0,9816005	0,016338	0.8	6	4	75
0,9996095	0,9958655	0,003744	0.8	6	6	75
0,9998995	0,998876	0,0010235	0.8	6	8	75
0,999956	0,999742	0,000214	0.8	6	10	75
0,972199	0,955997	0,016202	0.2	7	2	75
0,9963825	0,9920905	0,004292	0.2	7	4	75
0,9987805	0,9977285	0,001052	0.2	7	6	75
0,999902	0,99973	0,000172	0.2	7	8	75
0,999977	0,9999515	0,0000255	0.2	7	10	75
0,999964	0,9999	0,000064	0.3	7	10	75
0,9847855	0,95752	0,0272655	0.4	7	2	75
0,9974165	0,991362	0,0060545	0.4	7	4	75
0,9996055	0,998042	0,0015635	0.4	7	6	75
0,9999345	0,999616	0,0003185	0.4	7	8	75
0,9999745	0,999912	0,0000625	0.4	7	10	75
0,986712	0,9543985	0,0323135	0.5	7	2	75
0,9976935	0,9904765	0,007217	0.5	7	4	75
0,9997475	0,9984005	0,001347	0.5	7	6	75
0,9999565	0,999623	0,0003335	0.5	7	8	75
0,9999965	0,9999955	0,000001	0.5	7	10	75
0,991177	0,9995095	-0,0083325	0.6	7	2	75
0,998436	0,9999655	-0,0015295	0.6	7	4	75
0,9997965	0,9999945	-0,000198	0.6	7	6	75
0,9999525	1	-0,0000475	0.6	7	8	75
0,999983	1	-0,000017	0.6	7	10	75
0,993068	0,95555	0,037518	0.7	7	2	75
0,998883	0,990764	0,008119	0.7	7	4	75
0,9998305	0,9983755	0,001455	0.7	7	6	75
0,999992	0,999741	0,000251	0.7	7	8	75
0,99999	0,999833	0,000157	0.7	7	10	75
0,995626	0,9559495	0,0396765	0.8	7	2	75
0,9993715	0,992009	0,0073625	0.8	7	4	75
0,9997815	0,997719	0,0020625	0.8	7	6	75
0,9999735	0,9997135	0,00026	0.8	7	8	75
0,999996	0,9999005	0,0000955	0.8	7	10	75
0,989106	0,9688255	0,0202805	0.2	8	2	75
0,998677	0,993987	0,00469	0.2	8	4	75
0,999849	0,999049	0,0008	0.2	8	6	75
0,9999775	0,99974	0,0002375	0.2	8	8	75

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9999955	0,9999315	0,000064	0.2	8	10	75
0,9913395	0,9847665	0,006573	0.3	8	2	75
0,9989775	0,9972065	0,001771	0.3	8	4	75
0,999868	0,9995325	0,0003355	0.3	8	6	75
0,9999895	0,9999135	0,000076	0.3	8	8	75
0,999997	0,999929	0,000068	0.3	8	10	75
0,9942325	0,9725815	0,021651	0.4	8	2	75
0,9990835	0,994119	0,0049645	0.4	8	4	75
0,9999405	0,999137	0,0008035	0.4	8	6	25
0,999979	0,999827	0,000152	0.4	8	8	75
0,9999995	0,9999935	0,000006	0.4	8	10	75
0,9946055	0,966197	0,0284085	0.5	8	2	75
0,999396	0,9944195	0,0049765	0.5	8	4	75
0,999713	0,9980635	0,0016495	0.5	8	6	75
0,999991	0,9999305	0,0000605	0.5	8	8	75
0,999991	1	-0,000009	0.6	8	8	75
1	1	0	0.6	8	10	75
0,9971465	0,9695925	0,027554	0.7	8	2	75
0,99962	0,9946165	0,0050035	0.7	8	4	75
0,999956	0,99911	0,000846	0.7	8	6	75
0,999996	0,9998255	0,0001705	0.7	8	8	75
0,999999	0,999909	0,00009	0.7	8	10	75
0,998137	0,9690105	0,0291265	0.8	8	2	75
0,9997265	0,993935	0,0057915	0.8	8	4	75
0,999964	0,999035	0,000929	0.8	8	6	75
0,999996	0,999741	0,000255	0.8	8	8	25
0,999999	0,999993	0,000006	0.8	8	10	75
0,996031	0,9819755	0,0140555	0.2	9	2	75
0,9998425	0,9969515	0,002891	0.2	9	4	75
0,9999435	0,9994915	0,000452	0.2	9	6	75
0,999999	0,999944	0,000055	0.2	9	8	75
1	0,999996	0,000004	0.2	9	10	75
0,9973355	0,99094	0,0063955	0.3	9	2	75
0,999767	0,998342	0,001425	0.3	9	4	75
0,9999595	0,999761	0,0001985	0.3	9	6	75
0,9999975	0,9999565	0,000041	0.3	9	8	75
1	0,999998	0,000002	0.3	9	10	75
0,9978035	0,982187	0,0156165	0.4	9	2	75
0,9997765	0,99661	0,0031665	0.4	9	4	75
0,999984	0,9994745	0,0005095	0.4	9	6	75
0,9999995	0,999951	0,0000485	0.4	9	8	75
0,9999785	0,999997	-0,0000185	0.4	9	10	75
0,9979725	0,979064	0,0189085	0.5	9	2	75
0,999928	0,9971715	0,0027565	0.5	9	4	75
0,9999775	0,9996205	0,000357	0.5	9	6	75
0,9999945	0,999945	0,0000495	0.5	9	8	75
0,9999995	0,999999	0,0000005	0.5	9	10	25
0,9986645	0,9999725	-0,001308	0.6	9	2	75
0,999871	0,999999	-0,000128	0.6	9	4	75

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS PAUTA 3						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9999905	0,9999995	-0,000009	0.6	9	6	75
0,9999995	1	-0,0000005	0.6	9	8	75
0,999989	1	-0,000011	0.6	9	10	75
0,999047	0,9815715	0,0174755	0.7	9	2	75
0,999919	0,9968445	0,0030745	0.7	9	4	75
0,999989	0,999565	0,000424	0.7	9	6	75
1	0,9999395	0,0000605	0.7	9	8	75
1	0,9999975	0,0000025	0.7	9	10	75
0,9992815	0,9819825	0,017299	0.8	9	2	75
0,9999625	0,9970045	0,002958	0.8	9	4	75
0,999989	0,999465	0,000524	0.8	9	6	75
0,9999995	0,9999155	0,000084	0.8	9	8	75
1	0,9999975	0,0000025	0.8	9	10	75
0,999178	0,9929645	0,0062135	0.2	10	2	75
0,999844	0,998513	0,001331	0.2	10	4	75
0,9999945	0,9999385	0,000056	0.2	10	6	75
0,9999975	0,999989	0,0000085	0.2	10	8	75
0,9999995	1	-0,0000005	0.2	10	10	75
0,9990465	0,9895055	0,009541	0.3	10	2	25
0,9999595	0,99833	0,0016295	0.3	10	4	75
0,9999995	0,999904	0,0000955	0.3	10	6	75
1	0,9999915	0,0000085	0.3	10	8	75
0,9999995	0,9999985	0,000001	0.3	10	10	75
0,9994285	0,9892435	0,010185	0.4	10	2	75
0,999944	0,999007	0,000937	0.4	10	4	75
0,999996	0,999849	0,000147	0.4	10	6	75
1	0,999994	0,000006	0.4	10	8	75
0,9999995	0,999999	0,0000005	0.4	10	10	75
0,999547	0,9895635	0,0099835	0.5	10	2	75
0,9999545	0,9982745	0,00168	0.5	10	4	75
0,999979	0,999829	0,00015	0.5	10	6	75
0,999999	0,9999755	0,0000235	0.5	10	8	75
1	0,9999995	0,0000005	0.5	10	10	75
0,999638	0,9999955	-0,0003575	0.6	10	2	75
0,999967	0,999999	-0,000032	0.6	10	4	75
0,999997	0,999999	-0,000002	0.6	10	6	75
0,9999995	0,9999995	0	0.6	10	8	75
0,9999995	1	-0,0000005	0.6	10	10	75
0,9996735	0,989093	0,0105805	0.7	10	2	75
0,999981	0,9985225	0,0014585	0.7	10	4	25
0,9999995	0,9998265	0,000173	0.7	10	6	75
0,999999	0,9999815	0,0000175	0.7	10	8	75
0,9999995	0,9999985	0,000001	0.7	10	10	75
0,994987	0,9905965	0,0043905	0.8	10	2	75
0,998894	0,9980955	0,0007985	0.8	10	4	75
0,999776	0,999939	-0,000163	0.8	10	6	75
0,999975	0,9999945	-0,0000195	0.8	10	8	75
0,9999995	1	-0,0000005	0.8	10	10	75

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro D.1. Proporción de unidades estadísticas correctamente clasificadas por las dos reglas sin reducir el espacio p-dimensional

MEJOR DESEMPEÑO EN LA CLASIFICACIÓN DE UNIDADES POR LA REGLA PROPUESTA						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9191565	0,9258025	-0,006646	0.2	6	2	25
0,9418225	0,947237	-0,0054145	0.3	6	2	25
0,9997515	0,999852	-0,0001005	0.3	6	10	25
0,9755805	0,998333	-0,0227525	0.6	6	2	25
0,9948685	0,9998155	-0,004947	0.6	6	4	25
0,9989855	0,9999595	-0,000974	0.6	6	6	25
0,999587	0,999987	-0,0004	0.6	6	8	25
0,999802	0,9999995	-0,0001975	0.6	6	10	25
0,940585	0,9486625	-0,0080775	0.2	7	2	25
0,988545	0,9886525	-0,0001075	0.2	7	4	25
0,998466	0,9981475	0,0003185	0.2	7	6	25
0,999379	0,999758	-0,000379	0.2	7	8	25
0,998614	0,999067	-0,000453	0.3	7	6	25
0,9994215	0,9996445	-0,000223	0.3	7	8	25
0,999883	0,9998995	-0,0000165	0.3	7	10	25
0,952528	0,958048	-0,00552	0.5	7	2	25
0,999993	0,9999935	-0,0000005	0.5	7	10	25
0,9814645	0,9995625	-0,018098	0.6	7	2	25
0,996551	0,9999475	-0,0033965	0.6	7	4	25
0,999184	0,999996	-0,000812	0.6	7	6	25
0,9998385	0,9999995	-0,000161	0.6	7	8	25
0,99989	1	-0,00011	0.6	7	10	25
0,999932	0,999948	-0,000016	0.2	8	10	25
0,9741915	0,9765405	-0,002349	0.3	8	2	25
0,9961105	0,9963555	-0,000245	0.3	8	4	25
0,999031	0,9993985	-0,0003675	0.3	8	6	25
0,9999075	0,9999135	-0,000006	0.3	8	8	25
0,988955	0,9999155	-0,0109605	0.6	8	2	25
0,998123	0,9999955	-0,0018725	0.6	8	4	25
0,9997525	0,999999	-0,0002465	0.6	8	6	25
0,999977	1	-0,000023	0.6	8	8	25
0,9999955	1	-0,0000045	0.6	8	10	25
0,980594	0,981344	-0,00075	0.2	9	2	25
0,999997	0,9999975	-0,0000005	0.3	9	10	25
0,9924085	0,999988	-0,0075795	0.6	9	2	25
0,9989035	0,999999	-0,0010955	0.6	9	4	25
0,999865	0,9999995	-0,0001345	0.6	9	6	25
0,9999815	1	-0,0000185	0.6	9	8	25
1	1	0	0.6	9	10	25
1	1	0	0.6	10	10	25
0,9999825	0,999995	-0,0000125	0.5	10	10	25
0,9960685	0,999996	-0,0039275	0.6	10	2	25
0,9996515	0,999999	-0,0003475	0.6	10	4	25
0,9999665	0,999999	-0,0000325	0.6	10	6	25
0,999995	1	-0,000005	0.6	10	8	25
0,9999825	0,999995	-0,0000125	0.5	10	10	25
0,941004	0,9620075	-0,0210035	0.3	6	2	50
0,9874805	0,9905105	-0,00303	0.3	6	4	50

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

continua.

MEJOR DESEMPEÑO EN LA CLASIFICACIÓN DE UNIDADES POR LA REGLA PROPUESTA						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,997838	0,997921	-0,000083	0.3	6	6	50
0,9991525	0,9993845	-0,000232	0.3	6	8	50
0,976807	0,9984595	-0,0216525	0.6	6	2	50
0,9948145	0,9997655	-0,004951	0.6	6	4	50
0,9986345	0,9999425	-0,001308	0.6	6	6	50
0,999582	0,9999755	-0,0003935	0.6	6	8	50
0,999749	0,9999915	-0,0002425	0.6	6	10	50
0,9670705	0,9709805	-0,00391	0.3	7	2	50
0,994267	0,994577	-0,00031	0.3	7	4	50
0,9988945	0,9988995	-0,000005	0.3	7	6	50
0,9863905	0,9996425	-0,013252	0.6	7	2	50
0,997412	0,999951	-0,002539	0.6	7	4	50
0,9995505	0,999992	-0,0004415	0.6	7	6	50
0,999957	0,999999	-0,000042	0.6	7	8	50
0,9999905	0,9999995	-0,000009	0.6	7	10	50
0,999887	0,999987	-0,0001	0.3	8	8	50
0,999998	0,9999985	-0,0000005	0.3	8	10	50
0,9930785	0,9999195	-0,006841	0.6	8	2	50
0,999008	0,999991	-0,000983	0.6	8	4	50
0,999881	0,9999975	-0,0001165	0.6	8	6	50
0,9999805	0,9999995	-0,000019	0.6	8	8	50
0,9999925	1	-0,0000075	0.6	8	10	50
0,999954	0,999964	-0,00001	0.5	9	8	50
0,997201	0,999982	-0,002781	0.6	9	2	50
0,9995775	0,999999	-0,0004215	0.6	9	4	50
0,99993	0,999999	-0,000069	0.6	9	6	50
0,999984	0,9999995	-0,0000155	0.6	9	8	50
0,9999985	1	-0,0000015	0.6	9	10	50
0,998598	0,999997	-0,001399	0.6	10	2	50
0,999888	0,9999995	-0,0001115	0.6	10	4	50
0,9999905	1	-0,0000095	0.6	10	6	50
0,9999995	1	-0,0000005	0.6	10	8	50
1	1	0	0.6	10	10	50
1	1	0	0.8	10	10	50
0,9521555	0,9642265	-0,012071	0.3	6	2	75
0,990108	0,9907295	-0,0006215	0.3	6	4	75
0,978235	0,9981625	-0,0199275	0.6	6	2	75
0,995554	0,9997885	-0,0042345	0.6	6	4	75
0,9992005	0,9999815	-0,000781	0.6	6	6	75
0,9997465	0,999977	-0,0002305	0.6	6	8	75
0,999862	1	-0,000138	0.6	6	10	75
0,991177	0,9995095	-0,0083325	0.6	7	2	75
0,998436	0,9999655	-0,0015295	0.6	7	4	75
0,9997965	0,9999945	-0,000198	0.6	7	6	75
0,9999525	1	-0,0000475	0.6	7	8	75
0,999983	1	-0,000017	0.6	7	10	75
0,9965595	0,999906	-0,0033465	0.6	8	2	75

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

MEJOR DESEMPEÑO EN LA CLASIFICACIÓN DE UNIDADES POR LA REGLA PROPUESTA						
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables	Distancia	% var.
0,9994585	0,9999875	-0,000529	0.6	8	4	75
0,999956	0,9999995	-0,0000435	0.6	8	6	75
0,999991	1	-0,000009	0.6	8	8	75
1	1	0	0.6	8	10	75
0,9999785	0,999997	-0,0000185	0.4	9	10	75
0,9986645	0,9999725	-0,001308	0.6	9	2	75
0,999871	0,999999	-0,000128	0.6	9	4	75
0,9999905	0,9999995	-0,000009	0.6	9	6	75
0,9999995	1	-0,0000005	0.6	9	8	75
0,999989	1	-0,000011	0.6	9	10	75
0,9999995	1	-0,0000005	0.2	10	10	75
0,999638	0,9999955	-0,0003575	0.6	10	2	75
0,999967	0,999999	-0,000032	0.6	10	4	75
0,999997	0,999999	-0,000002	0.6	10	6	75
0,9999995	0,9999995	0	0.6	10	8	75
0,9999995	1	-0,0000005	0.6	10	10	75
0,999776	0,999939	-0,000163	0.8	10	6	75
0,999975	0,9999945	-0,0000195	0.8	10	8	75
0,9999995	1	-0,0000005	0.8	10	10	75

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro D.2. Lugares donde la regla propuesta tiene mejor desempeño frente a la regla de puntajes lineales discriminante

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS CON 25 %				
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables
0,999943	0,999739	0,000204	0.2	6
0,9997515	0,999852	-0,0001005	0.3	6
0,9996805	0,999315	0,0003655	0.4	6
0,9999155	0,999846	0,0000695	0.5	6
0,999802	0,9999995	-0,0001975	0.6	6
0,999924	0,9995755	0,0003485	0.7	6
0,9999855	0,999896	0,0000895	0.8	6
0,999975	0,9999375	0,0000375	0.2	7
0,999883	0,9998995	-0,0000165	0.3	7
0,999827	0,9996995	0,0001275	0.4	7
0,999993	0,9999935	-0,0000005	0.5	7
0,99989	1	-0,00011	0.6	7
0,9999665	0,99983	0,0001365	0.7	7
0,999994	0,9999795	0,0000145	0.8	7
0,999932	0,999948	-0,000016	0.2	8
0,999988	0,9999765	0,0000115	0.3	8
0,9999775	0,9999715	0,000006	0.4	8
0,999995	0,999934	0,000061	0.5	8
0,9999955	1	-0,0000045	0.6	8
0,999995	0,999955	0,00004	0.7	8
0,999991	0,9999515	0,0000395	0.8	8
0,999989	0,9999845	0,0000045	0.2	9
0,999997	0,9999975	-0,0000005	0.3	9
0,9999885	0,9999725	0,000016	0.4	9
0,9999995	0,9999925	0,000007	0.5	9
1	1	0	0.6	9
0,999999	0,999996	0,000003	0.7	9
0,999999	0,999972	0,000027	0.8	9
0,9999985	0,999995	0,0000035	0.2	10
0,9999995	0,9999975	0,000002	0.3	10
1	0,9999975	0,0000025	0.4	10
0,9999825	0,999995	-0,0000125	0.5	10
0,9999825	0,999995	-0,0000125	0.5	10
0,9999825	0,999995	-0,0000125	0.5	10
1	0,999997	0,000003	0.7	10
1	0,9999995	0,0000005	0.8	10

* *RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro D.3. Proporción de unidades estadísticas correctamente clasificadas por las regla de puntajes lineales y la regla propuesta al 25 % de variabilidad a una distancia de 10 sin reducir el espacio p-dimensional

PROPORCIÓN DE UNIDADES CORRECTAMENTE CLASIFICADAS CON 50 %				
Proporcion RPL	Proporcion RP	Diferencia	A-priori	Variables
0,9997235	0,999631	0,0000925	0.2	6
0,999839	0,9997745	0,0000645	0.3	6
0,999549	0,9992505	0,0002985	0.4	6
0,999944	0,9998425	0,0001015	0.5	6
0,999749	0,9999915	-0,0002425	0.6	6
0,9997945	0,9996115	0,000183	0.7	6
0,999959	0,9996295	0,0003295	0.8	6
0,999952	0,999892	0,00006	0.2	7
0,999905	0,9998905	0,0000145	0.3	7
0,9999815	0,9999255	0,000056	0.4	7
0,9999635	0,9998355	0,000128	0.5	7
0,9999905	0,9999995	-0,000009	0.6	7
0,9999615	0,9998335	0,000128	0.7	7
0,9999895	0,999884	0,0001055	0.8	7
0,9999995	0,9999885	0,000011	0.2	8
0,999998	0,9999985	-0,0000005	0.3	8
0,9999845	0,9999595	0,000025	0.4	8
0,999975	0,999949	0,000026	0.5	8
0,9999925	1	-0,0000075	0.6	8
0,9999995	0,999964	0,0000355	0.7	8
1	0,9999935	0,0000065	0.8	8
0,9999995	0,999999	0,0000005	0.2	9
0,999999	0,999998	0,000001	0.3	9
0,999998	0,9999765	0,0000215	0.4	9
0,9999995	0,999999	0,0000005	0.5	9
0,9999985	1	-0,0000015	0.6	9
1	0,9999945	0,0000055	0.7	9
1	0,9999995	0,0000005	0.8	9
0,9999995	0,9999985	0,000001	0.2	10
0,9999995	0,9999955	0,000004	0.3	10
1	0,9999995	0,0000005	0.4	10
1	0,9999995	0,0000005	0.5	10
1	1	0	0.6	10
0,999999	0,9999935	0,0000055	0.7	10
1	1	0	0.8	10

** RPL hace referencia a la regla de puntajes lineales, RP hace referencia a la regla propuesta .*

Cuadro D.4. Proporción de unidades estadísticas correctamente clasificadas por las regla de puntajes lineales y la regla propuesta al 50 % de variabilidad a una distancia de 10 sin reducir el espacio p-dimensional