MÉTODOS NÃO PARAMÉTRICOS DE DISCRIMINA-
ÇÃO BASEADOS EM REGIÕES DE TOLERÂNCIA

DENIŞE APARECIDA BOTTER

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA DE CONCENTRAÇÃO: ESTATÍSTICA

ORIENTADOR: PROF.DR.ABRAHAM LAREDO SICSU

- São Paulo, junho de 1984 -

Aos meus pais

Euclides e Dirce

e ao meu irmão

Rui

AGRADECIMENTOS

Queremos agradecer a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, em particular:

— Ao Professor Doutor Abraham Laredo Sicsū, pela paciência e dedicação na orientação e transmissão de seus conhecimentos.

— Ao Professor Rui Carlos Botter do Departamento de Engenharia Naval da Escola Politécnica da Universidade de São Paulo, pela elaboração dos programas de computação e pelo incentivo constante.

— Às amigas Raquel Cymrot e Emília Sakurai, pelas valiosas colaborações.

— Ao senhor João Baptista Esteves de Oliveira e senhora Maria Aparecida Oliveira De Pierro, pelo excelente trabalho de datilografia.

Í N D I C E

APRESENTAÇÃO.	ix
CAP. 1. INTRODUÇÃO.	1
1. Fundamentos.	1
2. Análise Discriminante Paramétrica.	4
3. Análise Discriminante Não Paramétrica.	9
4. Classificação dos Métodos Não Paramétricos de Discrimi- nação.	11
5. Algumas Definições	14
6. Métodos de Estimação da Probabilidade de Erro de Classi- ficação.	15
CAP. 2. REGIÕES DE TOLERÂNCIA	20
2.1. Histórico e Definição.	20
2.2. Métodos de Construção de Regiões de Tolerância	21
2.3. Escolha das Funções de Corte $\phi_i(\underline{x})$ e da Sequência $\{K_i\}$ de Números Inteiros.	30
2.3.1. Método de subpartição de regiões de tolerância.	31
2.3.2. Método de Tukey (1947)	35
2.4. Probabilidades Associadas às Regiões de Tolerância	37
CAP. 3. MÉTODOS DE DISCRIMINAÇÃO NÃO PARAMÉTRICOS BASEADOS EM REGIÕES DE TOLERÂNCIA	40
3.1. Métodos de Discriminação Forçada	40
3.1.1. Regras discriminatórias — Síntese	41
3.1.2. Descrição dos principais métodos encontrados na literatura.	44
3.1.3. Considerações	47
3.1.4. Um exemplo numérico	48
3.1.5. Consistência dos métodos apresentados	55
3.1.6. Probabilidade de erro de classificação.	58
3.2. Métodos de Discriminação Parcial	59
3.2.1. Modelo geral.	60
3.2.1. Apresentação dos métodos.	61
3.2.3. Um exemplo numérico	66
3.2.4. Probabilidades de erro de classificação e de não classificação	70
CAP. 4. MÉTODOS NÃO PARAMÉTRICOS DE DISCRIMINAÇÃO BASEADOS NOS v VIZINHOS MAIS PRÓXIMOS.	73
4.1. Introdução	73
4.2. Método de Discriminação Forçada.	75

4.2.1. Estimaco de $f_i(x)$, $i=1,2$ e Regra discriminatria	75
4.2.2. Consideraces	78
4.2.3. Um exemplo numrico	82
4.2.4. Consistncia do mtodo v-VMP.	82
4.2.5. Probabilidade de erro de classificaco.	84
4.3. Mtodo de Discriminaco Parcial.	87
4.3.1. Regra discriminatria	87
4.3.2. Probabilidades de erro de classificaco e de no classificaco	89
CAP. 5. RESULTADOS EXPERIMENTAIS.	91
5.1. Notaco.	91
5.2. Resultados Experimentais Encontrados na Literatura Pesquisada	92
5.2.1. Comparaco entre o mtodo discriminante linear amostral e o baseado nos v vizinhos mais prximos (Fix e Hodges, 1952)	93
5.2.2. Comparaco entre o mtodo Discriminante linear amostral e o de Gessaman e Gessaman (1972) baseado em regies de tolerncia - Discriminaco forçada (Gessaman e Gessaman, 1972)	93
5.2.3. Comparaco entre os mtodos de discriminaco parcial de Gessaman e Gessaman (1972), Quesenberry e Gessaman (1968) e Kendall (1966) (Gessaman e Gessaman, 1972)	96
5.3. Outros Resultados Experimentais.	99
5.3.1. Mtodo v-VMP: Discusso sobre o valor de v.	99
5.3.2. Comparaco entre os mtodos de discriminaco linear amostral, quadrtico amostral, v vizinhos mais prximos e de Gessaman e Gessaman (1972) - Discriminaco forçada.	102
APNDICE A.	108
APNDICE B.	110
APNDICE C.	113
APNDICE D.	118
APNDICE E.	124
BIBLIOGRAFIA.	148

APRESENTAÇÃO

A Análise Discriminante é um conjunto de métodos estatísticos cuja finalidade é a locação de um elemento em uma das populações, supondo que esse elemento pertence realmente a uma delas.

Na Análise Discriminante Não Paramétrica, a construção dos métodos de discriminação independe do conhecimento da função de distribuição da variável aleatória associada a cada uma das populações, exceto, talvez, quanto à existência de densidade.

Suponhamos disponível uma amostra de observações de cada uma das populações. Uma região de tolerância é um conjunto definido no espaço amostral da variável aleatória acima mencionada, através das observações amostrais de pelo menos uma das populações.

Os métodos não paramétricos de discriminação construídos através de regiões de tolerância baseiam-se no número de observações amostrais de cada uma das populações, encontradas na região de tolerância que contém o elemento a ser discriminado.

No presente trabalho apresentamos os principais métodos de discriminação baseados em regiões de tolerância. O primeiro Capítulo se destina a dar um breve histórico da análise dis-

criminante, apresentar um resumo da análise discriminante paramétrica, justificar a razão do estudo da análise discriminante não paramétrica baseada em regiões de tolerância e apontar as principais dificuldades que ocorrem na sua aplicação prática. No Capítulo 2 abordaremos os métodos de construção de regiões de tolerância. No Capítulo 3 é feito um estudo dos métodos não paramétricos de discriminação baseados em regiões de tolerância encontrados na literatura pesquisada, apontando suas características e estabelecendo comparações entre os mesmos. O Capítulo 4 trata de um método particular de discriminação baseado em regiões de tolerância que particiona o espaço amostral, através de uma função distância, em regiões de tolerância mutuamente exclusivas e discrimina um elemento com base na região de tolerância à qual pertencem esse elemento e seus vizinhos mais próximos. Finalmente, o Capítulo 5 consiste da apresentação de trabalhos de aplicação das técnicas desenvolvidas nos capítulos anteriores, tanto os encontrados na literatura pesquisada quanto os por nós desenvolvidos com o auxílio de programas computacionais que elaboramos e processamos no CCE-USP.

CAPÍTULO 1

INTRODUÇÃO

1. FUNDAMENTOS

Consideraremos, neste trabalho, duas populações distintas, π_1 e π_2 , sendo que extensões para mais de duas populações não são muito trabalhosas. A cada elemento dessas populações associaremos um vetor $\underline{x} = (x_1, \dots, x_p)$ de valores da variável aleatória p-dimensional $\underline{X} = (X_1, \dots, X_p)$, cuja função de distribuição será assumida contínua. Denotaremos por F_i , a função de distribuição e, por f_i , a função de densidade de probabilidade de $\underline{X}$, na população π_i , $i=1,2$. Denotaremos o espaço amostral de $\underline{X}$ por S_p , sendo este um subconjunto de R^p .

Admitiremos conhecida uma amostra de π_1 com n_1 elementos e outra de π_2 com n_2 elementos.

Denotaremos por $\underline{x}_{ij}$, o vetor associado ao j-ésimo elemento da amostra da população π_i , $i=1,2$; $j=1, \dots, n_i$.

Seja E um elemento, escolhido ao acaso entre todos os elementos de π_1 e π_2 , ao qual corresponde o vetor de medidas $\underline{x} = (x_1, \dots, x_p)$.

O problema básico da Análise Discriminante consiste em

alocar o elemento E em uma das duas populações. Para tanto, particionaremos S_p em duas regiões D_1 e D_2 de modo que se $\underline{x}$ pertencer a D_1 , classificaremos E na população π_1 e caso contrário, se $\underline{x}$ pertencer a D_2 , classificaremos E em π_2 .

Denotaremos por p_i , a probabilidade *a priori* do elemento E pertencer à população π_i e por c_{ji} , o custo incorrido ao classificar o elemento E em π_j quando, na realidade, ele pertence à população π_i , $i, j=1, 2$.

A partição $D = (D_1, D_2)$ é denominada regra discriminatória. Existem vários critérios de determinação de uma "boa" regra discriminatória. Apresentaremos os dois mais utilizados.

CRITÉRIO 1 - Minimização do custo total de classificação

Consiste em determinar a regra $D = (D_1, D_2)$ que minimiza a seguinte função custo:

$$C(D_1, D_2) = P(D_2 | \pi_1) \cdot p_1 \cdot c_{21} + P(D_1 | \pi_2) \cdot p_2 \cdot c_{12} \quad (1.1)$$

onde $P(D_j | \pi_i) = P(\pi_j | \pi_i)$ é a probabilidade de classificar um elemento em π_j , quando ele pertence a π_i , $i, j=1, 2$.

Assumindo $c_{ii} = 0$ e definindo $c_{ij} = 1$; $i, j=1, 2$; $i \neq j$, o critério do custo total mínimo equivale a minimizar a probabilidade total de erro de classificação, isto é, as regiões de classificação D_j , $j=1, 2$, são escolhidas de maneira a minimizar a seguinte função:

$$P(D_1, D_2) = P(D_2 | \pi_1) \cdot p_1 + P(D_1 | \pi_2) \cdot p_2 \quad (1.2)$$

O critério da minimização da probabilidade total de erro de classificação conduz, em algumas situações, a valores de $p_1 P(D_2 | \pi_1)$ e $p_2 P(D_1 | \pi_2)$ desproporcionais. Por esta razão, muitas vezes adota-se um critério de determinação da partição D diferente do critério de minimização do custo total ou da probabilidade total de erro de classificação.

O mais utilizado destes critérios é apresentado a seguir.

CRITÉRIO 2- Minimização da maior probabilidade de erro de classificação (Critério minimax)

O problema consiste em determinar $D = (D_1, D_2)$ que minimize a expressão:

$$\max \{ P(D_1 | \pi_2), P(D_2 | \pi_1) \} \quad (1.3)$$

No desenvolvimento deste trabalho utilizaremos, somente, o critério da minimização da probabilidade total de erro de classificação.

Na obtenção de uma regra discriminante é necessário estabelecer hipóteses sobre a função de distribuição F_i , $i=1,2$.

Podemos distinguir três casos:

- a) F_i é completamente conhecida;
- b) F_i é conhecida, exceto quanto aos valores de um ou mais parâmetros, e,
- c) F_i é completamente desconhecida, exceto, talvez, quanto à suposição sobre a existência de densidade.

2. ANÁLISE DISCRIMINANTE PARAMÉTRICA

Considerando o caso a), onde F_i , $i=1,2$, é completamente conhecida, a obtenção da regra ótima de classificação é simples. Minimizar a função dada em (1.2) é equivalente a minimizar a expressão

$$p_1 \int_{D_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{D_1} f_2(\underline{x}) d\underline{x} \quad (2.1)$$

A solução deste problema (Welch, 1939) leva à seguinte regra ótima de classificação:

$$\left\{ \begin{array}{l} \text{Se } \frac{f_1(\underline{x})}{f_2(\underline{x})} > \frac{p_2}{p_1}, \text{ classificamos o elemento } E \text{ em } \pi_1 \\ \text{Se } \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_2}{p_1}, \text{ classificamos o elemento } E \text{ em } \pi_2 \end{array} \right. \quad (2.2)$$

Observemos que:

$$\begin{aligned} D_1 &= \left\{ \underline{x} \in S_p \mid \frac{f_1(\underline{x})}{f_2(\underline{x})} > \frac{p_2}{p_1} \right\} \\ D_2 &= \left\{ \underline{x} \in S_p \mid \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_2}{p_1} \right\} \end{aligned} \quad (2.3)$$

Esta regra ótima de discriminação é denominada regra da razão de verossimilhança.

Quando F_1 e F_2 são, respectivamente, normais p -variadas com médias $\underline{\mu}_1$ e $\underline{\mu}_2$ e matrizes de covariância $\underline{\Sigma}_1$ e $\underline{\Sigma}_2$ e quando supomos que $\underline{\Sigma}_1 = \underline{\Sigma}_2 = \underline{\Sigma}$, a regra ótima de discriminação é dada

por:

$$\left\{ \begin{array}{l} \text{Se } [\underline{x} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)]' \underline{\Sigma}^{-1}(\underline{\mu}_1 - \underline{\mu}_2) > \ln \frac{p_2}{p_1}, \text{ classificamos} \\ \quad \text{o elemento E em } \pi_1 \\ \\ \text{Se } [\underline{x} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)]' \underline{\Sigma}^{-1}(\underline{\mu}_1 - \underline{\mu}_2) < \ln \frac{p_2}{p_1}, \text{ classificamos} \\ \quad \text{o elemento E em } \pi_2. \end{array} \right. \quad (2.4)$$

Esta regra ótima de discriminação é denominada regra discriminante linear.

As probabilidades exatas de erro de classificação da regra discriminante linear são dadas por:

$$\begin{aligned} P_L = (\pi_1/\pi_2) &= \phi \left\{ [-\ln(p_2/p_1) - \delta^2/2]/\delta \right\} \\ P_L = (\pi_2/\pi_1) &= \phi \left\{ [\ln(p_2/p_1) - \delta^2/2]/\delta \right\} \end{aligned} \quad (2.5)$$

onde

$$\delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

é a distância generalizada de Mahalanobis e $\phi(\cdot)$ denota a função de distribuição da variável aleatória normal padrão, $N(0,1)$.

Quando F_1 e F_2 são, respectivamente, normais p -variadas, mas as matrizes de covariância são diferentes, isto é, $\underline{\Sigma}_1 \neq \underline{\Sigma}_2$, temos a seguinte regra discriminante:

$$\left\{ \begin{array}{l} \text{Se } \frac{1}{2} \left[\ln \frac{|\underline{\Sigma}_2|}{|\underline{\Sigma}_1|} + (\underline{x} - \underline{\mu}_2)' \underline{\Sigma}_2^{-1} (\underline{x} - \underline{\mu}_2) - (\underline{x} - \underline{\mu}_1)' \underline{\Sigma}_1^{-1} (\underline{x} - \underline{\mu}_1) \right] > \ln \frac{p_2}{p_1}, \\ \text{classificamos o elemento E em } \pi_1 \\ \\ \text{Se } \frac{1}{2} \left[\ln \frac{|\underline{\Sigma}_2|}{|\underline{\Sigma}_1|} + (\underline{x} - \underline{\mu}_2)' \underline{\Sigma}_2^{-1} (\underline{x} - \underline{\mu}_2) - (\underline{x} - \underline{\mu}_1)' \underline{\Sigma}_1^{-1} (\underline{x} - \underline{\mu}_1) \right] < \ln \frac{p_2}{p_1}, \\ \text{classificamos o elemento E em } \pi_2. \end{array} \right. \quad (2.6)$$

Esta regra ótima de discriminação é denominada regra discriminante quadrática.

O cálculo das probabilidades exatas de erro de classificação para a regra discriminante quadrática é bastante complexo (ver, por exemplo, Anderson e Bahadur, 1962). Uma alternativa ao cálculo de tais probabilidades é estimá-las (ver seção 6).

* * *

Consideremos o caso b), onde F_i , $i=1,2$, é conhecida excepto quanto aos valores de alguns parâmetros que denotaremos conjuntamente por $\underline{\theta}$. Denotando as distribuições correspondentes a um dado $\underline{\theta}$ por $F_{i,\underline{\theta}}$, utilizamos as amostras de π_i , $i=1,2$, para estimar $\underline{\theta}$, através de $\hat{\underline{\theta}}$, e então classificamos o elemento E segundo a regra discriminatória dada em (2.2), usando a distribuição $F_{i,\hat{\underline{\theta}}}$ no lugar de $F_{i,\underline{\theta}}$.

Por exemplo, no caso da regra discriminante linear, ao substituírmos $\underline{\mu}_1$, $\underline{\mu}_2$ e $\underline{\Sigma}$ por seus estimadores de máxima veros-

semelhança, $\bar{X}_1$, $\bar{X}_2$ e $\underline{S}$, teremos a seguinte regra discriminante:

$$\left\{ \begin{array}{l} \text{Se } [\underline{x} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)]' \underline{S}^{-1} (\bar{x}_1 - \bar{x}_2) > \ln \frac{p_2}{p_1}, \text{ classificamos o} \\ \text{elemento E em } \pi_1 \\ \\ \text{Se } [\underline{x} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)]' \underline{S}^{-1} (\bar{x}_1 - \bar{x}_2) < \ln \frac{p_2}{p_1}, \text{ classificamos o} \\ \text{elemento E em } \pi_2. \end{array} \right. \quad (2.7)$$

Esta regra discriminante é denominada regra discriminante linear amostral.

Os coeficientes $\underline{S}^{-1}(\bar{x}_1 - \bar{x}_2)$ são denominados coeficientes discriminantes.

Fisher (1936) propôs, para resolução do problema discriminante, a obtenção de uma função linear Y,

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_p X_p \quad (2.8)$$

que maximizasse o quociente entre a diferença dos valores de Y calculados para as médias amostrais e o desvio padrão de Y estimado "dentro" das amostras.

O vetor dos coeficientes b_i da função Y, obtido por Fisher, é igual a $\underline{W}^{-1}(\bar{x}_1 - \bar{x}_2)$ onde $\underline{W}$ é a matriz da soma de quadrados e produtos dentro das amostras, isto é,

$$(n_1 + n_2 - 2) \underline{S} = \underline{W} \quad (2.9)$$

Observamos que os coeficientes obtidos por Fisher são iguais aos coeficientes discriminantes $\underline{S}^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$ multiplicados por uma constante. Deste modo, justifica-se o fato da regra discriminante dada em (2.7) ser denominada regra discriminante linear de Fisher.

As probabilidades exatas de erro de classificação da regra discriminante linear amostral, deduzidas em Lachenbruch (1975) são dadas por:

$$P_{LS}(\pi_2 | \pi_1) = \phi \left\{ \frac{\ln \left(\frac{p_2}{p_1} \right) - \left[\mu_1 - \frac{1}{2} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2) \right]' \underline{S}^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}{\sqrt{V_S}} \right\} \quad (2.10)$$

$$P_{LS}(\pi_1 | \pi_2) = \phi \left\{ \frac{\ln \left(\frac{p_2}{p_1} \right) + \left[\mu_2 - \frac{1}{2} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2) \right]' \underline{S}^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}{\sqrt{V_S}} \right\}$$

onde $V_S = (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' \underline{S}^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$

Notemos que o cálculo de $P_{LS}(\pi_2 | \pi_1)$ e de $P_{LS}(\pi_1 | \pi_2)$ só é possível se os parâmetros populacionais são conhecidos; caso contrário não podemos calcular tais probabilidades. Nesta situação, uma alternativa é estimar $P_{LS}(\pi_2 | \pi_1)$ e $P_{LS}(\pi_1 | \pi_2)$ substituindo μ_i , $i = 1, 2$, e $\underline{S}$ por seus estimadores de máxima verossimilhança nas expressões dadas em (2.10).

Teremos:

$$P_{LS}(\pi_2 | \pi_1) = \phi \left\{ \left[\ln \left(\frac{p_2}{p_1} \right) - D^2 \right] / D \right\} \quad (2.11)$$

$$P_{LS}(\pi_1|\pi_2) = \phi \left\{ \left[-\ln \left(\frac{p_2}{p_1} \right) - D^2 \right] | D \right\}$$

onde

$$D^2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$$

Notemos que $P_{LS}(\pi_2|\pi_1)$ e $P_{LS}(\pi_1|\pi_2)$ não só estimam $P_{LS}(\pi_2|\pi_1)$ e $P_{LS}(\pi_1|\pi_2)$ como também estimam $P_L(\pi_2|\pi_1)$ e $P_L(\pi_1|\pi_2)$.

Para o caso da regra discriminante quadrática, procedemos de maneira análoga à regra discriminante linear, isto é, substituímos os parâmetros μ_1 , μ_2 , Σ_1 e Σ_2 por seus estimadores de máxima verossimilhança, obtendo a regra discriminante quadrática amostral.

Não encontramos, na literatura pesquisada, uma expressão para $P_{QS}(\pi_i|\pi_j)$, $i, j=1, 2; i \neq j$, probabilidade de erro de classificação da regra discriminante quadrática amostral. Nesta situação, podemos estimar $P_{QS}(\pi_i|\pi_j)$ através de algum dos métodos apresentados na seção 6.

3. ANÁLISE DISCRIMINANTE NÃO PARAMÉTRICA

A robustez de uma regra discriminante se reflete no efeito causado nas probabilidades de erro de classificação, quando as suposições básicas para a construção de tal regra não se verificam.

Lachenbruch, Sneeringer e Revo (1973) estudaram a robustez das funções discriminantes linear e quadrática, quando aplicadas a distribuições contínuas não normais. Seus estudos mostraram

que, nestas condições, as probabilidades totais de erro de classificação, $\sum_i p_i P(\pi_j | \pi_i)$, $i, j = 1, 2$; $i \neq j$, superam, em muitos, o valor ótimo.

Gilbert (1969) e Marks e Dunn (1974) estudaram o comportamento da função discriminante linear, quando as matrizes de covariância são diferentes, supondo distribuições normais. Para tanto, os parâmetros das distribuições normais envolvidas foram colocados na forma canônica, através de uma transformação linear, de forma que foram obtidos:

$\underline{\mu}_1 = \underline{0}$, $\underline{\mu}_2 = \underline{v}$, $\underline{\Sigma}_1 = \underline{I}$ (matriz identidade) e $\underline{\Sigma}_2 = \underline{\Lambda}$, onde $\underline{\Lambda}$ é uma matriz diagonal denotada por $\underline{\lambda}' = (\lambda_1, \lambda_2, \dots, \lambda_p)$.

Considerando três vetores $\underline{\lambda}'$ distintos

$$\underline{\lambda}' = (\lambda, \dots, \lambda); \quad \underline{\lambda}' = \frac{(\lambda, \dots, \lambda, 1, \dots, 1)}{p/2} \quad \text{e} \quad \underline{\lambda}' = \frac{(\lambda, 2\lambda, \dots, p\lambda)}{p/2},$$

estes autores mostraram que:

- a função discriminante linear é satisfatória quando λ está próximo de 1, ou quando os vetores de médias estão bastante afastados;
- a função discriminante quadrática amostral é bastante pobre, isto é, a probabilidade total de erro de classificação é muito superior ao valor ótimo, quando amostras pequenas são consideradas.

Através das conclusões dos estudos de robustez justifica-se a necessidade de pesquisa de métodos de discriminação que independam da suposição de normalidade, da suposição de homocedasticidade, ou de qualquer outra suposição sobre a função de distribuição F_i , $i=1, 2$, e que, nestes casos, forneçam

pequenos valores para as probabilidades de erro de classificação.

4. CLASSIFICAÇÃO DOS MÉTODOS NÃO PARAMÉTRICOS DE DISCRIMINAÇÃO

A família de métodos de discriminação não paramétricos que mais tem motivado atividades de pesquisa, trabalhos aplicados e, portanto, a realização desta monografia, é a família dos métodos de discriminação baseados em regiões de tolerância, estudados inicialmente por Fix e Hodges (1951).

Podemos resumir os métodos de discriminação baseados em regiões de tolerância da seguinte forma:

A partir das observações amostrais de π_1 e π_2 , particionemos S_p em conjuntos mutuamente exclusivos, denominados regiões de tolerância. Sejam N_1 e N_2 , os números de observações amostrais de π_1 e π_2 , respectivamente, pertencentes à região de tolerância que contém $\underline{x}$, vetor associado ao elemento E a ser classificado. A discriminação do elemento E basear-se-á nos números N_1 e N_2 encontrados.

Analisando a construção dos métodos de discriminação baseados em regiões de tolerância, verificamos que a partição conveniente de S_p , através das observações amostrais de π_1 e π_2 , em conjuntos mutuamente exclusivos, tem por objetivo delimitar as regiões onde predomina a densidade de probabilidade de π_1 ou π_2 , isto é, regiões onde a classificação de um elemento E em π_1 seja mais provável do que em π_2 e vice-versa.

Analisando os trabalhos encontrados sobre os métodos de discriminação baseados em regiões de tolerância, verificamos que a maioria de seus autores :

i) preocupam-se muito mais com a construção dos métodos, do que com a existência ou não de consistência com o método ótimo da razão de verossimilhança (ver seção 5). Somente dois métodos de discriminação forçada (ver seção 5), dentre os encontrados na literatura pesquisada, são consistentes com o método ótimo da razão de verossimilhança:

- método dos v vizinhos mais próximos, desenvolvido por Fix e Hodges (1951); a demonstração de consistência foi elaborada por esses autores.
- método de Patrick e Fischer (1970); a demonstração de consistência foi por nós desenvolvida.

ii) não propõem expressões para o cálculo exato da probabilidade de erro de classificação dos métodos por eles desenvolvidos. Somente para dois métodos de discriminação forçada encontramos o cálculo da probabilidade exata de erro de classificação, que apresentamos no apêndice D:

- método dos v vizinhos mais próximos, de Cover e Hart (1967);
- método de Patrick e Fischer (1970).

* * *

Uma outra família de métodos de discriminação não para-

métricos é a dos baseados em distâncias entre distribuições, que conta com pouca literatura. Um trabalho importante nesta área, deve-se a Gupta (1963). Em seu trabalho, Gupta desenvolve uma regra de discriminação utilizando a distância de Kolmogorov entre funções de distribuição. Uma amostra de observações é discriminada em π_i , $i=1,2$, se a distância entre a sua função de distribuição e a função de distribuição formada pelas observações amostrais de π_i for mínima.

* . * *

Kendall (1966) desenvolveu um método não paramétrico de discriminação parcial que não pertence à família dos métodos de discriminação baseados em regiões de tolerância, nem à família dos métodos de discriminação baseados em distâncias entre distribuições. Consideramos importante citá-lo neste trabalho pois é um método não paramétrico de discriminação de fácil construção.

* * *

Na literatura pesquisada encontramos vários trabalhos sobre estimação não paramétrica de funções densidade de probabilidade (ver por exemplo, Parzen, 1962; Cacoullos, 1966; Schwartz, 1967 e Wegman, 1972).

Substituindo tais estimadores na regra ótima da razão de verossimilhança dada em (2.2), obtemos os métodos não paramétricos de discriminação baseados em estimadores de densidade de probabilidade.

No entanto, a utilização destes métodos de discriminação

é praticamente inexistente pois poucos resultados sobre estimação das funções densidade de probabilidade de fácil aplicação, podem ser encontrados, principalmente para mais de uma dimensão ($p > 1$).

5. ALGUMAS DEFINIÇÕES

DEFINIÇÃO 1. Um método de *discriminação forçada* é um procedimento através do qual o elemento E deve ser, *necessariamente*, classificado em π_1 ou π_2 .

DEFINIÇÃO 2. Um método de *discriminação parcial* é um procedimento que permite deixar de classificar o elemento E em π_1 ou π_2 , isto é, podemos nos abster de tomar uma decisão quanto à origem populacional de E.

DEFINIÇÃO 3. Um critério utilizado para a obtenção de um *método ótimo de discriminação parcial*, com pleno conhecimento de F_i , $i = 1, 2$, é minimizar a probabilidade de não classificação, fixados valores máximos para as probabilidades de erro de classificação.

DEFINIÇÃO 4. Seja M_{n_i} um método de discriminação não paramétrico construído através de uma amostra de n_i observações de π_i , $i = 1, 2$, e M_0 , um método ótimo de discriminação.

Dizemos que o método não paramétrico é *consistente* com o método ótimo se, quando n_i cresce, M_{n_i} e M_0 tendem a discriminar o elemento E na mesma população.

Formalmente, M_{n_i} e M_0 são consistentes se, para todo $0 < \epsilon < 1$, existe n_0 tal que para $n_i > n_0$, $i = 1, 2$,
 $P \left\{ M_{n_i} \text{ e } M_0 \text{ classificarem } E \text{ na mesma população} \right\} > \epsilon.$

DEFINIÇÃO 5. Seja $\hat{f}_n(\underline{x})$ um estimador da função densidade de probabilidade f calculada no ponto $\underline{x}$ obtido a partir de uma amostra de n observações da variável aleatória p -dimensional $\underline{X}$.

Dizemos que $\hat{f}_n(\underline{x})$ é consistente com $f(\underline{x})$ se, para todo $\epsilon > 0$, temos

$$\lim_{n \rightarrow \infty} P[|\hat{f}_n(\underline{x}) - f(\underline{x})| < \epsilon] = 1$$

6. MÉTODOS DE ESTIMAÇÃO DA PROBABILIDADE DE ERRO DE CLASSIFICAÇÃO

Seja M uma regra discriminante ótima e MS uma regra discriminante amostral, isto é, uma regra discriminatória cuja construção baseia-se numa amostra de n_1 observações de π_1 e numa amostra de n_2 observações amostrais de π_2 .

Denotemos por $P_M(\pi_i|\pi_j)$ e $P_{MS}(\pi_i|\pi_j)$, $i, j = 1, 2$; $i \neq j$, as probabilidades exatas de erro de classificação dos métodos de discriminação M e MS , respectivamente.

MÉTODO DE ESTIMAÇÃO POR SIMULAÇÃO

O método de estimação por simulação da probabilidade de erro de classificação é uma alternativa ao cálculo de $P_M(\pi_i|\pi_j)$,

quando este é muito trabalhoso.

Consideremos disponível uma amostra de n_1 observações de π_1 e de n_2 observações de π_2 .

Utilizando a regra discriminante M, discriminemos cada uma das n_1 observações amostrais de π_1 e das n_2 observações amostrais de π_2 . O valor estimado por simulação de $P_M(\pi_1|\pi_2)$ será dado por:

$$P_M(\pi_i|\pi_j) = \frac{n_{ji}}{n_j}, \quad i, j = 1, 2; i \neq j \quad (6.1)$$

onde n_{ij} denota o número de observações amostrais de π_i classificadas erradamente em π_j .

MÉTODO DE ESTIMAÇÃO POR RESUBSTITUIÇÃO

Discriminemos, através da regra discriminante MS, cada uma das (n_1+n_2) observações amostrais de π_1 e π_2 que a construíram. O valor de $P_{MS}(\pi_i|\pi_j)$ estimado por resubstituição será dado por:

$$P_{MS}^R(\pi_i|\pi_j) = \frac{n_{ji}}{n_j}, \quad i, j = 1, 2; i \neq j \quad (6.2)$$

onde n_{ij} denota o número de observações amostrais de π_i classificadas erradamente em π_j .

MÉTODO DE ESTIMAÇÃO POR EXCLUSÃO

Excluamos, sequencialmente, cada uma das $(n_1 + n_2)$ obser-

vações amostrais obtidas de π_1 e π_2 . Construímos a regra discriminante MS através das (n_1+n_2-1) observações amostrais restantes. Discriminemos, através da regra construída, a observação amostral excluída. O valor de $P_{MS}(\pi_i|\pi_j)$ estimado por exclusão será dado pela expressão (6.2) e denotado por $\hat{P}_{MS}^E(\pi_i|\pi_j)$.

MÉTODO DE ESTIMAÇÃO POR PARTIÇÃO

Separemos dentre as n_i observações amostrais de π_i , uma quantidade pré-fixada, n'_i , $i=1,2$. Construímos a regra discriminante MS através das $(n_1+n_2-n'_1-n'_2)$ observações amostrais restantes. Discriminemos, através da regra construída, cada uma das $(n'_1+n'_2)$ observações amostrais separadas. O valor de $P_{MS}(\pi_i|\pi_j)$ estimado por partição será dado por:

$$\hat{P}_{MS}^P(\pi_i|\pi_j) = \frac{n'_{ji}}{n'_j}, \quad i, j = 1, 2; i \neq j \quad (6.3)$$

onde n'_{ij} denota o número de observações amostrais de π_i , encontradas entre as n'_j , classificadas erradamente em π_j .

Convém observar que $P_{MS}(\pi_i|\pi_j)$ estimada por resubstituição, exclusão ou partição também estima $P_M(\pi_i|\pi_j)$, $i, j=1, 2; i \neq j$.

Segundo Lachenbruch (1975), vários trabalhos empíricos mostram que o método de estimação por exclusão é o melhor método de estimação da probabilidade de erro de classificação, no sentido de que as probabilidades por ele produzidas estão bastante próximas das reais, mesmo que sejam utilizadas amostras de pequeno tamanho; o método de estimação por resubsti -

tuição apresenta a desvantagem de não ser um bom método de estimação, isto é, de produzir estimativas da probabilidade de erro de classificação distantes das reais, quando utilizadas amostras de tamanho pequeno.

7. ORGANIZAÇÃO DA MONOGRAFIA

A seguir faremos um resumo do trabalho a ser desenvolvido nesta monografia:

- descreveremos, de maneira didática, os métodos não paramétricos de discriminação baseados em regiões de tolerância, encontrados na literatura pesquisada;
- apontaremos as características dos métodos apresentados;
- demonstraremos, quando possível, a consistência dos métodos apresentados com o método ótimo da razão de verossimilhança;
- compararemos alguns dos métodos apresentados entre si e com os métodos discriminantes quadrático e linear, através de suas probabilidades exatas ou estimadas de erro de classificação (discriminação forçada) e de não classificação (discriminação parcial).

* * *

Um problema de ordem prática apresentado por alguns dos métodos não paramétricos baseados em regiões de tolerância é exigir que todos os valores das observações amostrais de π_1 e π_2 sejam utilizadas a cada discriminação de um novo elemento. Em outras palavras, é necessário refazer a regra discriminante cada vez que quisermos classificar um novo elemento. Desse mo-

do, aplicações práticas desses métodos podem consumir demasiado tempo computacional.

Além disso, notamos que estes métodos de discriminação não paramétricos encontram-se em estágios iniciais no que se refere à seleção de variáveis.

CAPÍTULO 2

REGIÕES DE TOLERÂNCIA

2.1. HISTÓRICO E DEFINIÇÃO

A teoria sobre regiões de tolerância foi iniciada por Wilks (1941). Outros trabalhos nesta área foram apresentados por: Wilks (1942, 1962), Wald (1943), Tukey (1947), Murphy (1948), Fraser (1951, 1957), Kemperman (1956), Fraser e Guttman (1956) e Anderson (1966). Após 1966 nada encontramos sobre a teoria de construção de regiões de tolerância. A partir desta data começaram a surgir trabalhos sobre estimação de funções densidade de probabilidade, baseados no uso de regiões de tolerância, como, por exemplo, o trabalho de Gessaman (1970).

Seja $\{ \underline{x}_1, \underline{x}_2, \dots, \underline{x}_n \}$ uma amostra de tamanho n , da variável aleatória p -dimensional $\underline{X}$, com função de distribuição F e função densidade de probabilidade f .

Denotemos o espaço amostral de $\underline{X}$ por S_p , sendo este um subconjunto de $\mathbb{R}^p$.

Um conjunto S , contido em S_p , é denominado *região de tolerância*, se for determinado por uma função das observações amostrais $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$.

Exemplo. Seja $\underline{x}_A$ um vetor pertencente a S_p . Ordenemos as

observações amostrais $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$, através de suas distâncias até $\underline{x}_A$. Uma região de tolerância, S_A , é o conjunto determinado pelas v , $v \leq n$, observações amostrais mais próximas de $\underline{x}_A$, isto é

$$S_A = \left\{ \underline{x} \in S_p \mid d(\underline{x}, \underline{x}_A) \leq d(\underline{x}^{(v)}, \underline{x}_A) \right\}.$$

onde

d é uma função distância definida em S_p e $\underline{x}^{(v)}$ é o vetor amostral correspondente ao valor da v -ésima distância até $\underline{x}_A$, uma vez colocadas em ordem crescente as distâncias entre cada vetor amostral e $\underline{x}_A$.

2.2. MÉTODOS DE CONSTRUÇÃO DE REGIÕES DE TOLERÂNCIA

Na literatura pesquisada encontramos diversos métodos de construção de regiões de tolerância.

Fraser (1957) uniformizou todos esses métodos; Anderson (1966) ilustrou o método de Fraser com alguns exemplos, para torná-lo mais compreensível. Em linhas gerais, Fraser particiona S_p através de divisões sucessivas gerando, a cada divisão, regiões de tolerância constituídas de um certo número de observações amostrais.

Formalmente, podemos descrever o método de Fraser da seguinte maneira:

Consideremos uma amostra de n observações, $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$, de $\underline{X}$. Definamos, em S_p , n funções reais e contínuas $\phi_1(\underline{x}), \dots,$

$\phi_n(\underline{x})$, não necessariamente distintas, denominadas funções de corte. Definamos uma sequência de números inteiros $\{k_1, \dots, k_n\}$ que seja uma permutação de 1, 2, ..., n. Esta sequência de números inteiros será utilizada para indicar os pontos onde S_p será particionado.

PASSO 1. Coloquemos em ordem crescente os vetores observados $\underline{x}_1, \dots, \underline{x}_n$, segundo $\phi_{k_1}(\underline{x})$, e denotemos por $\underline{x}^{(k_1)}$ o vetor correspondente ao k_1 -ésimo valor de $\phi_{k_1}(\underline{x})$ nesta ordenação.

Teremos dividido S_p em dois conjuntos:

$$B(1, \dots, k_1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) \leq \phi_{k_1}(\underline{x}^{(k_1)}) \right\} \text{ e}$$

$$B(k_1 + 1, \dots, n+1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) > \phi_{k_1}(\underline{x}^{(k_1)}) \right\}.$$

Em caso de empate, $\underline{x}^{(k_1)}$ será o vetor de menor índice entre os empatados. Este critério de escolha é arbitrário.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(k_1)}$ pertencente a $B(1, \dots, k_1)$, não será considerado.

PASSO 2. Se K_2 for índice do conjunto $B(1, \dots, k_1)$, coloquemos em ordem crescente os vetores amostrais deste conjunto, segundo $\phi_{k_2}(\underline{x})$, e denotemos por $\underline{x}^{(k_2)}$ o vetor amostral correspondente ao k_2 -ésimo valor de $\phi_{k_2}(\underline{x})$ nesta ordenação.

Teremos dividido S_p em três conjuntos:

$$B(1, \dots, K_2) = \left\{ \underline{x} \in S_p \mid \begin{aligned} &\phi_{k_1}(\underline{x}) \leq \phi_{k_1}(\underline{x}^{(k_1)}) \text{ e} \\ &\phi_{k_2}(\underline{x}) \leq \phi_{k_2}(\underline{x}^{(k_2)}) \end{aligned} \right\},$$

$$B(k_2+1, \dots, k_1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) \leq \phi_{k_1}(\underline{x}^{(k_1)}) \text{ e } \right. \\ \left. \phi_{k_2}(\underline{x}) > \phi_{k_2}(\underline{x}^{(k_2)}) \right\} \text{ e}$$

$$B(k_1+1, \dots, n+1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) > \phi_{k_1}(\underline{x}^{(k_1)}) \right\}.$$

Em caso de empate, $\underline{x}^{(k_2)}$ será o vetor de menor índice entre os empatados. Este critério de escolha é arbitrário.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(k_2)}$, pertencente a $B(1, \dots, k_2)$, não será considerado.

Se k_2 for índice do conjunto $B(k_1+1, \dots, n+1)$, coloquemos em ordem crescente os vetores amostrais deste conjunto, segundo $\phi_{k_2}(\underline{x})$, e denotemos por $\underline{x}^{(k_2)}$ o vetor amostral correspondente ao $(k_2 - k_1)$ -ésimo valor de $\phi_{k_2}(\underline{x})$ nesta ordenação.

Teremos os seguintes conjuntos:

$$B(1, \dots, k_1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) \leq \phi_{k_1}(\underline{x}^{(k_1)}) \right\},$$

$$B(k_1+1, \dots, k_2) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) > \phi_{k_1}(\underline{x}^{(k_1)}) \text{ e } \right. \\ \left. \phi_{k_2}(\underline{x}) \leq \phi_{k_2}(\underline{x}^{(k_2)}) \right\} \text{ e}$$

$$B(k_2+1, \dots, n+1) = \left\{ \underline{x} \in S_p \mid \phi_{k_1}(\underline{x}) > \phi_{k_1}(\underline{x}^{(k_1)}) \text{ e } \right. \\ \left. \phi_{k_2}(\underline{x}) > \phi_{k_2}(\underline{x}^{(k_2)}) \right\}.$$

Em caso de empate, $\underline{x}^{(k_2)}$ será o vetor de menor índice en-

tre os empatados. Este critério de escolha é arbitrário.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(k_2)}$, pertencente a $B(k_1+1, \dots, K_2)$, não será considerado.

Caso o número de vetores amostrais pertencentes a $B(k_1+1, \dots, n+1)$ seja menor que $(k_2 - k_1)$, o número inteiro k_2 e a função de corte $\phi_{k_2}(\underline{x})$ não serão utilizados, isto é, não particionaremos S_p através $\phi_{k_2}(\underline{x})$.

PASSO(m+1). O (m+1)-ésimo corte será realizado no $(1 < m \leq n+1)$

conjunto que tiver k_{m+1} como índice. Devemos colocar em ordem crescente as observações amostrais desse conjunto, segundo

$\phi_{k_{m+1}}(\underline{x})$, e denotar por $\underline{x}^{(k_{m+1})}$ o vetor amostral correspondente ao $(k_{m+1} - k_A + 1)$ -ésimo valor de $\phi_{k_{m+1}}(\underline{x})$ nesta ordenação, onde k_A é igual ao menor índice do conjunto.

O conjunto que tem k_{m+1} como índice será substituído por sua intersecção com

$$B(k_A, \dots, k_{m+1}) = \left\{ \underline{x} \in S_p \mid \phi_{k_{m+1}}(\underline{x}) \leq \phi_{k_{m+1}}(\underline{x}^{(k_{m+1})}) \right\} \text{ e}$$

$$B(k_{m+1}+1, \dots, k_B) = \left\{ \underline{x} \in S_p \mid \phi_{k_{m+1}}(\underline{x}) > \phi_{k_{m+1}}(\underline{x}^{(k_{m+1})}) \right\},$$

onde

k_B é o maior índice do conjunto.

Em caso de empate, $\underline{x}^{(k_{m+1})}$ será o vetor de menor índice entre os empatados. Este critério de escolha é arbitrário.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(k_{m+1})}$, pertencente a $B(k_A, \dots, k_{m+1})$, não será considerado.

Caso o número de vetores amostrais pertencentes a $B(k_A, \dots, k_{m+1}, \dots, k_B)$ seja menor que $(k_{m+1} - k_A + 1)$, o número inteiro k_{m+1} e a função de corte $\phi_{k_{m+1}}(\underline{x})$ não serão utilizados, isto é, não particionaremos S_p através de $\phi_{k_{m+1}}(\underline{x})$.

* * *

Se não ocorrerem empates em relação a cada uma das n funções de corte $\phi_1(\underline{x}), \dots, \phi_n(\underline{x})$, após n passos teremos dividido S_p em $(n+1)$ conjuntos $B(1), \dots, B(n+1)$ tais que $B(1), \dots, B(n)$ contêm, cada um, uma observação amostral e $B(n+1)$ nenhuma observação amostral. Estes conjuntos são denominados na literatura pesquisada, *blocos estatisticamente equivalentes*.

Se ocorrerem empates em relação a pelo menos uma das funções de corte $\phi_1(\underline{x}), \phi_2(\underline{x}), \dots, \phi_n(\underline{x})$, podemos encontrar uma de duas situações:

a) S_p será particionado em $(n+1)$ conjuntos $B(1), \dots, B(n+1)$, mas entre os conjuntos $B(1), \dots, B(n)$ teremos pelo menos um contendo nenhuma observação amostral e pelo menos um contendo mais de uma observação amostral.

b) S_p será particionado num número menor que $(n+1)$ de conjuntos, isto é, não utilizaremos todos os números inteiros da sequência $\{k_1, \dots, k_n\}$, nem todas as n funções de corte $\phi_{k_i}(\underline{x})$ definidas. Entre os conjuntos construídos pelo menos um conterá nenhuma observação amostral e pelo menos um conterá mais de uma observação amostral.

Observemos que o caso (a) está incluído no caso (b) se, no caso (a) unirmos os conjuntos vazios construídos.

Para construirmos $1 < r < (n+1)$, $n > 1$, regiões de tolerância, basta fazermos a união de blocos estatísticos equivalentes, o que equivale a aplicarmos o método de Fraser supondo a existência de um número $(r-1)$ de funções de corte. Quando $n=1$, temos um único número inteiro k_1 e uma única função de corte $\phi_{k_1}(\underline{x})$. Nesta situação, particionamos S_p em duas regiões de tolerância que são blocos estatisticamente equivalentes,

Exemplo. Para ilustrar a teoria exposta sobre o método de Fraser daremos agora um exemplo numérico. Este basear-se-á nos dados apresentados por Fisher (1936).

A tabela que segue apresenta quatro medidas, comprimento do caule (CC), largura do caule (LC), comprimento da pétala (CP) e largura da pétala (LP), para uma amostra de oito elementos de cada um de dois tipos de Iris (planta), a saber Iris Versicolor e Iris Virginica.

Tabela. Medidas associadas às plantas do tipo Iris Versicolor e Iris Virginica

Iris Versicolor				Iris Virginica			
CC	LC	CP	LP	CC	LC	CP	LP (1)
7,0	3,2	4,7	1,4	6,3	3,3	6,0	2,5
6,4	3,2	4,5	1,5	5,8	2,7	5,1	1,9
6,9	3,1	4,9	1,5	7,1	3,0	5,9	2,1
5,5	2,3	4,0	1,3	6,3	2,9	5,6	1,8
6,5	2,8	4,6	1,5	6,5	3,0	5,8	2,2
5,7	2,8	4,5	1,3	7,6	3,0	6,6	2,1
6,3	3,3	4,7	1,6	4,9	2,5	4,5	1,7
4,9	2,4	3,3	1,0	7,3	2,9	6,3	1,8

- (1) CC : Comprimento do caule
LC : Largura do caule
CP : Comprimento da pétala
LP : Largura da pétala

Tomemos, como exemplo, as medidas comprimento do caule e largura do caule associadas às oito plantas do tipo Iris Versicolor.

Temos $n=8$ e $p=2$. Seja $\{\underline{x}_1, \dots, \underline{x}_8\}$ a amostra considerada, onde $\underline{x}' = (x_1, x_2)$, com x_1 denotando o comprimento do caule e x_2 a largura do caule.

A sequência de inteiros $\{k_1, \dots, k_8\}$, escolhida arbitrariamente, é dada por $\{3, 6, 1, 2, 4, 5, 7, 8\}$ e as funções de corte, também escolhidas arbitrariamente, são dadas por $\phi_3(\underline{x}) = \phi_6(\underline{x}) = x_1$ e $\phi_i(\underline{x}) = x_2$ para $i \neq 3, 6$.

PASSO 1. Colocamos em ordem crescente os vetores observados $\underline{x}_1, \dots, \underline{x}_8$, segundo $\phi_3(\underline{x}) = x_1$, e denotamos por $\underline{x}^{(3)}$, o vetor correspondente ao terceiro valor de x_1 nesta ordenação. Temos $\underline{x}^{(3)} = \underline{x}_6 = \begin{bmatrix} 5,7 \\ 2,8 \end{bmatrix}$.

Dividimos S_p em dois conjuntos:

$$B(1, 2, 3) = \{\underline{x} \in S_p \mid x_1 \leq 5,7\} \text{ e}$$

$$B(4, 5, 6, 7, 8, 9) = \{\underline{x} \in S_p \mid x_1 > 5,7\}$$

Além do vetor $\underline{x}^{(3)} = \underline{x}_6$, temos que os vetores amostrais $\underline{x}_4$ e $\underline{x}_8$ pertencem a $B(1, 2, 3)$ e os vetores $\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_5, \underline{x}_7$ per-

tencem a $B(4,5,6,7,8,9)$.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(3)}$ não será considerado.

PASSO 2. Como $k_2=6$ é índice do conjunto $B(4,5,6,7,8,9)$, colocamos em ordem crescente os cinco vetores amostrais deste conjunto, segundo $\phi_6(\underline{x}) = x_1$, e denotamos por $\underline{x}^{(6)}$ o vetor correspondente ao terceiro valor de x_1 nesta ordenação. Temos $\underline{x}^{(6)} = \underline{x}_5 = \begin{bmatrix} 6,5 \\ 2,8 \end{bmatrix}$.

Dividimos S_p em três conjuntos:

$$B(1,2,3) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \},$$

$$B(4,5,6) = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,5 \} \text{ e}$$

$$B(7,8,9) = \{ \underline{x} \in S_p \mid x_1 > 6,5 \}.$$

Os vetores amostrais $\underline{x}^{(3)} = \underline{x}_6$, $\underline{x}_4$ e $\underline{x}_8$ pertencem a $B(1,2,3)$, os vetores $\underline{x}^{(6)} = \underline{x}_5$, $\underline{x}_2$ e $\underline{x}_7$ pertencem a $B(4,5,6)$ e os vetores $\underline{x}_1$ e $\underline{x}_3$ pertencem a $B(7,8,9)$.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(6)}$ não será considerado.

PASSO 3. Como $k_3 = 1$, o terceiro corte foi realizado no conjunto $B(1,2,3)$. Colocamos em ordem crescente as observações amostrais desse conjunto, segundo $\phi_1(\underline{x}) = x_2$, e denotamos por $\underline{x}^{(1)}$ o vetor correspondente ao menor valor de x_2 nesta ordenação. Temos $\underline{x}^{(1)} = \underline{x}_4 = \begin{bmatrix} 5,5 \\ 2,3 \end{bmatrix}$.

Dividimos S_p em quatro conjuntos:

$$B(1) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 \leq 2,3 \},$$

$$B(2,3) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 > 2,3 \},$$

$$B(4,5,6) = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,5 \} \text{ e}$$

$$B(7,8,9) = \{ \underline{x} \in S_p \mid x_1 > 6,5 \}.$$

O vetor $\underline{x}^{(1)} = \underline{x}_4$ pertence a $B(1)$, os vetores $\underline{x}^{(3)} = \underline{x}_6$ e $\underline{x}_8$ pertencem a $B(2,3)$, os vetores $\underline{x}^{(6)} = \underline{x}_5$, $\underline{x}_2$ e $\underline{x}_7$ pertencem a $B(4,5,6)$ e os vetores $\underline{x}_1$ e $\underline{x}_3$ pertencem a $B(7,8,9)$.

Para a determinação de futuros cortes em S_p , o vetor $\underline{x}^{(1)}$ não será considerado.

Após oito passos dividimos S_p em nove conjuntos mutuamente exclusivos:

$$B(1) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 \leq 2,3 \},$$

$$B(2) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } 2,3 < x_2 \leq 2,4 \},$$

$$B(3) = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 > 2,4 \},$$

$$B(4) = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,5 \text{ e } x_2 \leq 3,2 \},$$

$$B(5) = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,5 \text{ e } 3,2 < x_2 \leq 3,3 \},$$

$$B(6) = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,5 \text{ e } x_2 > 3,3 \},$$

$$B(7) = \{ \underline{x} \in S_p \mid x_1 > 6,5 \text{ e } x_2 \leq 3,1 \},$$

$$B(8) = \{ \underline{x} \in S_p \mid x_1 > 6,5 \text{ e } 3,1 < x_2 \leq 3,2 \} \text{ e}$$

$$B(9) = \{ \underline{x} \in S_p \mid x_1 > 6,5 \text{ e } x_2 > 3,2 \} .$$

Observemos que o vetor $\underline{x}_4$ pertence a $B(1)$ e localiza-se na fronteira entre $B(1)$ e $B(2)$; o vetor $\underline{x}_8$ pertence a $B(2)$ e localiza-se na fronteira entre $B(2)$ e $B(3)$; o vetor $\underline{x}_6$ pertence a $B(3)$ e localiza-se na fronteira entre $B(3)$ e $B(4)$; os vetores $\underline{x}_2$ e $\underline{x}_5$ pertencem a $B(4)$, sendo que $\underline{x}_2$ localiza-se na fronteira entre $B(4)$ e $B(5)$ e $\underline{x}_5$, na fronteira entre $B(4)$ e $B(7)$; $\underline{x}_7$ pertence a $B(5)$ e localiza-se na fronteira entre $B(5)$ e $B(6)$; $\underline{x}_3$ pertence a $B(7)$ e localiza-se na fronteira entre $B(7)$ e $B(8)$ e $\underline{x}_1$ pertence a $B(8)$, localizando-se na fronteira entre $B(8)$ e $B(9)$.

Os conjuntos $B(6)$ e $B(9)$ não contêm observações amostrais. O fato de $B(6)$ não conter observações amostrais deve-se à ocorrência de empates na coordenada x_2 dos vetores amostrais.

2.3 ESCOLHA DAS FUNÇÕES DE CORTE $\phi_i(\underline{x})$ E DA SEQUÊNCIA $\{k_i\}$ DE NÚMEROS INTEIROS

A aplicação do método de Fraser depende da escolha das funções de corte ϕ_i e da sequência $\{k_i\}$ de números inteiros. Existem diversas maneiras de realizarmos estas escolhas, uma vez que podemos utilizar qualquer uma das possíveis permutações dos números inteiros que constituem a sequência $\{k_i\}$ e qualquer função real e contínua do vetor $\underline{x}$ como função de corte.

Apresentaremos os dois métodos de realizar tais escolhas

mais encontrados na literatura.

2.3.1. MÉTODO DE SUBPARTIÇÃO DE REGIÕES DE TOLERÂNCIA

As funções de corte $\phi_i(\underline{x})$ e a sequência $\{k_i\}$ de números inteiros devem ser escolhidas de modo que obtenhamos regiões de tolerância contendo, aproximadamente, o mesmo número "m" de observações amostrais. Para isto, deve ser obedecido o seguinte esquema:

ETAPA INICIAL. Particionamos S_p em um número L (inteiro e maior que 1) de regiões de tolerância, onde cada região de tolerância construída contém, aproximadamente, o mesmo número de observações amostrais.

ETAPAS SEGUINTE. Subparticionamos cada uma das regiões de tolerância obtidas anteriormente em L novas sub-regiões, cada uma contendo, aproximadamente, o mesmo número de observações amostrais.

TÉRMINO DO PROCESSO. O processo de subpartição prossegue até obter-se sub-regiões de tolerância que contenham, aproximadamente, o mesmo número "m" de observações amostrais.

Na literatura sobre regiões de tolerância, os autores utilizam o método de subpartição de regiões de tolerância escolhendo as funções de corte $\phi_i(\underline{x})$ e a sequência $\{k_i\}$ de números inteiros através de exemplos numéricos. Procuramos, neste trabalho, formalizar o método de determinação de tais funções

e sequência de números inteiros, da seguinte maneira:

Seja q o número de vezes em que cada região de tolerância, a partir de S_p , é dividida em L novas regiões. Denotemos por $[y]$ o maior inteiro menor ou igual a y .

Através do método de subpartição de regiões de tolerância construiremos, no total, L^q regiões de tolerância.

Se o valor de L for fixado a priori, cada uma das L^q regiões de tolerância conterá, aproximadamente, m observações amostrais, onde

$$m = \left[\frac{n}{L^q} \right] \quad (2.3.1)$$

Se o valor de m for fixado a priori temos o valor de L dado por

$$L = \left[\left(\frac{n+1}{m} \right)^{1/q} \right] \quad (2.3.2)$$

Para obter as L^q regiões de tolerância necessárias apliquemos o método de Fraser. Na j -ésima etapa de subpartição, $j = 1, \dots, q$, obteremos L^j sub-regiões de tolerância, utilizando $L^{j-1}(L-1)$ números inteiros (as L^j sub-regiões de tolerância são obtidas particionando-se cada uma das L^{j-1} sub-regiões, obtidas na $(j-1)$ -ésima etapa, em L novas sub-regiões através de $(L-1)$ números inteiros) dados, sequencialmente, por

$$\left[\frac{i(n+1)}{L^j} \right], \quad i = 1, \dots, (L^j-1) \quad (2.3.3)$$

e por L^{j-1} ($L-1$) funções de corte dadas pela j -ésima coordenada de $\underline{x}$. Observemos que se $q > p$, repetimos as coordenadas de $\underline{x}$, como funções de corte, após todas as p terem sido utilizadas uma vez.

Quando a aplicação da expressão (2.3.3) resultar em números inteiros repetidos, as repetições devem ser desconsideradas.

A figura abaixo ilustra a situação em que temos $L=3$, $q=2$ e $p=2$.

Para a construção das regiões de tolerância S_1 , S_2 e S_3 , utilizamos dois números inteiros e duas funções de corte dadas pela primeira coordenada de $\underline{x}$. Para a construção das sub-regiões de tolerância $S_4, S_5, S_6, S_7, S_8, S_9, S_{10}, S_{11}$ e S_{12} utilizamos seis números inteiros e seis funções de corte dadas pela segunda coordenada de $\underline{x}$.

Se $(n+1) = L^q$ e $m=1$, construiremos blocos estatística-

mente equivalentes (na ausência de empates) pois a sequência de inteiros $\{k_i\}$, obtida através de (2.3.3), terá n inteiros, isto é, será uma permutação dos inteiros $1, 2, \dots, n$.

Se $(n+1) \neq L^q$ e $m=1$, construiremos blocos estatisticamente equivalentes (na ausência de empates) aumentando o número de etapas do processo de subpartição, isto é, adotando $j = 1, \dots, q, q+1, \dots$ etapas de subpartição, de modo que a sequência de inteiros $\{k_i\}$, obtida através de (2.3.3), seja uma permutação dos inteiros $1, 2, \dots, n$.

Observemos que na construção do método de subpartição de regiões de tolerância, utilizamos, nas expressões (2.3.2) e (2.3.3), no lugar de n , como era esperado, o valor $(n+1)$. A utilização deste valor facilita:

- a partição de S_p em regiões de tolerância contendo, aproximadamente, o mesmo número " m " de observações amostrais;
- a determinação da sequência de números inteiros $\{k_i\}$.

Exemplo. Suponhamos $n=8$, $p=2$ e $m=1$ (situação apresentada no exemplo da seção 2.2). Seja $q=2$.

$$\text{Temos } L = \left\lceil \left(\frac{n+1}{m} \right)^{1/q} \right\rceil = \lceil 3 \rceil = 3.$$

Observemos que, neste caso, $(n+1) = L^q$, isto é, a sequência de números inteiros $\{k_i\}$ será uma permutação dos inteiros $1, 2, \dots, 8$.

Para $j = 1$, temos:

$$k_1 = \left[\frac{1 \cdot (n+1)}{3} \right] = [3] = 3,$$

$$k_2 = \left[\frac{2 \cdot (n+1)}{3} \right] = [6] = 6 \text{ e}$$

$$\phi_3(\underline{x}) = \phi_6(\underline{x}) = x_1.$$

Para $j = 2$, temos:

$$k_3 = \left[\frac{1 \cdot (n+1)}{9} \right] = [1] = 1, \quad k_4 = \left[\frac{2 \cdot (n+1)}{9} \right] = [2] = 2,$$

$$k_5 = \left[\frac{4 \cdot (n+1)}{9} \right] = [4] = 4, \quad k_6 = \left[\frac{5 \cdot (n+1)}{9} \right] = [5] = 5,$$

$$k_7 = \left[\frac{7 \cdot (n+1)}{9} \right] = [7] = 7, \quad k_8 = \left[\frac{8 \cdot (n+1)}{9} \right] = [8] = 8 \text{ e}$$

$$\phi_1(\underline{x}) = \phi_2(\underline{x}) = \phi_4(\underline{x}) = \phi_5(\underline{x}) = \phi_7(\underline{x}) = \phi_8(\underline{x}) = x_2.$$

2.3.2. MÉTODO DE TUKEY (1947)

Para construir uma região de tolerância contendo m observações amostrais devemos particionar S_p em $(n-m+1)$ regiões de tolerância, de modo que as $(n-m)$ primeiras sejam blocos estatisticamente equivalentes e a última seja a região de tolerância desejada.

Para tanto, deve ser obedecido o seguinte esquema:

Apliquemos o método de Fraser utilizando $(n-m)$ números inteiros dados por $k_1=1, k_2=2, \dots, k_{n-m}=n-m$ e $(n-m)$ funções de corte dadas pelas coordenadas do vetor $\underline{x}$ e por funções destas

coordenadas. Se $p=2$, temos a seguinte sequência de funções de corte:

$$\{x_1, x_2, -x_1, -x_2, x_1+x_2, x_1-x_2, -x_1+x_2, -x_1-x_2, x_1, x_2, \dots\}.$$

Se $p > 2$ as funções de corte são determinadas de maneira análoga às utilizadas quando $p = 2$. Teremos, em cada sequência, $(2p + 2^p)$ funções de corte distintas.

Exemplo. Consideremos os mesmos vetores amostrais do exemplo dado na seção 2.2.

Queremos construir uma região de tolerância contendo $m=5$ observações amostrais. Para isso, utilizaremos três números inteiros dados por $k_1=1$, $k_2=2$ e $k_3=3$ e três funções de corte dadas por $\phi_1(\underline{x}) = x_1$, $\phi_2(\underline{x}) = x_2$ e $\phi_3(\underline{x}) = -x_1$.

Aplicando o método de Fraser dividimos S_p em quatro conjuntos mutuamente exclusivos:

$$B(1) = \{\underline{x} \in S_p \mid x_1 \leq 4,9\},$$

$$B(2) = \{\underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 \leq 2,3\},$$

$$B(3) = \{\underline{x} \in S_p \mid x_1 \geq 7,0 \text{ e } x_2 > 2,3\}$$

$$B(4, \dots, 9) = \{\underline{x} \in S_p \mid 4,9 < x_1 < 7,0 \text{ e } x_2 > 2,3\}.$$

Observemos que o vetor $\underline{x}_8$ pertence a $B(1)$, o vetor $\underline{x}_4$ pertence a $B(2)$, o vetor $\underline{x}_1$ pertence a $B(3)$ e os vetores $\underline{x}_2, \underline{x}_3, \underline{x}_4, \underline{x}_5, \underline{x}_6$ e $\underline{x}_7$ pertencem a $B(4, \dots, 9)$. As três primeiras regiões de to-

lerância construídas, $B(1)$, $B(2)$ e $B(3)$, são blocos estatisticamente equivalentes (contêm uma observação amostral) e a última, $B(4, \dots, 9)$, é a região de tolerância desejada.

2.4. PROBABILIDADES ASSOCIADAS ÀS REGIÕES DE TOLERÂNCIA

DEFINIÇÃO 1. Seja S uma região de tolerância definida através da amostra $\{\underline{x}_1, \dots, \underline{x}_n\}$ de $\underline{X}$ e P_S , a probabilidade de que uma nova observação amostral $\underline{x}^*$ pertença a S . Temos

$$P_S = \int_S f(\underline{x}) \, d\underline{x} \quad (2.4.1)$$

Como S é função da amostra, P_S será uma variável aleatória.

DEFINIÇÃO 2. Uma região de tolerância, S , é denominada região de tolerância não paramétrica, se a função densidade de probabilidade de P_S , é independente de f .

Uma vez exposto o método de Fraser (seção 2.2) para construção de regiões de tolerância, podemos enunciar os seguintes teoremas:

TEOREMA 1. Se $\phi_1(\underline{x}), \dots, \phi_n(\underline{x})$ têm, cada uma, distribuição contínua, então as probabilidades $P_{B(1)}, \dots, P_{B(n+1)}$ dos $(n+1)$ blocos estatisticamente equivalentes construídos, através da amostra $\{\underline{x}_1, \dots, \underline{x}_n\}$ de $\underline{x}$, têm distribuição conjunta uniforme. $\square$

A demonstração deste teorema encontra-se em Fraser (1957).

TEOREMA 2. Seja S_m uma região de tolerância formada pela união

de m blocos estatisticamente equivalentes. A probabilidade P_{S_m} de S_m é a soma das probabilidades dos m blocos que a compõem. Além disso, P_{S_m} é uma variável aleatória com distribuição beta de parâmetros m e $(n-m+1)$. $\square$

A demonstração de que P_{S_m} tem distribuição beta de parâmetros m e $(n-m+1)$ está em Fraser (1957) e Wilks (1962).

No capítulo 3 veremos que Patrick e Fischer (1970) utilizam os resultados do Teorema 2 para obter a probabilidade total de erro de classificação de seu método não paramétrico de discriminação.

Através do Teorema 2 notamos que S_m é uma região de tolerância não paramétrica, pois a função densidade de probabilidade de P_{S_m} é independente de f , qualquer que seja f .

Uma vez conhecida a distribuição de P_{S_m} podemos calcular $P(P_{S_m} \geq \beta)$, isto é, calcular a probabilidade de S_m conter, no mínimo, $100\beta\%$ de observações amostrais ou, então, fixada uma probabilidade γ , calcular a porcentagem mínima β de observações amostrais que devem pertencer a S_m .

Murphy (1948) construiu gráficos para β , fixando três níveis de tolerância $\gamma = 0,90; 0,95$ e $0,99$. Nestas construções, foram fixados os valores de $(n-m+1)$ (número de blocos que não pertencem à região de tolerância S_m) e de n (tamanho da amostra) que variou de $(n-m+1)$ até 500. Uma aplicação dos gráficos de Murphy aparece quando fixamos γ , β e n , pois nesse caso obtemos o valor de $(n-m+1)$, isto é, determinamos quantos

blocos não pertencem à região de tolerância e,consequentemente, o tamanho m dessa região.

No capítulo seguinte, veremos que Quesenberry e Gessaman (1968) construíram um método de discriminação não paramétrico baseado em regiões de tolerância utilizando o método de Tukey (1947) de construção de regiões de tolerância,sendo que na determinação do número m de observações amostrais pertencentes à região de tolerância construída foram utilizados os gráficos de Murphy.

CAPÍTULO 3

MÉTODOS DE DISCRIMINAÇÃO NÃO PARAMÉTRICOS BASEADOS EM REGIÕES DE TOLERÂNCIA

Dividiremos os métodos de discriminação não paramétricos baseados em regiões de tolerância em dois grupos:

MÉTODOS DE DISCRIMINAÇÃO FORÇADA

Neste grupo situam-se os métodos de discriminação de Anderson (1966), Patrick e Fischer (1970) e Gessaman e Gessaman (1972).

MÉTODOS DE DISCRIMINAÇÃO PARCIAL

Neste grupo incluímos os métodos de discriminação de Gessaman e Gessaman (1972) e de Quesenberry e Gessaman (1968).

Neste capítulo, apresentaremos os métodos de discriminação não paramétricos baseados em regiões de tolerância, apontando suas propriedades, vantagens e desvantagens e comparando tais métodos em função de seus procedimentos de construção e das estimativas de suas probabilidades de erro de classificação (discriminação forçada e parcial) e de não classificação (discriminação parcial).

3.1. MÉTODOS DE DISCRIMINAÇÃO FORÇADA

Os métodos não paramétricos de discriminação forçada ba-

seiam-se na utilização de estimadores de $f_i(\underline{x})$, $i = 1, 2$, na regra (ótima) de discriminação paramétrica da razão de verossimilhança (Welch, 1939).

3.1.1. REGRAS DISCRIMINATÓRIAS - SÍNTESE

Procuramos, nesta seção, uniformizar os principais métodos de discriminação forçada, encontrados na literatura pesquisada, através da seguinte forma geral:

PARTIÇÃO DE S_p . Particionamos S_p segundo o método de subpartição de regiões de tolerância, apresentado na seção 2.3.1 do capítulo 2, utilizando:

Primeira alternativa: construímos uma (única) partição de S_p através da amostra $\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}\}$ de π_1 ou da amostra $\{\underline{x}_{21}, \dots, \underline{x}_{2n_2}\}$ de π_2 ou, da amostra conjunta $\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}, \underline{x}_{21}, \dots, \underline{x}_{2n_2}\}$ de π_1 e π_2 .

Segunda alternativa: construímos duas partições de S_p , sendo uma através da amostra $\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}\}$ de π_1 e a outra através da amostra $\{\underline{x}_{21}, \dots, \underline{x}_{2n_2}\}$ de π_2 .

ESTIMAÇÃO DE $f_i(\underline{x})$, $i=1, 2$ E REGRA DISCRIMINATÓRIA

Partição única de S_p (primeira alternativa)

Denotamos por $S_{\underline{x}}$ a região de tolerância que contém o vetor $\underline{x}$, associado ao elemento E a ser classificado em π_1 ou π_2 .

Contamos o número N_i de observações amostrais de π_i , $i=1, 2$,

encontradas em $S_{\underline{x}}$.

O estimador de $f_i(\underline{x})$, $i=1,2$, é dado por

$$\hat{f}_i(\underline{x}) = \frac{N_i}{n_i} \cdot \frac{1}{V_{S_{\underline{x}}}}, \quad (3.1.1)$$

onde $V_{S_{\underline{x}}}$ denota o "volume" de $S_{\underline{x}}$.

Uma vez obtido $\hat{f}_i(\underline{x})$, $i = 1,2$, este estimador é substituído na regra (ótima) de discriminação paramétrica da razão de verossimilhança.

Obtemos:

Se $p_1 \frac{N_1}{n_1} > p_2 \frac{N_2}{n_2}$, classificamos o elemento E em π_1

Se $p_1 \frac{N_1}{n_1} < p_2 \frac{N_2}{n_2}$, classificamos o elemento E em π_2

(3.1.2)

Como já foi definido no capítulo 1, p_i é a probabilidade a priori do elemento E pertencer à população π_i , $i=1,2$.

Duas partições de S_p (segunda alternativa)

Denotamos por $S_{1\underline{x}}$ a região de tolerância à qual $\underline{x}$ pertence, construída através da amostra de π_1 e por $S_{2\underline{x}}$ a região de tolerância à qual $\underline{x}$ pertence, construída através da amostra de π_2 .

Contamos o número N_i observações de π_i encontradas em

$S_{i\bar{x}}$, $i = 1, 2$.

O estimador de $f_i(\underline{x})$, $i = 1, 2$, é dado por:

$$\hat{f}_i(\underline{x}) = \frac{N_i}{n_i} \cdot \frac{1}{V_{S_{i\bar{x}}}}, \quad (3.1.3)$$

onde $V_{S_{i\bar{x}}}$, denota o "volume" de $S_{i\bar{x}}$, $i = 1, 2$.

Substituindo o estimador dado em (3.1.3) na regra (ótima) de discriminação paramétrica da razão de verossimilhança, obtemos:

Se $p_1 \frac{N_1}{n_1} \cdot \frac{1}{V_{S_{1\bar{x}}}} > p_2 \frac{N_2}{n_2} \cdot \frac{1}{V_{S_{2\bar{x}}}}$, classificamos o elemento E em π_1

Se $p_1 \frac{N_1}{n_1} \cdot \frac{1}{V_{S_{1\bar{x}}}} < p_2 \frac{N_2}{n_2} \cdot \frac{1}{V_{S_{2\bar{x}}}}$, Classificamos o elemento E em π_2 (3.1.4)

PROBABILIDADE A PRIORI. Observemos que se p_i for desconhecida e não puder ser estimada (amostras separadas), será substituída por $p_i = 0,5$, $i=1,2$; se p_i for desconhecida e puder ser estimada (amostras misturadas), será substituída por $p_i = \frac{n_i}{n}$, onde $n = \sum n_i, i=1,2$.

IGUALDADES. Se, para a regra discriminante dada em (3.1.2), $N_1 = N_2$ e $p_1 \frac{N_2}{n_2} = p_2 \frac{N_2}{n_2}$, ou para a regra discriminante dada em (3.1.4),

$N_1 = N_2$ e $p_1 \frac{N_1}{n_1} \cdot \frac{1}{V_{S_{1\bar{x}}}} = p_2 \frac{N_2}{n_2} \cdot \frac{1}{V_{S_{2\bar{x}}}}$, podemos solucionar estas igualdades examinando as observações amostrais das regiões

de tolerância adjacentes àquela(s) que contém $\underline{x}$ e redefinindo $N_i = N'_i$ como o número total de observações amostrais π_i encontradas nas regiões examinadas, incluindo as daquela(s) que contém $\underline{x}$. Nestes casos, V_{S_x} ou $V_{S_{ix}}$ serão redefinidos pela soma dos volumes das regiões de tolerância examinadas, incluindo o volume de S_x ou de S_{ix} , $i = 1, 2$, respectivamente. Se encontrarmos $N'_1 \neq N'_2$, classificamos o elemento E através da substituição desses novos valores de N_1 e N_2 na regra discriminante dada em (3.1.2) e/ou (3.1.4); se $N'_1 = N'_2$, o elemento E será classificado aleatoriamente em π_1 ou π_2 .

Por outro lado, se, para a regra discriminante em (3.1.2), $N_1 \neq N_2$ e $p_1 \frac{N_1}{n_1} = p_2 \frac{N_2}{n_2}$ ou para a regra discriminante dada em (3.1.4), $N_1 \neq N_2$ e $p_1 \frac{N_1}{n_1} \cdot \frac{1}{V_{S_{1x}}} = p_2 \frac{N_2}{n_2} \cdot \frac{1}{V_{S_{2x}}}$, classificamos o elemento E aleatoriamente em π_1 ou π_2 .

3.1.2. DESCRIÇÃO DOS PRINCIPAIS MÉTODOS ENCONTRADOS NA LITERATURA

Os métodos de Anderson (1966) e o de Gessaman e Gessaman (1972), apresentados nesta seção, utilizam estimadores de $f_i(\underline{x})$, $i = 1, 2$, dados pela expressão (3.1.1) e regra discriminante dada em (3.1.2); o método de Patrick e Fischer (1970) utiliza os estimadores $\hat{f}_i(\underline{x})$, $i = 1, 2$, dados pela expressão (3.1.3) e a regra discriminante dada em (3.1.4).

A diferença entre estes métodos está na maneira pela qual particionam S_p , segundo o método de subpartição de regiões de

tolerância, apresentado na seção 2.3.1 do capítulo 2. Convém observar que todos os métodos apresentados fixam $p = q$, onde q é o número de etapas de subpartição, pois dessa forma utilizam como funções de corte todas as coordenadas do vetor $\underline{x}$, uma em cada etapa de subpartição (na construção de blocos estatisticamente equivalentes, estas coordenadas podem vir a ser repetidas).

MÉTODO 1 DE ANDERSON (1966)

Particionamos S_p em blocos estatisticamente equivalentes, utilizando, conjuntamente, a amostra

$$\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}, \underline{x}_{21}, \dots, \underline{x}_{2n_2}\} \text{ de } \pi_1 \text{ e } \pi_2.$$

N_i será o número de observações amostrais de π_i , $i=1,2$, pertencentes a $S_{\underline{x}}$ ou situadas em sua fronteira, caso $S_{\underline{x}}$ não contenha observações amostrais.

MÉTODO 2 DE ANDERSON (1966)

Utilizando, conjuntamente, a amostra

$\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}, \underline{x}_{21}, \dots, \underline{x}_{2n_2}\}$ de π_1 e π_2 , particionamos S_p em regiões de tolerância, de modo que cada uma contenha, aproximadamente, um número " m ", ímpar e maior que um, de observações amostrais.

MÉTODO DE GESSAMAN E GESSAMAN (1972)

Através da amostra $\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}\}$ de π_1 , particionamos

S_p em L_1 regiões de tolerância, onde

$$L_1 = [n_1^{1/(q+1)}], \quad (3.1.5)$$

tais que cada uma contenha, aproximadamente,

$$m_1 = \left[\frac{n_1}{L_1^q} \right] = \left[\frac{n_1}{[n_1^{1/(q+1)}]^q} \right] \quad (3.1.6)$$

observações amostrais de π_1 (vide seção 2.3.1 do capítulo 2).

N_i será o número de observações amostrais de $\pi_i, i = 1, 2$, pertencentes a S_x .

Observemos que se utilizássemos a amostra $\{x_{21}, \dots, x_{2n_2}\}$ de π_2 , obteríamos, de maneira análoga, outra partição de S_p que, de modo geral, é diferente daquela obtida através da amostra de π_1 . Nesta situação o resultado da discriminação baseada na partição de S_p através da amostra de π_1 , pode diferir do resultado da discriminação baseada na partição de S_p através da amostra de π_2 .

MÉTODO DE PATRICK FISCHER (1970)

Através da amostra $\{x_{11}, \dots, x_{1n_1}\}$ de π_1 , particionamos S_p em L_1^q regiões de tolerância, onde

$$L_1 = [n_1^{1/(q+1)}], \quad (3.1.7)$$

tais que cada uma contenha, aproximadamente,

$$m_1 = \left[\frac{n_1}{L_1^q} \right] = \left[\frac{n_1}{[n_1^{1/(q+1)}]_1^q} \right] \quad (3.1.8)$$

observações amostrais de π_1 , (vide seção 2.3.1 do capítulo 2).

N_1 será o número de observações amostrais de π_1 pertencentes a $S_{1\bar{x}}$.

De modo análogo, particionamos S_p através da amostra $\{x_{21}, \dots, x_{2n_2}\}$ de π_2 .

N_2 será o número de observações amostrais de π_2 pertencentes a $S_{2\bar{x}}$.

3.1.3. CONSIDERAÇÕES

- os métodos apresentados na seção 3.1.2 baseiam-se no princípio do "vizinho mais próximo", isto é, a estimação de $f_i(\underline{x})$ baseia-se na contagem do número de observações amostrais de π_i , $i = 1, 2$, encontradas na região de tolerância que contém $\underline{x}$ ("vizinhos de $\underline{x}$ ").

- estes métodos diferem quanto ao número de observações amostrais pertencentes a cada região de tolerância construída. Com referência a este aspecto, notamos que o método de Gessaman e Gessaman e o de Patrick e Fischer particionam S_p em regiões de tolerância contendo um número m_i de observações amostrais que é função do tamanho n_i , $i=1, 2$, das amostras. Pela definição de m_i , verificamos que se n_i é grande, m_i também é grande, mas bem menor que n_i . cremos que qualquer outra função real de n_i poderia ser utilizada na definição de m_i . A que a-

presentamos, sugerida por Gessaman (1970), é a comumente utilizada na literatura pesquisada.

- em termos computacionais verificamos que os métodos mais trabalhosos são o método 1 de Anderson, pois requer a construção de todos os possíveis blocos estatisticamente equivalentes e o método de Patrick e Fischer, pois particiona S_p duas vezes, primeiramente através da amostra de π_1 e depois através da amostra de π_2 , e requer, na construção da regra discriminante dada pela expressão (3.1.4), o cálculo do volume das regiões de tolerância que contêm o vetor $\underline{x}$.

3.1.4. UM EXEMPLO NUMÉRICO

Para ilustrar a teoria sobre métodos não paramétricos de discriminação forçada baseados em regiões de tolerância, exposta nos itens anteriores, consideramos um exemplo numérico baseado nos dados da tabela^(*) apresentada a seguir.

Tabela. Medidas associadas às plantas do tipo Iris Versicolor e Iris Virginica

Iris Versicolor				Iris Virginica			
CC	LC	CP	LP	CC	LC	CP	LP ⁽¹⁾
7,0	3,2	4,7	1,4	6,3	3,3	6,0	2,5
6,4	3,2	4,5	1,5	2,7	5,1	1,9	2,5
6,9	3,1	4,9	1,5	7,1	3,0	5,9	2,1
5,5	2,3	4,0	1,3	6,3	2,9	5,6	1,8
6,5	2,8	4,6	1,5	6,5	3,0	5,8	2,2
5,7	2,8	4,5	1,3	7,6	3,0	6,6	2,1
6,3	3,3	4,7	1,6	4,9	2,5	4,5	1,7
4,9	2,4	3,3	1,0	7,3	2,9	6,3	1,8

(*) Fonte: Fisher (1936)

- (1) CC : Comprimento do caule
LC : Largura do caule
CP : Comprimento da pétala
LP : Largura da pétala

Utilizando as medidas comprimento do caule e largura do caule, consideramos as oito plantas do tipo Iris Versicolor (população π_1) e as oito plantas do tipo Iris Virginica (população π_2).

Temos $n_1 = n_2 = 8$ e $p = 2$. Seja $\{\underline{x}_{11}, \dots, \underline{x}_{18}\}$ a amostra de π_1 e $\{\underline{x}_{21}, \dots, \underline{x}_{28}\}$ a amostra de π_2 , onde $\underline{x}' = (x_1, x_2)$ com x_1 denotando o comprimento do caule e x_2 , a largura do caule.

Seja E uma planta à qual associamos o vetor de medidas $\underline{x}'_0 = (5, 7, 2, 8)$.

Discriminamos E em π_1 ou π_2 , através dos métodos 1 e 2 de Anderson, do método de Gessaman e Gessaman e do de Patrick e Fischer, supondo $p_1 = p_2 = 0,5$.

APLICAÇÃO DO MÉTODO 1 DE ANDERSON

Para utilizarmos o método de subpartição de regiões de tolerância fixamos o número aproximado de observações pertencentes a cada região de tolerância construída em $m=1$, e o número de etapas de subpartição em $q=2$. Através da expressão (2.3.3) apresentada na seção 2.3.1 do capítulo 2, a sequência de inteiros $\{k_1, \dots, k_{16}\}$ foi definida por $\{4, 8, 12, 1, 2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16\}$ e as funções de corte por $\phi_4(\underline{x}) =$

$\phi_8(\underline{x}) = \phi_{12}(\underline{x}) = x_1$ e $\phi_i(\underline{x}) = x_2$ para $i \neq 4, 8, 12$.

Após 15 passos dividimos S_p em 16 conjuntos mutuamente exclusivos:

$$B_{(1)} = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 \leq 2,3 \}$$

$$B_{(2)} = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } 2,3 < x_2 \leq 2,4 \}$$

$$B_{(3)} = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } 2,4 < x_2 \leq 2,5 \}$$

$$B_{(4)} = \{ \underline{x} \in S_p \mid x_1 \leq 5,7 \text{ e } x_2 > 2,5 \}$$

$$B_{(5)} = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,3 \text{ e } x_2 \leq 2,7 \}$$

$$B_{(6)} = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,3 \text{ e } 2,7 < x_2 \leq 2,9 \}$$

$$B_{(7)} = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,3 \text{ e } 2,9 < x_2 \leq 3,3 \}$$

$$B_{(8)} = \{ \underline{x} \in S_p \mid 5,7 < x_1 \leq 6,3 \text{ e } x_2 > 3,3 \}$$

$$B_{(9)} = \{ \underline{x} \in S_p \mid 6,3 < x_1 \leq 6,9 \text{ e } x_2 \leq 2,8 \}$$

$$B_{(10)} = \{ \underline{x} \in S_p \mid 6,3 < x_1 \leq 6,9 \text{ e } 2,8 < x_2 \leq 3,0 \}$$

$$B_{(11)} = \{ \underline{x} \in S_p \mid 6,3 < x_1 \leq 6,9 \text{ e } 3,0 < x_2 \leq 3,2 \}$$

$$B_{(12)} = \{ \underline{x} \in S_p \mid 6,3 < x_1 \leq 6,9 \text{ e } x_2 > 3,2 \}$$

$$B(13) = \{ \underline{x} \in S_p \mid x_1 > 6,9 \text{ e } x_2 \leq 2,9 \}$$

$$B(14) = \{ \underline{x} \in S_p \mid x_1 > 6,9 \text{ e } 2,9 < x_2 \leq 3,0 \}$$

$$B(15) = \{ \underline{x} \in S_p \mid x_1 > 6,9 \text{ e } 3,0 < x_2 \leq 3,2 \}$$

$$B(16) = \{ \underline{x} \in S_p \mid x_1 > 6,9 \text{ e } x_2 > 3,2 \}$$

O vetor $\underline{x}_0$ pertence a $B(4)$. Logo, $S_{\underline{x}} = B(4)$. A observação amostral (única!) pertencente a $S_{\underline{x}}$ é $\underline{x}_{16}$. Neste caso $N_1=1$ e $N_2=0$.

Pela regra de discriminação dada em (3.1.2), que é a que deve ser aplicada neste método, temos que E foi discriminada em π_1 , isto é, como do tipo Iris Versicolor.

APLICAÇÃO DO MÉTODO 2 DE ANDERSON

Segundo o método de subpartição de regiões de tolerância, uma vez fixado o número de etapas de subpartição em $q=2$ e escolhido, arbitrariamente, $m=3$, onde m é o número aproximado de observações amostrais pertencentes a cada uma das $L^q=L^2$ regiões de tolerância construídas, temos

$$L = \left\lceil \left(\frac{n+1}{m} \right)^{1/q} \right\rceil = \lceil 2,38 \rceil = 2,$$

onde $n = n_1 + n_2$.

A sequência de inteiros $\{k_1, k_2, k_3\}$ necessária para a cons-

trução das quatro regiões de tolerância é dada por $\{8,4,12\}$ e as funções de corte por $\phi_8(\underline{x}) = x_1$ e $\phi_4(\underline{x}) = \phi_8(\underline{x}) = x_2$ (vide seção 2.3.1 do capítulo 2).

Dividimos S_p em quatro conjuntos mutuamente exclusivos:

$$B(1,2,3,4) = \{\underline{x} \in S_p | x_1 \leq 6,3 \text{ e } x_2 \leq 2,7\}$$

$$B(5,6,7,8) = \{\underline{x} \in S_p | x_1 \leq 6,3 \text{ e } x_2 > 2,7\}$$

$$B(9,10,11,12) = \{\underline{x} \in S_p | x_1 > 6,3 \text{ e } x_2 \leq 3,0\}$$

$$B(13,16,15,16) = \{\underline{x} \in S_p | x_1 > 6,3 \text{ e } x_2 > 3,0\}$$

O vetor $\underline{x}_0$ pertence a $B(5,6,7,8)$, logo $S_{\underline{x}} = B(5,6,7,8)$. Os vetores amostrais pertencentes a $S_{\underline{x}}$ são $\underline{x}_{16}$, $\underline{x}_{17}$ e $\underline{x}_{21}$. Neste caso, $N_1 = 2$ e $N_2 = 1$. Pela regra de discriminação dada em (3.1.2), que é a que deve ser aplicada neste método, temos que E foi discriminada em π_1 , isto é, como do tipo Iris Versicolor.

APLICAÇÃO DO MÉTODO DE GESSAMAN E GESSAMAN

Através da amostra de π_1 e fixando $q = 2$, onde q é o número de etapas de subpartição, particionamos S_p , segundo o método de subpartição de regiões de tolerância, em $L_1^q = L_1^2$ regiões de tolerância, onde $L_1 = \lfloor n_1^{1/(q+1)} \rfloor = \lfloor 2 \rfloor = 2$, cada uma contendo aproximadamente $m_1 = \left\lceil \frac{n_1}{\lfloor n_1^{1/(q+1)} \rfloor^q} \right\rceil = \lceil 2 \rceil = 2$ observações amostrais de π_1 .

A sequência de inteiros $\{k_1, k_2, k_3\}$ necessária para a cons-

trução das quatro regiões de tolerância é definida por $\{4,2,6\}$ e as funções de corte por $\phi_4(\underline{x}) = x_1$ e $\phi_2(\underline{x}) = \phi_6(\underline{x}) = x_2$ (vide seção 2.3.1 do capítulo 2).

Dividimos S_p nos seguintes conjuntos mutuamente exclusivos:

$$B(1,2) = \{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 \leq 2,4 \}$$

$$B(3,4) = \{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 > 2,4 \}$$

$$B(5,6) = \{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 \leq 3,1 \}$$

$$B(7,8) = \{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 > 3,1 \}$$

Temos $S_{\underline{x}} = B(3,4)$. As observações amostrais pertencentes a $S_{\underline{x}}$ são $\underline{x}_{16}$, $\underline{x}_{17}$, $\underline{x}_{21}$, $\underline{x}_{22}$, $\underline{x}_{24}$ e $\underline{x}_{27}$; neste caso, $N_1 = 2$ e $N_2 = 4$. Pela regra de discriminação dada em (3.1.2), que é a que deve ser aplicada neste método, temos que E foi discriminada em π_2 , isto é, como do tipo Iris Virginica.

Fazendo uso da amostra de π_2 e utilizando a mesma sequência de inteiros e funções de corte obtidas para a amostra de π_1 , pois $n_1 = n_2$, concluímos que E deve ser discriminada aleatoriamente em π_1 ou π_2 .

APLICAÇÃO DO MÉTODO DE PATRICK E FISCHER

A regra de discriminação do método de Patrick e Fischer envolve o cálculo dos volumes das regiões de tolerância que contém $\underline{x}_0$. Ao particionarmos S_p segundo o método de subpartição de regiões de tolerância, algumas das regiões de tolerân-

cia construídas terão volumes infinitos. Para contornar este problema, iniciamos a partição de S_p limitando-o através de quatro retas pois $p = 2$; aplicamos, em seguida, o método de subpartição de regiões de tolerância na região limitada.

Para limitarmos S_p , através da amostra de π_1 , aplicamos o método de construção de regiões de tolerância de Fraser (vide seção 2.2 do capítulo 2), utilizando a sequência de números inteiros dada por $\{1, 2, 3, 4\}$ e as funções de corte definidas, arbitrariamente, por $\phi_1(\underline{x}) = x_1$, $\phi_2(\underline{x}) = x_2$, $\phi_3(\underline{x}) = -x_1$ e $\phi_4(\underline{x}) = -x_2$.

Obtivemos:

$$B(1) = \{\underline{x} \in S_p | x_1 \leq 4,9\}$$

$$B(2) = \{\underline{x} \in S_p | x_1 > 4,9 \text{ e } x_2 \leq 2,3\}$$

$$B(3) = \{\underline{x} \in S_p | x_1 \geq 7,0 \text{ e } x_2 > 2,3\}$$

$$B(4) = \{\underline{x} \in S_p | 4,9 < x_1 < 7,0 \text{ e } x_2 \geq 3,3\}$$

$$B(5,6,7,8,9) = \{\underline{x} \in S_p | 4,9 < x_1 < 7,0 \text{ e } 2,3 < x_2 < 3,3\}$$

As observações amostrais $\underline{x}_{12}$, $\underline{x}_{13}$, $\underline{x}_{15}$ e $\underline{x}_{16}$ pertencem a $B(5,6,7,8,9)$ (região limitada). Aplicando o método de subpartição de regiões de tolerância neste conjunto obtivemos $L_1 = [(n_1 - 4)^{1/(q+1)}] = 1$ e $m_1 = \left[\frac{(n_1 - 4)}{[(n_1 - 4)^{1/(q+1)}]q} \right] = [4] = 4$ e, neste caso, não particionamos $B(5,6,7,8,9)$. Logo, $S_{1\underline{x}} = B(5,6,7,8,9)$, $N_1 = 4$ e $V_{S_{1\underline{x}}} = 2,1$.

Através da amostra de π_2 , particionamos S_p segundo o mes-

mo esquema utilizado para a amostra de π_1 .

Obtivemos:

$$B(1) = \{\underline{x} \in S_p | x_1 \leq 4,9\}$$

$$B(2) = \{\underline{x} \in S_p | x_1 > 4,9 \text{ e } x_2 \leq 2,7\}$$

$$B(3) = \{\underline{x} \in S_p | x_1 \geq 7,6 \text{ e } x_2 > 2,7\}$$

$$B(4) = \{\underline{x} \in S_p | 4,9 < x_1 < 7,6 \text{ e } x_2 \geq 3,3\}$$

$$B(5,6,7,8,9) = \{\underline{x} \in S_p | 4,9 < x_1 < 7,6 \text{ e } 2,7 < x_2 < 3,3\}$$

Nesta situação temos $S_{2\underline{x}} = B(5,6,7,8,9)$, $N_2 = 4$ e $V_{S_{2\underline{x}}} = 1,62$.

Pela regra de discriminação dada em (3.1.4), que é a que deve ser aplicada neste método, temos que E foi discriminada em π_1 , isto é, como do tipo Iris Versicolor.

3.1.5. CONSISTÊNCIA DOS MÉTODOS APRESENTADOS

De acordo com os teoremas que citaremos a seguir mostraremos que o método de Patrick e Fischer é consistente com o método ótimo de discriminação paramétrico da razão de verossimilhança. Apesar de extensiva busca na literatura, nada encontramos sobre a consistência dos demais métodos não paramétricos de discriminação forçada baseados em regiões de tolerância, apresentados nas seções 3.1.1 e 3.1.2 deste capítulo.

Consideremos o teorema que segue, cuja apresentação e demonstração (vide apêndice B) baseamos no teorema encontrado em

Gessaman (1970).

TEOREMA 1. Seja $\{\underline{x}_1, \dots, \underline{x}_n\}$ uma amostra da variável aleatória p -dimensional $\underline{X}$, cuja função de distribuição denotaremos por F . Suponhamos que F seja contínua e que a correspondente função densidade de probabilidade seja denotada por f . Seja $\underline{x}$, um ponto pertencente a S_p onde $f(\underline{x})$ é positiva.

Através da amostra de $\underline{X}$, construíamos regiões de tolerância segundo o método de subpartição de regiões de tolerância, de modo que cada uma contenha, aproximadamente, m observações amostrais, sendo m uma função de n que satisfaz a seguinte condição:

m cresce sempre que n cresce, mas de modo mais lento que n .

Denotemos por $S_{\underline{x}}$, a região de tolerância que contém $\underline{x}$ e por $V_{S_{\underline{x}}}$, o volume de $S_{\underline{x}}$.

Contemos o número de observações amostrais pertencentes a $S_{\underline{x}}$ e denotemos este número por N .

Um estimador consistente de $f(\underline{x})$ é dado por

$$\hat{f}(\underline{x}) = \frac{N}{n} \cdot \frac{1}{V_{S_{\underline{x}}}} \quad \square$$

O teorema que citamos a seguir foi apresentado e demonstrado (vide apêndice A) por Fix e Hodges (1951).

TEOREMA 2. Seja $f_i(\underline{x})$ um estimador de $f_i(\underline{x})$ que depende do tamanho n_i da amostra retirada da população π_i , $i = 1, 2$.

Se $f_i(\underline{x})$ for consistente com $f_i(\underline{x})$ para todo $\underline{x}$, então o

método de discriminação baseado em $\hat{f}_i(\underline{x})$, $i = 1, 2$, é consistente com o método ótimo de discriminação paramétrico da razão de verossimilhança. $\square$

Pelo teorema 1, verificamos que o estimador $\hat{f}_i(\underline{x})$ dado em (3.1.3) é consistente com $f_i(\underline{x})$, $i = 1, 2$. A partir disto, concluimos, pelo teorema 2, que o método não paramétrico de discriminação forçada de Patrick e Fischer é consistente com o método ótimo de discriminação paramétrico da razão de verossimilhança.

Nossa tentativa em demonstrar a consistência dos métodos 1 e 2 de Anderson e de Gessaman e Gessaman, seguindo o raciocínio adotado na demonstração de consistência do método de Patrick e Fischer, foi posta de lado porque nestes três métodos a estimação de $f_i(\underline{x})$, $i = 1, 2$, baseia-se numa *única* região de tolerância. Ao adotarmos outras linhas de raciocínio na demonstração de consistência desses três métodos de discriminação, notamos que o tratamento do problema tornou-se bastante complicado justificando, portanto, o fato de ter sido abandonado. No entanto, observando a semelhança na construção de $\hat{f}_i(\underline{x})$, $i = 1, 2$, nos métodos 1 e 2 de Anderson e na do método de Patrick e Fischer cremos, intuitivamente, que os métodos de Anderson são consistentes com o método ótimo de discriminação paramétrico da razão de verossimilhança. Quanto ao método de Gessaman e Gessaman, observamos que ao particionarmos S_p somente através da amostra de π_1 , por exemplo, pudemos demonstrar que $\hat{f}_1(\underline{x})$ é consistente com $f_1(\underline{x})$ mas nada conseguimos concluir sobre a consistência de $\hat{f}_2(\underline{x})$ e, consequentemente,

deste método de discriminação.

3.1.6. PROBABILIDADE DE ERRO DE CLASSIFICAÇÃO

Na literatura pesquisada encontramos um (único!) resultado sobre a probabilidade exata de erro de classificação dos métodos não paramétricos de discriminação forçada baseados em regiões de tolerância.

Admitindo conhecida a função densidade de probabilidade $f_i(\underline{x})$, $i = 1, 2$, Patrick e Fischer (1970) deduziram uma expressão para a probabilidade exata de erro de classificação de seu método discriminante (vide apêndice D).

Nos métodos de Anderson e de Gessaman e Gessaman, nossa tentativa em deduzir a probabilidade exata de erro de classificação, utilizando o raciocínio adotado por Patrick e Fischer, foi posta de lado, pois, para estes métodos, a estimação de $f_i(\underline{x})$, $i = 1, 2$, é realizada numa única região de tolerância.

Nestes casos, uma solução consiste em estimar a probabilidade exata de erro de classificação. Convém observar que esta solução deve ser adotada quando $f_i(\underline{x})$, $i = 1, 2$, for desconhecida ou quando for conhecida, mas o cálculo da probabilidade exata de erro de classificação torna-se muito trabalhoso.

Na literatura pesquisada, apesar de exaustiva busca, nada encontramos sobre comparações entre as probabilidades de erro de classificação dos métodos de discriminação forçada basea-

dos em regiões de tolerância. Todavia, Gessaman e Gessaman (1972) compararam a probabilidade de erro de classificação de seu método discriminante com a do método de discriminação linear amostral (Fischer), através do método de estimação por partição. Para tanto, trabalharam, na construção dos métodos de discriminação, com três pares de diferentes distribuições normais sendo que, para cada par, geraram amostras de tamanhos diferentes. Os resultados desta comparação, apresentados na seção 5.2.2 do capítulo 5, mostram que o método de Gessaman e Gessaman tem um bom desempenho, isto é, tem menor probabilidade estimada de erro de classificação que o método discriminante linear amostral nos casos em que este último método não deve ser aplicado (distribuições normais com matrizes de covariância diferentes).

3.2. MÉTODOS DE DISCRIMINAÇÃO PARCIAL

Na literatura pesquisada, verificamos que:

- os autores não se preocupam em demonstrar se seus métodos não paramétricos de discriminação parcial são ou não consistentes com o método ótimo de discriminação parcial, definido na seção 5 do capítulo 1 (caso paramétrico). Em Quesenberry e Gessaman (1968) encontramos algumas justificativas para tal fato, sendo que não as citamos neste trabalho por considerá-las pouco convincentes.

- os autores não deduzem expressões para a probabilidade exata de erro de classificação e de não classificação; em alguns

métodos, os autores fixam limites superiores para a probabilidade de erro de classificação, mas não indicam como calcular a probabilidade de não classificação. Desse modo, para termos idéia do comportamento de tais probabilidades uma solução é estimá-las através de algum dos métodos apresentados na seção 6 do capítulo 1.

3.2.1. MODELO GERAL

Todos os métodos de discriminação parcial baseados em regiões de tolerância podem ser construídos da seguinte maneira:

Definamos, em S_p , duas regiões de tolerância, A_1 e A_2 , tais que A_1 contenha, se possível, mais observações de π_1 que de π_2 e A_2 contenha, se possível, mais observações amostrais de π_2 que de π_1 .

Seja E um elemento a ser discriminado em π_1 ou π_2 , ao qual associamos o vetor de medidas $\underline{x}$.

Definimos a seguinte regra discriminatória:

Se $\underline{x}$ pertencer a $A_1 \cap \bar{A}_2$, o elemento E será classificado em π_1 .

Se $\underline{x}$ pertencer a $\bar{A}_1 \cap A_2$, o elemento E será classificado em π_2 . (3.2.1)

Se $\underline{x}$ pertencer a $A_3 = (A_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2)$, o elemento E não será classificado em π_1 ou π_2 .

A região de tolerância A_3 será denominada *região de dúvida* (ou região cinza).

Observemos que os métodos de discriminação parcial basea-

dos em regiões de tolerância, definidos pela regra discriminatória dada em (3.2.1) não levam em consideração o tamanho n_i da amostra de π_i nem a probabilidade a priori, p_i , do elemento E pertencer a π_i , $i = 1, 2$. Tal característica pode acarretar uma diminuição do desempenho destas regras discriminatórias (desempenho medido em termos da probabilidade de erro de classificação e de não classificação) quando n_1 é muito diferente de n_2 e/ou p_1 é muito diferente de p_2 .

3.2.2. APRESENTAÇÃO DOS MÉTODOS

Os métodos não paramétricos de discriminação parcial baseados em regiões de tolerância encontrados na literatura dividem-se em dois grupos:

- *métodos sem controle da probabilidade de erro de classificação*: método 1 de Gessaman e Gessaman (1972).
- *métodos com controle da probabilidade de erro de classificação através de limites superiores pré-fixados*: método de Quesenberry e Gessaman (1968) e o método 2 de Gessaman e Gessaman (1972).

Apresentaremos, a seguir, a maneira pela qual cada um destes métodos constrói as regiões de tolerância A_1 e A_2 .

MÉTODO 1 DE GESSAMAN E GESSAMAN (1972)

Através da amostra de π_1 , particionamos S_p , segundo o método de subpartição de regiões de tolerância, em L_1^q regiões de

tolerância, onde

$$L_1 = [n_1^{1/(q+1)}], \quad (3.2.2)$$

tais que cada uma contenha, aproximadamente,

$$m_1 = \left[\frac{n_1}{L_1^q} \right] = \left[\frac{n_1}{[n_1^{1/(q+1)}]^q} \right] \quad (3.2.3)$$

observações amostrais de π_1 (vide seção 2.3.1 do capítulo 2 e seção 3.1.2 deste capítulo).

Fixemos um número inteiro t , tal que $1 \leq t \leq m_1$.

A_1 será formada pela união das regiões de tolerância que tiverem pelo menos t observações amostrais de π_1 a mais que de π_2 ;

A_2 será formada pela união das regiões de tolerância que tiverem pelo menos t observações amostrais de π_2 a mais que de π_1 e

A_3 será formada pela união das demais regiões de tolerância.

Uma vez construídas as regiões de tolerância A_1, A_2 e A_3 , discriminamos o elemento E segundo a regra discriminatória dada em (3.2.1).

Fazendo uso da amostra de π_2 podemos, de maneira análoga, determinar outras regiões A_1, A_2 e A_3 que, de modo geral, são diferentes daquelas obtidas através da amostra de π_1 . Nesta situação, o resultado obtido quando aplicamos a regra discriminatória construída através da amostra de π_1 pode diferir do

obtido pela construída através da amostra de π_2 .

Para grandes valores de t , as regiões A_1 e A_2 ficam, nitidamente, demarcadas como regiões contendo um grande número de observações amostrais de π_1 e um grande número de observações amostrais de π_2 , respectivamente. Porém, a região A_3 aumenta, pois será a união das regiões de tolerância que tiverem i , $i = 0, \dots, t-1$, observações a mais de π_1 que de π_2 ou de π_2 a mais que de π_1 . Podemos concluir, portanto, que a probabilidade de erro de classificação será pequena e a probabilidade de não classificação será grande.

Para valores pequenos de t , o processo se inverte, isto é, concluímos que a probabilidade de erro de classificação será grande e que a probabilidade de não classificação será pequena. Uma saída para a situação é escolher, se possível, um valor de t onde as probabilidades se equilibrem e não sejam muito grandes. Nossa sugestão é calcular a probabilidade estimada de erro de classificação e de não classificação, deste método discriminante, para cada conjunto de diferentes distribuições de $f_i(\underline{x})$, $i = 1, 2$, diferentes tamanhos de amostras e diferentes valores de t , verificando se é possível encontrar, para cada situação analisada, o valor de t desejado.

MÉTODO DE QUESENBERRY E GESSAMAN (1968)

Na seção 1 do capítulo 1 definimos que $P(\pi_j | \pi_i)$, $i, j = 1, 2$; $i \neq j$, é a probabilidade de classificar erradamente um elemento da população π_i na população π_j .

Fixemos um número real $\alpha_i, 0 < \alpha_i < 1$, de modo que $P(\pi_j | \pi_i) \leq \alpha_i$, $i, j = 1, 2; i \neq j$.

Utilizando o método de Tukey, definido na seção 2.3.2 do capítulo 2, construíamos a região de tolerância A_1 , através da amostra de π_1 , de modo que contenha

$$a_1 = [(1-\alpha_1)(n_1+1)] \quad (3.2.4)$$

observações amostrais de π_1 .

Através da amostra de π_2 , construíamos, de maneira análoga, a região de tolerância A_2 de modo que contenha

$$[(1-\alpha_2)(n_2+1)] \quad (3.2.5)$$

observações amostrais de π_2 .

Construídas as regiões A_1 e A_2 , procedamos à classificação do elemento E através da regra discriminante dada em (3.2.1).

Observemos que:

(a) através dos resultados apresentados na seção 2.4 do capítulo 2, verificamos que A_i é uma região de tolerância não paramétrica e que P_{A_i} (probabilidade de que uma nova observação amostral pertença a A_i) é uma variável aleatória com distribuição beta de parâmetros a_i e $(n_i - a_i + 1)$, cujo valor esperado é dado por

$$E(P_{A_i}) = \frac{a_i}{n_i + 1} = (1 - \alpha_i) + o(1/n_i), o(1/n_i) \geq 0, i=1, 2$$

(vide Quesenberry e Gessaman, 1968).

(b) $(1-\alpha_i) \leq 1-P(\pi_j|\pi_i) = P(\pi_i|\pi_i) = P(A_i \cap \bar{A}_j|\pi_i) \leq P(A_i|\pi_i) = P_{A_i}$,
 $i, j=1, 2; i \neq j$, onde $P(\pi_i|\pi_i)$ é a probabilidade de classificarmos corretamente um elemento pertencente a π_i .

Quando n_i cresce, podemos interpretar a_i , através das considerações (a) e (b), como o número mínimo esperado de observações amostrais de π_i classificadas corretamente pelo método de Quesenberry e Gessaman.

MÉTODO 2 DE GESSAMAN E GESSAMAN (1972)

Fixemos um limite superior, α_i , para

$$P(\pi_j|\pi_i), i, j=1, 2; i \neq j.$$

Através da amostra de π_1 , particionemos S_p , segundo o método de subpartição de regiões de tolerância, em L_1^q regiões de tolerância, onde

$$L_1 = [n_1^{1/(q+1)}], \quad (3.2.6)$$

tais que cada uma contenha, aproximadamente,

$$m_1 = \left[\frac{n_1}{L_1^q} \right] = \left[\frac{n_1}{[n_1^{1/(q+1)}]^q} \right] \quad (3.2.7)$$

observações amostrais de π_1 (vide seção 2.3.1 do capítulo 2 e seção 3.1.2 deste capítulo).

A_1 será formada pela união de

$$a_1 = [(1-\alpha_1)L_1^q] \quad (3.2.8)$$

regiões de tolerância escolhidas entre aquelas que contenham

mais observações amostrais de π_1 que de π_2 .

Através da amostra de π_2 , construímos, de maneira análoga, a região de tolerância A_2 , que será formada pela união de

$$a_2 = [(1-\alpha_2)L_2^q] \quad (3.2.9)$$

regiões de tolerância escolhidas entre aquelas que contenham mais observações amostrais de π_2 que de π_1 .

Construídas as regiões A_1 e A_2 , classificamos o elemento E através da regra de discriminação dada em (3.2.1).

Observemos que o método 2 de Gessaman e Gessaman pode ser considerado uma extensão do método de Quesenberry e Gessaman, com a diferença de que as observações amostrais que vão para a região de tolerância A_i são, agora, escolhidas indiretamente através das n_i observações amostrais de $\pi_i, i=1,2$, isto é, a partir de determinadas regiões de tolerância construídas.

3.2.3. UM EXEMPLO NUMÉRICO

Consideremos a situação apresentada na seção 3.1.4.

Discriminamos o elemento E ao qual associamos o vetor de medidas $\underline{x}_0^1 = (5,7, 2,8)$, através dos três métodos de discriminação parcial apresentados na seção 3.2.2.

APLICAÇÃO DO MÉTODO 1 DE GESSAMAN E GESSAMAN

Através da amostra de π_1 , particionamos S_p nos seguintes

conjuntos mutuamente exclusivos:

$$B(1,2) = \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 \leq 2,4 \right\}$$

$$B(3,4) = \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 > 2,4 \right\}$$

$$B(5,6) = \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 \leq 3,1 \right\}$$

$$B(7,8) = \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 > 3,1 \right\}$$

Observemos que estes conjuntos são os construídos pelo método de discriminação forçada de Gessaman e Gessaman, na seção 3.1.4.

As observações amostrais $\underline{x}_{14}$ e $\underline{x}_{18}$ pertencem a $B(1,2)$; $\underline{x}_{16}$, $\underline{x}_{17}$, $\underline{x}_{21}$, $\underline{x}_{22}$, $\underline{x}_{24}$ e $\underline{x}_{27}$ pertencem a $B(3,4)$; $\underline{x}_{13}$, $\underline{x}_{15}$, $\underline{x}_{23}$, $\underline{x}_{25}$, $\underline{x}_{26}$ e $\underline{x}_{28}$ pertencem a $B(5,6)$ e $\underline{x}_{11}$ e $\underline{x}_{12}$ pertencem a $B(7,8)$.

Fixemos $t = 1$.

A região de tolerância A_1 é formada pela união de $B(1,2)$ e $B(7,8)$. A região A_2 é formada pela união de $B(3,4)$ e $B(5,6)$. Observemos que A_3 é vazia.

Segundo a regra discriminatória dada em (3.2.1), classificamos E em π_2 , isto é, como do tipo Iris Virginica, pois $\underline{x}_0$ pertence a A_2 .

Se particionarmos S_p através da amostra de π_2 , verificaremos que as observações amostrais $\underline{x}_{14}$, $\underline{x}_{18}$, $\underline{x}_{22}$ e $\underline{x}_{27}$ pertencem a $B(1,2) = \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 \leq 2,7 \right\}$; $\underline{x}_{16}$, $\underline{x}_{17}$, $\underline{x}_{21}$ e $\underline{x}_{24}$ pertencem a $B(3,4) = \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 > 2,7 \right\}$; $\underline{x}_{15}$, $\underline{x}_{23}$, $\underline{x}_{25}$,

$\underline{x}_{26}$ e $\underline{x}_{28}$ pertencem a $B(5,6) = \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 \leq 3,0 \right\}$ e $\underline{x}_{11}$,
 $\underline{x}_{12}$ e $\underline{x}_{13}$ pertencem a $B(7,8) = \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 > 3,0 \right\}$.

Fixando $t = 1$, temos

$$A_1 = B(7,8), A_2 = B(5,6) \text{ e } A_3 = B(1,2) \cup B(3,4).$$

Através da regra discriminatória dada em (3.2.1) não classificamos E em π_1 nem em π_2 , pois $\underline{x}_0$ pertence a A_3 .

Comparando este último resultado sobre a classificação do elemento E , com o resultado obtido através do método de discriminação forçada de Gessaman e Gessaman, construído através da amostra de π_2 , na seção 3.1.4, verificamos que o elemento E discriminado aleatoriamente deveria, na realidade, não ser discriminado.

APLICAÇÃO DO MÉTODO DE QUESENBERY E GESSAMAN

Fixemos $\alpha_1 = \alpha_2 = 0,3$.

Teremos $a_1 = a_2 = [(1-0,3) \cdot (8+1)] = [6,3] = 6$.

Através da amostra de π_1 e utilizando o método de Tukey, particionamos S_p nas seguintes regiões de tolerância:

$$B(1) = \left\{ \underline{x} \in S_p \mid x_1 \leq 4,9 \right\}$$

$$B(2) = \left\{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 \leq 2,3 \right\}$$

$$B(3,4,5,6,7,8,9) = \left\{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 > 2,3 \right\}$$

$$\text{Logo, } A_1 = \left\{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 > 2,3 \right\}$$

Através da amostra de π_2 , particionamos S_p , de maneira análoga e encontramos

$$A_2 = \{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 > 2,7 \}$$

Pela regra discriminante dada em (3.2.1) temos:

$$A_1 \cap \bar{A}_2 = \{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } 2,3 < x_2 \leq 2,7 \} ;$$

$$\bar{A}_1 \cap A_2 = \phi \quad \text{e}$$

$$A_3 = (A_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2) = \{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 > 2,7 \} \cup \\ \{ \underline{x} \in S_p \mid x_1 \leq 4,9 \} \cup \{ \underline{x} \in S_p \mid x_1 > 4,9 \text{ e } x_2 \leq 2,3 \}$$

Observemos que por esta regra discriminante, nunca classificamos um elemento em π_2 . Como $\underline{x}_0$ pertence a A_3 , não classificamos o elemento E.

APLICAÇÃO DO MÉTODO 2 DE GESSAMAN E GESSAMAN

Fixemos $\alpha_1 = \alpha_2 = 0,3$

Através da amostra de π_1 , particionamos S_p nas quatro regiões de tolerância apresentadas na página 67.

Para $a_1 = [(1-0,3).4] = [2,8] = 2$, temos que A_1 é formada pela união de duas regiões de tolerância escolhidas entre aquelas que contêm mais observações amostrais de π_1 que de π_2 . Logo, $A_1 = B(1,2) \cup B(7,8)$.

Através da amostra de π_2 , particionamos S_p nas quatro regiões de tolerância apresentadas nas páginas 67 e 68.

Para $a_2 = a_1 = 2$, temos que A_2 é formada pela união de duas regiões de tolerância escolhidas entre aquelas que contêm mais observações amostrais de π_2 que de π_1 . Observando que, somente $B(5,6)$ tem maior número de observações amostrais de π_2 que de π_1 fizemos $A_2 = B(5,6)$.

Segundo a regra discriminante dada em (3.1.2) verificamos que:

$$\begin{aligned} A_1 \cap \bar{A}_2 = A_1 &= \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 \leq 2,4 \right\} \cup \\ &\cup \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 > 3,1 \right\}; \\ \bar{A}_1 \cap A_2 = A_2 &= \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } x_2 \leq 3,0 \right\} \text{ e} \\ A_3 &= (A_1 \cap A_2) \cup (\bar{A}_1 \cap \bar{A}_2) = \left\{ \underline{x} \in S_p \mid x_1 \leq 6,3 \text{ e } x_2 > 2,4 \right\} \cup \\ &\cup \left\{ \underline{x} \in S_p \mid x_1 > 6,3 \text{ e } 3,0 < x_2 \leq 3,1 \right\}. \end{aligned}$$

Como $\underline{x}_0$ pertence a A_3 , não classificamos o elemento E em π_1 nem em π_2 .

3.2.4. PROBABILIDADES DE ERRO DE CLASSIFICAÇÃO E DE NÃO CLASSIFICAÇÃO

Consideremos o método de discriminação de Quesenberry e Gessaman. Temos que $P(\pi_2 \mid \pi_1) = P(\bar{A}_1 \cap A_2 \mid \pi_1) = P_{\bar{A}_1 \cap A_2} \leq P_{\bar{A}_1}$.

Sabendo que $P_{\bar{A}_1}$ tem distribuição beta de parâmetros $(n_1 - a_1 + 1)$ e a_1 , Quesenberry e Gessaman (1968) demonstraram que, quando n_1 cresce, $P_{\bar{A}_1}$ converge, em probabilidade, para α_1 . Porém, para casos práticos onde n_1 é pequeno, temos que um limite superior para $P(\pi_2 \mid \pi_1)$ é $P_{\bar{A}_1}$.

Através dos gráficos de Murphy (1948), fixados n_1, a_1 e γ , podemos calcular o valor de β tal que $P(P_{\bar{A}_1} \leq \beta) \leq \gamma$, isto é, podemos obter um valor aproximado para $P_{\bar{A}_1}$ com $\gamma\%$ de confiança. Resultados análogos podem ser deduzidos para $P(\pi_1 | \pi_2)$. Além disso, podemos estender estes resultados para o método 2 de discriminação de Gessaman e Gessaman observando que, agora, $P_{\bar{A}_i}$ tem distribuição beta de parâmetros $(n_i - a_i m_i + 1)$ e $a_i m_i$, onde a_i e m_i são dados, respectivamente, por:

$$a_i = [(1 - \alpha_i) [n_i^{1/(q+1)}]^q] \text{ e } m_i = \left[\frac{n_i}{[n_i^{1/(q+1)}]^q} \right], i=1,2 \quad (3.2.10)$$

(vide teorema 2 da seção 2.4 do capítulo 2 e seção 3.2.2 deste capítulo).

Outra maneira de analisarmos as probabilidades de erro de classificação e de não classificação dos métodos não paramétricos de discriminação parcial, apresentados neste capítulo, é estimá-las através de algum método apresentado na seção 6 do capítulo 1.

Gessaman e Gessaman (1972) compararam as probabilidades de erro de classificação e de não classificação de seu método 2 de discriminação parcial com as do método de Quesenberry e Gessaman, através do método de estimação por partição. Para isso, supuseram que $f_i(\underline{x})$, $i = 1, 2$, tivesse distribuição normal e utilizaram diferentes tamanhos de amostras na construção dos métodos de discriminação. Os resultados desta comparação, apresentados na seção 5.2.3 do capítulo 5, mostram que o método 2 de Gessaman e Gessaman tem maior probabilidade es-

timada de classificação correta e menor probabilidade estimada de não classificação que o método de Quesenberry e Gessaman. Estes resultados podem ser justificados se observarmos que Quesenberry e Gessaman não utilizam o princípio do "vizinho mais próximo" na construção de A_1 e A_2 , isto é, escolhem A_1 e A_2 sem levar em consideração se elas contêm ou não um maior número de observações amostrais de π_1 e π_2 , respectivamente. Desse modo, pode acontecer, por exemplo, de $\bar{A}_1$ e $\bar{A}_2$ conterem muitas observações de π_1 , o que implica na região de dúvida, A_3 , conter muitas observações de π_1 e, portanto, numa grande probabilidade de não classificação.

Concluindo, temos que o método 2 de Gessaman e Gessaman é preferível ao de Quesenberry e Gessaman quando a probabilidade de erro de classificação é controlada.

Não encontramos, na literatura pesquisada, nem conseguimos deduzir uma expressão para as probabilidades exatas de erro de classificação e de não classificação do método 1 de Gessaman e Gessaman. Nesta situação, uma solução consiste, como dissemos anteriormente, em estimar tais probabilidades através de algum método apresentado na seção 6 do capítulo 1.

CAPÍTULO 4

MÉTODOS NÃO PARAMÉTRICOS DE DISCRIMINAÇÃO

BASEADOS NOS v VIZINHOS MAIS PRÓXIMOS

4.1. INTRODUÇÃO

Seja E um elemento ao qual associamos o vetor de medidas $\underline{x}$. Seja $\{\underline{x}_{11}, \dots, \underline{x}_{1n_1}, \underline{x}_{12}, \dots, \underline{x}_{2n_2}\}$ a amostra combinada de π_1 e π_2 . Seja v um número inteiro tal que $1 \leq v \leq n_1 + n_2$.

Definamos, em S_p , uma função distância real, que denotaremos por d . Particionemos S_p , através da função distância d e da amostra combinada de π_1 e π_2 , em dois conjuntos mutuamente exclusivos tais que um deles, denotado por $S_{\underline{x}}$, contenha as v observações amostrais mais próximas de $\underline{x}$ e o outro contenha as $(n_1 + n_2 - v)$ observações amostrais restantes.

Para discriminarmos E em π_1 ou π_2 , utilizaremos as v observações amostrais mais próximas de $\underline{x}$, pertencentes a $S_{\underline{x}}$.

Observando a maneira pela qual S_p foi particionado, verificamos que $S_{\underline{x}}$ é uma região de tolerância. Portanto, os métodos não paramétricos de discriminação baseados nas v observações amostrais mais próximas de $\underline{x}$, denominados métodos de discriminação baseados nos v vizinhos mais próximos ou, sim-

plesmente, métodos "v-VMP" (v-NN em inglês), são métodos de discriminação baseados em regiões de tolerância.

Os motivos que nos levaram a dedicar um capítulo a parte aos métodos v-VMP foram os seguintes:

i) o valor histórico destes métodos de discriminação como os primeiros métodos não paramétricos de discriminação baseados em regiões de tolerância a serem utilizados na Análise Discriminante.

ii) o fato de encontrarmos, no caso de discriminação forçada, um método baseado numa única região de tolerância que é consistente com o método (ótimo) da razão de verossimilhança. Notemos que este fato merece destaque pois, dos métodos de discriminação forçada apresentados no capítulo anterior, que se baseavam numa única região de tolerância, nada se sabe quanto à consistência com o método (ótimo) da razão de verossimilhança.

Neste capítulo, apresentaremos e justificaremos dois métodos de discriminação v-VMP, sendo um método de discriminação forçada e outro de discriminação parcial. Comentaremos as propriedades destes métodos e apresentaremos processos para o cálculo da probabilidade exata de erro de classificação do método de discriminação forçada, bem como estimadores dessa probabilidade, para ambos os métodos; apresentaremos, também, estimadores da probabilidade de não classificação do método de discriminação parcial.

4.2. MÉTODO DE DISCRIMINAÇÃO FORÇADA

O método não paramétrico de discriminação forçada v-VMP foi primeiramente estudado por Fix e Hodges (1951,1952). Considerando amostras separadas e estabelecendo condições sobre v , estes autores demonstraram a consistência de seu método discriminante com o método (ótimo) da razão de verossimilhança (Welch, 1939). Além disso, calcularam a probabilidade exata de erro de classificação para pequenas amostras e diversos valores de v , supondo que a distribuição de probabilidade em π_i , $i = 1,2$, fosse normal.

Cover e Hart (1967) trabalharam com amostras misturadas e estudaram um caso particular do método de Fix e Hodges. Além disso, obtiveram limites assintóticos para a probabilidade de erro de classificação de seu método v-VMP.

Wagner (1971, 1973), Fritz (1975), Rogers e Wagner (1978), Devroye e Wagner (1979a e 1979b) e Penrod e Wagner (1979) estudaram os métodos de estimação por exclusão e por resubstituição da probabilidade de erro de classificação do método v-VMP de Cover e Hart. Na seção 4.2.5 apresentamos alguns dos resultados obtidos por estes autores.

Fukunaga e Kessell (1973), Fukunaga e Hostetler (1975) e Garnett e Yau (1977) obtiveram estimadores da probabilidade de erro de classificação do método de Cover e Hart, através da estimação do risco de Bayes.

4.2.1. ESTIMAÇÃO DE $f_i(\underline{x})$, $i=1,2$, e REGRA DISCRIMINATÓRIA

Consideremos a construção da região de tolerância $S_{\underline{x}}$, apresentada na seção 4.1.

Seja $d(\underline{x}_{ij}, \underline{x})$ a distância do vetor amostral $\underline{x}_{ij}, i=1,2; j=1,\dots,n_i$, ao vetor $\underline{x}$ a ser classificado.

Seja v um número inteiro e positivo, função dos tamanhos n_1 e n_2 das amostras retiradas, respectivamente, de π_1 e π_2 , satisfazendo às seguintes condições:

a) $1 \leq v \leq n_1 + n_2$

b) quando $n_i, i = 1,2$, cresce, v cresce, e $\frac{v}{n_i}, i = 1,2$, tende a zero. (4.2.1)

As condições estabelecidas em (4.2.1) são necessárias para a demonstração de consistência do método v -VMP com o método (ótimo) da razão de verossimilhança (vide apêndice C).

Combinemos as amostras de π_1 e π_2 e coloquemos em ordem crescente as distâncias $d(\underline{x}_{ij}, \underline{x}), i=1,2; j=1,\dots,n_i$. Definamos

U : v -ésima menor distância até $\underline{x}$ obtida das amostras de π_1 e π_2 combinadas

$S_{\underline{x}}$: região de tolerância que contém as observações das amostras de π_1 e π_2 combinadas, cujas distâncias até $\underline{x}$ são menores ou iguais a U .

Contemos o número N_i de observações amostrais de $\pi_i, i = 1,2$, encontradas em $S_{\underline{x}}$.

O estimador de $f_i(\underline{x})$, $i = 1, 2$, é dado por:

$$\hat{f}_i(\underline{x}) = \frac{N_i}{n_i} \cdot \frac{1}{V_{S_{\underline{x}}}}, \quad (4.2.2)$$

onde $V_{S_{\underline{x}}}$ denota o "volume" de $S_{\underline{x}}$.

Uma vez obtido o estimador $\hat{f}_i(\underline{x})$, $i = 1, 2$, este é aplicado na regra discriminatória da razão de verossimilhança.

Obtemos a seguinte regra discriminatória para o método v-VMP:

Se $p_1 \frac{N_1}{n_1} \cdot \frac{1}{U} > p_2 \frac{N_2}{n_2} \cdot \frac{1}{U}$, classificamos o elemento E em π_1 .

Se $p_1 \frac{N_1}{n_1} \cdot \frac{1}{U} < p_2 \frac{N_2}{n_2} \cdot \frac{1}{U}$, classificamos o elemento E em π_2 .

Como já foi definido no capítulo 1, p_i é a probabilidade a priori do elemento E pertencer à população π_i , $i = 1, 2$.

Eliminando o valor de U em ambos os membros da regra apresentada obtemos:

Se $p_1 \frac{N_1}{n_1} > p_2 \frac{N_2}{n_2}$, classificamos o elemento E em π_1 .
(4.2.3)

Se $p_1 \frac{N_1}{n_1} < p_2 \frac{N_2}{n_2}$, classificamos o elemento E em π_2 .

Observemos que se p_i for desconhecida e não puder ser estimada (amostras separadas), será substituída por $p_i = 0,5$, $i = 1, 2$; se p_i for desconhecida e puder ser estimada (amos-

tras misturadas), será substituída por $p_i = \frac{n_i}{n}$, onde $n = \sum n_i$, $i=1,2$.

4.2.2. CONSIDERAÇÕES

a) Analisando a regra discriminatória dada em (4.2.3) notamos que, quando $\frac{p_1}{n_1} = \frac{p_2}{n_2}$, esta regra discriminatória baseia-se somente nos números N_1 e N_2 de observações amostrais de π_1 e π_2 , respectivamente, encontradas entre as v mais próximas de $\underline{x}$, isto é,

Se $N_1 > N_2$, classificamos o elemento E em π_1 .

(4.2.4)

Se $N_1 < N_2$, classificamos o elemento E em π_2 .

A regra discriminatória dada em (4.2.3) é conhecida, na literatura pesquisada, como regra v -VMP de Fix e Hodges (1951, 1952), pois foram estes autores os que primeiro as desenvolveram. Embora a regra dada em (4.2.4) seja um caso particular da regra dada em (4.2.3), ela é conhecida, na literatura pesquisada, como regra v -VMP de Cover e Hart (1967), pois estes autores a estudaram amplamente devido à sua simplicidade de construção.

b) Ao aplicarmos o método de discriminação v -VMP notamos que três tipos de empates podem ocorrer:

O primeiro tipo surge quando duas ou mais observações amostrais estão situadas à mesma distância de $\underline{x}$. Neste caso, as v observações amostrais mais próximas de $\underline{x}$ podem não estar unicamente determinadas. Uma solução para este problema, su-

gerida por Cover e Hart (1967), é considerar como observação mais próxima de $\underline{x}$, entre as empatadas, aquela observação $\underline{x}_{ij}$ que tiver menor índice i (ou j , se os índices i forem iguais), $i = 1, 2$; $j = 1, \dots, n_i$. cremos, no entanto, que uma solução melhor seria considerar como observação mais próxima de $\underline{x}$, entre as empatadas:

- a observação $\underline{x}_{ij}$ que tiver menor índice j , se os índices i forem iguais, $i=1, 2$; $j=1, \dots, n_i$;
- a observação $\underline{x}_{ij}$ escolhida aleatoriamente, se os índices i forem diferentes, $i=1, 2$; $j=1, \dots, n_i$.

O segundo tipo de empate acontece quando $N_1 = N_2$ e as expressões de cada lado da regra discriminatória dada em (4.3.2) se igualam. Observemos que, nestas situações, o valor de v é par permitindo, portanto, que o número de observações de π_1 seja igual ao número de observações de π_2 entre as v mais próximas de $\underline{x}$. Uma solução para este problema, que nos parece bastante lógica, foi sugerida por Rogers e Wagner (1978): o elemento E , ao qual corresponde o vetor de medidas $\underline{x}$, será discriminado na população que deu origem à observação mais próxima de $\underline{x}$ entre as v consideradas.

O terceiro tipo de empate ocorre quando $N_1 \neq N_2$ e as expressões de cada lado da regra discriminatória dada em (4.2.3) se igualam. Observemos que este tipo de empate é solucionado classificando-se o elemento E , aleatoriamente, em π_1 ou π_2 .

c) Na escolha do valor de v , a primeira regra que levamos em consideração é a de que v deve ser fixado de modo a respeitar as condições dadas em (4.2.1).

Intuitivamente, parece-nos que quanto maior for o valor de v , respeitadas as condições acima, melhor será o comportamento do método de discriminação v -VMP (menor probabilidade de erro de classificação), pois teremos maior incidência de valores de π_i , $i = 1, 2$, na vizinhança de x .

Num trabalho de Loftsgaarden e Quesenberry (1965) sobre estimação de densidade de probabilidade não paramétrica baseada nos v vizinhos mais próximos, os autores concluíram que o valor de v deve ser escolhido próximo a $n^{1/2}$ (n é o tamanho da amostra na qual baseia-se a estimação). Tal conclusão foi obtida através de trabalhos empíricos, uma vez respeitadas as condições dadas em (4.2.1).

Ao analisarmos o problema de discriminação entre duas populações, estamos trabalhando com a estimação de duas funções densidade de probabilidade, obtidas a partir das amostras de π_1 e π_2 combinadas. Cremos que, neste caso, podemos escolher o valor de v através de uma função de $(n_1+n_2)^{1/2}$, uma vez que tal função satisfaz as condições dadas em (4.2.1). No capítulo 5 desenvolvemos um trabalho experimental, onde estimamos a probabilidade de erro de classificação do método v -VMP para diversos tamanhos de amostras e diversos valores de v , supondo que a distribuição de probabilidade de $\underline{X}$ em π_i , $i=1, 2$, fosse normal. Os resultados desse trabalho mostraram que, de modo

geral, o comportamento do método v-VMP não é influenciado pelo valor de v.

d) Na escolha da função distância d, observamos que a mais utilizada na literatura é a função distância euclidiana dada por:

$$d(\underline{x}, \underline{y}) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}, \text{ onde}$$

$$\underline{x} = (x_1, \dots, x_p) \text{ e } \underline{y} = (y_1, \dots, y_p).$$

Segundo Fix e Hodges (1952), qualquer função distância pode ser utilizada na construção do método de discriminação v-VMP, sem necessidade de tal função ser uma métrica. O fato da função distância euclidiana ser a mais utilizada é justificado por esta ser a métrica mais usual no espaço $\mathbb{R}^p$. Parece-nos que a melhor função distância é aquela que conduz à menor probabilidade de erro de classificação do método v-VMP nela baseado. No entanto, não encontramos, na literatura pesquisada, trabalhos experimentais que envolvessem comparações entre diversas funções distância, uma vez fixados diversos valores de v, p e n_i , $i = 1, 2$.

e) Patrick e Fischer (1970) demonstraram que o método v-VMP de Cover e Hart (1967) é um caso particular do método de discriminação baseado em regiões de tolerância por eles desenvolvido, apresentado na seção 3.1.2 do capítulo 3. Considerando seu método discriminante na situação em que $\frac{p_1}{n_1} = \frac{p_2}{n_2}$ e $m_1 = m_2 = m$, onde m_i , $i = 1, 2$, é o número aproximado de observações amostrais de π_i , $i = 1, 2$, encontradas na região de tolerância $S_{i\mathbf{x}}$, $i = 1, 2$, que contém $\underline{x}$ (vetor a ser classificado),

Patrick e Fischer estabeleceram a relação $m = \left\lceil \frac{v+1}{2} \right\rceil$ e concluíram que seu método de discriminação produz a mesma classificação que o método v-VMP de Cover e Hart.

4.2.3. UM EXEMPLO NUMÉRICO

Consideremos a situação apresentada na seção 3.1.4 do capítulo 3.

Discriminaremos o vetor $\underline{x}'_0 = (5, 7, 2, 8)$, associado ao elemento E, através do método v-VMP apresentado nas seções 4.2.1 e 4.2.2.

Temos $n_1 = n_2 = 8$ e $p_1 = p_2 = 0,5$. Fixemos $v = (n_1 + n_2)^{1/2} = 16^{1/2} = 4$, isto é, fixemos $v = 4$.

Utilizando a função distância euclidiana, obtivemos que os vetores amostrais $\underline{x}_{16}$, $\underline{x}_{22}$, $\underline{x}_{14}$ e $\underline{x}_{24}$ são os quatro mais próximos de $\underline{x}$, sendo que a quarta menor distância é dada por $U = 0,54$.

Neste caso, temos $N_1 = 2$ e $N_2 = 2$.

Através da consideração (b) dada na seção 4.2.2, classificamos o elemento E na população que deu origem ao vetor amostral mais próximo de $\underline{x}$, isto é, em π_1 (Iris Versicolor).

4.2.4. CONSISTÊNCIA DO MÉTODO v-VMP

Como vimos na seção 4.2.1, a regra discriminante dada em (4.2.3) baseia-se somente no número de observações das amos-

tras de π_1 e π_2 combinadas pertencentes a $S_{\underline{x}}$ (círculo de raio U centrado em $\underline{x}$), isto é, baseia-se somente no número de observações das amostras de π_1 e π_2 combinadas cujas distâncias até $\underline{x}$ são menores ou iguais a U .

Observemos que a regra discriminante dada em (4.2.3) seria a mesma se substituíssemos $\hat{f}_i(\underline{x})$, $i = 1, 2$, dado em (4.2.2), por $\hat{f}_{i\underline{x}}(0)$, $i = 1, 2$, na regra da razão de verossimilhança, sendo que

$$\hat{f}_{i\underline{x}}(0) = \frac{N_i}{n_i} \cdot \frac{1}{U} \quad (4.2.5)$$

é um estimador da função densidade de probabilidade de $d(\underline{X}_i, \underline{x})$ no ponto zero, $f_{i\underline{x}}(0)$.

Pelo fato de $\hat{f}_{i\underline{x}}(0)$, $i = 1, 2$, ser consistente com $f_{i\underline{x}}(0)$ (vide demonstração apresentada no apêndice C, desenvolvida por Fix e Hodges, 1951) e pelo teorema 2, apresentado na seção 3.1.5 do capítulo 3 e demonstrado no apêndice A, verificamos que o método não paramétrico de discriminação forçada v-VMP é consistente com o método (ótimo) da razão de verossimilhança.

Devemos observar que os dois únicos métodos de discriminação forçada baseados em regiões de tolerância consistentes com o método da razão de verossimilhança são o v-VMP e o de Patrick e Fischer (1970), apresentado na seção 3.1.2 do capítulo 3. No entanto, verificamos que somente o método v-VMP é utilizado em trabalhos aplicados. Tal fato é justificado pela simplicidade de construção deste método discriminante, que não exige o cálculo do volume da região de tolerância que contém $\underline{x}$.

4.2.5. PROBABILIDADE DE ERRO DE CLASSIFICAÇÃO

Uma vez exposto o método de discriminação forçada baseado nos v vizinhos mais próximos, nosso próximo passo será medir-lhe o desempenho, em termos da probabilidade de erro de classificação.

Em estudos realizados por Fix e Hodges (1952) e Cover e Hart (1967) verificamos que o cálculo da probabilidade exata de erro de classificação baseia-se, somente, no número de observações amostrais de π_i , $i = 1, 2$, encontradas entre as v mais próximas de $\underline{x}$, isto é, só foi desenvolvido para o caso em que $n_1 = n_2$ e $p_1 = p_2$. Embora supondo tais simplificações, estes autores não deduziram uma expressão única para a probabilidade exata de erro de classificação do método v -VMP devido à complexidade dos cálculos envolvidos.

Como apresentado na seção 4.2.2, o método de discriminação v -VMP de Cover e Hart (1967) é um caso particular do método de discriminação de Patrick e Fischer (1970) quando $\frac{p_1}{n_1} = \frac{p_2}{n_2}$ e $m = \left\lfloor \frac{v+1}{2} \right\rfloor$. Portanto, a probabilidade exata de erro de classificação deste método v -VMP pode ser calculada através da expressão da probabilidade exata de erro de classificação do método de Patrick e Fischer (vide apêndice D). Observemos, no entanto, que quando a dimensionalidade p das populações e/ou o número v de vizinhos mais próximos considerado aumenta, o cálculo da probabilidade exata de erro de classificação do método de discriminação v -VMP de Cover e Hart torna-se bastante

trabalhoso, exigindo o emprego de recursos de cálculo numérico para sua obtenção.

A seguir, apresentaremos o resultado assintótico, obtido por Cover e Hart (1967) sobre a probabilidade exata de erro de classificação de sua regra discriminante.

TEOREMA. Seja f_i , $i = 1, 2$, uma função densidade de probabilidade tal que, com probabilidade 1, $\underline{x}$ é um ponto de continuidade de f_i ou um ponto com probabilidade diferente de zero.

A probabilidade exata de erro de classificação, $P_{e,VMP}$, para a regra discriminatória de Cover e Hart, possui os seguintes limites assintóticos:

$$P^* \leq P_{e,VMP} \leq 2P^*(1-P^*),$$

onde P^* é a probabilidade de erro de classificação do método (ótimo) da razão de verossimilhança. $\square$

A demonstração deste teorema pode ser encontrada em Cover e Hart (1967).

Analisando o teorema acima temos que o valor máximo do limite superior para $P_{e,VMP}$ é $\frac{1}{2}$.

Observemos que os limites assintóticos para $P_{e,VMP}$ são preferíveis ao seu cálculo exato, se este for muito trabalhoso e o valor de P^* for de fácil obtenção.

Considerando o caso geral, isto é, o método de discriminação forçada v-VMP de Fix e Hodges (1951), definido pela regra discriminatória dada em (4.2.3), verificamos que a literatura pesquisada não traz indicação alguma sobre o método de

cálculo da probabilidade exata de erro de classificação, uma vez conhecida a função densidade de probabilidade $f_i, i=1,2$.

Nesta situação e na situação em que f_i é desconhecida, uma solução para o cálculo da probabilidade de erro de classificação da regra v-VMP é a sua estimação, através de algum método apresentado na seção 6 do capítulo 1.

Dos três métodos de estimação apresentados naquela seção, o método de estimação por resubstituição é o menos adequado, pois conduz a $\hat{P}_{e,VMP} = 0$, quando $v=1$. Mostraremos este resultado através do seguinte argumento:

Seja $\{x_{11}, \dots, x_{1n_1}\}$ uma amostra retirada da população π_1 e $\{x_{21}, \dots, x_{2n_2}\}$ uma amostra retirada da população π_2 . Construamos o método 1-VMP de discriminação, através das observações amostrais de π_1 e π_2 e discriminemos cada um dos vetores amostrais x_{ij} , $i = 1,2$; $j = 1, \dots, n_i$.

Supondo que empates não ocorram, a menor distância obtida, a cada discriminação de um vetor amostral, é igual a $d(x_{ij}, x_{ij}) = 0$.

Nesse caso, teremos cada elemento amostral classificado em sua população de origem e, portanto, uma fração de elementos classificados erradamente igual a zero, isto é, $P_{e,VMP}=0$.

Para $v > 1$, cremos que $P_{e,VMP}$, obtida pelo método de estimação por resubstituição, subestima $P_{e,VMP}$, pois o vetor amostral que está sendo classificado sempre contribui com a menor

distância entre as v consideradas, podendo deste modo implicar numa grande proporção de observações amostrais classificadas corretamente em sua população de origem.

Segundo Lachenbruch (1975) e Penrod e Wagner (1979), o melhor método de estimação da probabilidade de erro de classificação da regra discriminante v -VMP é o método de estimação por exclusão, uma vez que as probabilidades por ele geradas estão bastante próximas das reais, mesmo que sejam utilizadas amostras de pequenos tamanhos.

4.3. MÉTODO DE DISCRIMINAÇÃO PARCIAL

Consideremos o método de discriminação forçada baseado nos v vizinhos próximos até agora analisado. Notemos que muitas vezes, o número de observações de π_1 é próximo (v ímpar) ou igual (v par) ao número de observações de π_2 encontradas entre as v fixadas. Neste caso, a classificação do elemento E torna-se, praticamente, arbitrária. Nesta situação, seria melhor não classificar o elemento E , isto é, seria conveniente construir uma região de dúvida tal que se o elemento E a ela pertencer, não tomaremos qualquer decisão quanto à sua classificação.

4.3.1. REGRA DISCRIMINATÓRIA

O método de discriminação parcial baseado nos v vizinhos mais próximos desenvolvido por Hellman (1970) pode ser apre-

sentado da seguinte maneira:

Fixemos dois números inteiros e positivos, v e v' , tais que:

$$1 \leq \frac{v}{2} + 1 \leq v' < v$$

Combinemos as amostras de π_1 e π_2 , calculemos e ordene-mos as distâncias $d(\underline{x}_{ij}, \underline{x})$, $i = 1, 2$; $j = 1, \dots, n_i$, onde $\underline{x}$ é o vetor a ser classificado, associado ao elemento E .

Obtenhamos as v observações amostrais mais próximas de $\underline{x}$ e contemos o número N_i , $i = 1, 2$, de observações amostrais de π_i , $i = 1, 2$, entre as v obtidas ($N_1 + N_2 = v$).

Obteremos a seguinte regra discriminatória:

Se $N_1 \geq v'$, classificaremos o elemento E em π_1

Se $N_2 \geq v'$, classificaremos o elemento E em π_2

Se $v - v' < N_i < v'$, $i = 1, 2$, não classificaremos o elemento E

Para exemplificar este método de discriminação parcial consideremos uma situação na qual fixamos $v = 7$ e $v' = 5$. Na tabela abaixo estão registrados os possíveis pares de valores de N_1 e N_2 e a regra discriminatória associada:

N_1	N_2	
7	0	E será classificado em π_1
6	1	
5	2	
4	3	
3	4	região de dúvida
2	5	
1	6	E será classificado em π_2
0	7	

Observemos que o método de discriminação parcial baseado nos v vizinhos mais próximos não é ponderado pelo tamanho da amostra retirada de cada população, nem pela probabilidade a priori p_i , $i = 1, 2$, do elemento E pertencer à população π_i , $i = 1, 2$.

Esta característica pode acarretar uma grande probabilidade de erro de classificação quando os tamanhos das amostras retiradas de π_1 e π_2 forem muito diferentes.

4.3.2. PROBABILIDADES DE ERRO DE CLASSIFICAÇÃO E DE NÃO CLASSIFICAÇÃO

As probabilidades de erro de classificação e de não classificação do método de discriminação parcial baseado nos v vizinhos mais próximos dependem de v e v' .

Quanto maior for a região de dúvida, maior será a probabilidade de não classificação e menor a probabilidade de erro de classificação. Quanto menor for a região de dúvida, menor será a probabilidade de não classificação e maior será a probabilidade de erro de classificação. Embora a literatura pesquisada não traga trabalhos teóricos ou aplicados a esse respeito, cremos que o experimentador deve escolher v e v' de modo que tais probabilidades se equilibrem e não sejam muito grandes. Para isto, pode estimar as probabilidades de erro de classificação e de não classificação, através de algum método apresentado na seção 6 do capítulo 1, considerando diferentes situações experimentais, isto é, diferentes distribuições

de $f_i(x)$, $i = 1, 2$, diferentes tamanhos de amostras e diferentes valores de v e v' .

CAPÍTULO 5

RESULTADOS EXPERIMENTAIS

Este capítulo tem os seguintes objetivos:

- a) apresentar os resultados experimentais, encontrados na literatura pesquisada (vide seção 5.2);
- b) discutir o desempenho do método de discriminação baseado nos v vizinhos mais próximos (v -VMP) em função do valor de v (vide seção 5.3.1), e
- c) comparar os seguintes métodos de discriminação forçada: linear amostral, quadrático amostral, baseado nos v vizinhos mais próximos e baseado em regiões de tolerância de Gessaman e Gessaman (1972) (vide seção 5.3.2).

Os testes para as comparações citadas em (b) e (c) foram por nós realizados.

5.1. NOTAÇÃO

Assumimos que para indivíduos da população π_i , $i = 1, 2$, o vetor aleatório p -dimensional $\underline{X}$ é normalmente distribuído com média $\underline{\mu}_i$ e matriz de covariância $\underline{\Sigma}_i$, e que a probabilidade a priori de que um indivíduo seja de π_i é p_i . Construímos a regra discriminante dos métodos linear amostral, quadrático amostral, baseado nos v vizinhos mais próximos e baseado em regiões de

tolerância a partir de duas amostras de tamanho n_i , com médias amostrais $\bar{X}_i$ e matrizes de covariância amostrais S_i , $i=1,2$.

O desempenho de uma regra discriminante foi avaliado através da probabilidade total do erro de classificação $P = P_1(\pi_1|\pi_2) + P_2(\pi_1|\pi_2)$, onde $P(\pi_j|\pi_i)$ é a probabilidade de que um indivíduo de π_i seja classificado erradamente em π_j , $i,j=1,2$; $i \neq j$.

Utilizamos a seguinte notação para algumas probabilidades de interesse:

P_L e P_Q são, respectivamente, as probabilidades (totais) exatas do erro de classificação para os métodos ótimos de discriminação linear e quadrático.

$\hat{P}_Q$ é a probabilidade (total) exata do erro de classificação para o método quadrático ótimo de discriminação, estimada por simulação (vide seção 6 do capítulo 1).

5.2. RESULTADOS EXPERIMENTAIS ENCONTRADOS NA LITERATURA PESQUISADA

Na literatura pesquisada encontramos resultados experimentais onde são comparados:

- o método de discriminação linear amostral e o baseado nos vizinhos mais próximos;
- o método de discriminação linear amostral e o baseado em regiões de tolerância de Gessaman e Gessaman (1972), considerando

do discriminação forçada, e,

- o método de discriminação parcial baseado em regiões de tolerância e o de Kendall (1966).

5.2.1. COMPARAÇÃO ENTRE O MÉTODO DISCRIMINANTE LINEAR AMOSTRAL E O BASEADO NOS V VIZINHOS MAIS PRÓXIMOS (FIX E HODGES, 1952)

Os autores avaliaram o desempenho do método v-VMP em situações onde o método linear amostral era o mais apropriado. Para tal avaliação, analisaram casos muito particulares tais como, $p = 1$ ou $v = 1$ ou $n_1 = n_2 = 1$. Não apresentamos nesta seção, os resultados numéricos obtidos por Fix e Hodges, pois, através dos casos analisados por estes autores, não pudemos tirar conclusões sobre o desempenho do método v-VMP.

5.2.2. COMPARAÇÃO ENTRE O MÉTODO DISCRIMINANTE LINEAR AMOSTRAL E O DE GESSAMAN E GESSAMAN (1972) BASEADO EM REGIÕES DE TOLERÂNCIA-DISCRIMINAÇÃO FORÇADA (GESSAMAN E GESSAMAN, 1972).

Os autores compararam os desempenhos dos métodos linear amostral e de discriminação forçada baseado em regiões de tolerância, por eles desenvolvido.

Ao longo deste estudo utilizou-se $p = 2$, $p_1 = 0,5$ e três pares de distribuições normais:

(a) $\mu_1 = \underline{0}$ e $\Sigma_1 = \underline{I}$ (matriz identidade)

$$\underline{\mu}_2' = (2, 0), \text{ e } \underline{\Sigma}_2 = \begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}$$

(b) $\underline{\mu}_1 = \underline{0} \text{ e } \underline{\Sigma}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$

$$\underline{\mu}_2 = \underline{0} \text{ e } \underline{\Sigma}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

(c) $\underline{\mu}_1' = (1, 1) \text{ e } \underline{\Sigma}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$

$$\underline{\mu}_2' = (-1, 0) \text{ e } \underline{\Sigma}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

Convém observar que não encontramos quaisquer relações entre as matrizes de covariância adotadas.

Os métodos de discriminação linear amostral e de Gessaman e Gessaman foram construídos através de amostras de tamanhos $n = n_1 = n_2 = 64, 200 \text{ e } 729$, geradas pelos pares (a), (b) e (c) de distribuições normais. Os desempenhos destes métodos foram medidos, calculando-se as estimativas, por partição, das probabilidades totais de erro de classificação através da discriminação de 500 observações adicionais, 250 de π_1 e 250 de π_2 .

TABELA 5.2.1. PROBABILIDADE TOTAL DE ERRO DE CLASSIFICAÇÃO, ESTIMADA POR PARTIÇÃO, PARA OS MÉTODOS LINEAR AMOSTRAL E DE GESSAMAN E GESSAMAN (1972), SUPONDO DISTRIBUIÇÃO NORMAL BIVARIADA

Método	(a)			(b)			(c)*		
	64	200	729	64	200	729	64	200	729**
Linear amostral	0,112	0,118	0,116	0,498	0,496	0,502	0,166	0,158	0,150
Gess. e Gess.	0,098	0,096	0,090	0,304	0,292	0,246	0,178	0,158	0,142

* casos analisados

** tamanho da amostra (n)

Analisando os resultados da Tabela 5.2.1, observamos que o método linear amostral tem pior desempenho que o método de Gessaman e Gessaman para as distribuições normais dos casos (a) e (b), por causa da desigualdade entre as matrizes de covariância. Contudo, seu desempenho no caso (c) é melhor que o método de Gessaman e Gessaman, por causa da igualdade entre as matrizes de covariância.

Uma falha deste estudo reside no fato dele não apresentar as probabilidades (totais) exatas do erro de classificação para os métodos ótimos de discriminação quadrático, apropriado aos casos (a) e (b), e linear, apropriado ao caso (c), a fim de que pudéssemos compará-las às probabilidades (totais) estimadas do erro de classificação dos métodos linear amostral e de Gessaman e Gessaman. Na Tabela 5.3.2, apresentada na seção 5.3.2, estão os valores que calculamos para as probabilidades (totais) exatas do erro de classificação dos métodos ótimos de discriminação apropriados aos casos (a), (b) e (c) analisados; comparando os valores destas probabilidades com os da Tabela

5.2.1, notamos que o método de Gessaman e Gessaman tem um bom desempenho em todos os casos analisados.

5.2.3. COMPARAÇÃO ENTRE OS MÉTODOS DE DISCRIMINAÇÃO PARCIAL DE GESSEMAN e GESSEMAN (1972), QUESENBERRY e GESSEMAN (1968) E KENDAL (1966) (GESSAMAN e GESSEMAN, 1972)

Consideremos as seguintes notações:

GG(1): método de Gessaman e Gessaman (1972), baseado em regiões de tolerância, sem controle das probabilidades de erro de classificação;

K : método de Kendall (1966), sem controle das probabilidades de erro de classificação;

GG(2): método de Gessaman e Gessaman (1972), baseado em regiões de tolerância, com controle das probabilidades de erro de classificação, e

QG : método de Quesenberry e Gessaman (1968), baseado em regiões de tolerância, com controle das probabilidades de erro de classificação.

Gessaman e Gessaman (1972) compararam os desempenhos dos métodos GG(1), K, GG(2) e QG. Os casos teste escolhidos para esta comparação, foram os mesmos apresentados na seção 5.2.2, sendo que foi incluída a estimativa, por partição, da probabilidade de não classificação.

A Tabela 5.2.2 fornece as estimativas por partição das

probabilidades totais do erro de classificação ($\hat{P}_{EC}$) e de não classificação ($\hat{P}_{NC}$), para os métodos GG(1), K, GG(2) e QG, construídas a partir de amostras de tamanhos 64, 200 e 729, geradas pelos pares (a), (b) e (c) de distribuições normais.

Dividimos a análise dos resultados apresentados na Tabela 5.2.2 em dois casos:

CASO A. COMPARAÇÃO ENTRE OS MÉTODOS K E GG(1)

O método K conduz a um valor alto de $\hat{P}_{NC}$ mesmo nas situações (a) e (c) onde as distribuições têm médias diferentes. Quando n decresce de 729 para 200, nota-se pouca variação no desempenho do método K. Contudo, quando n decresce para 64, surgem algumas variações; $\hat{P}_{EC}$ cresce e $\hat{P}_{NC}$ decresce. Em comparação com o método K, o método GG(1) mostra valores de $\hat{P}_{NC}$ bem menores. Esta melhora em $\hat{P}_{NC}$ se deve, parcialmente, ao aumento em $\hat{P}_{EC}$. É interessante notar que, em geral, quando n cresce, $\hat{P}_{EC}$ decresce e $\hat{P}_{NC}$ cresce.

CASO B. COMPARAÇÃO ENTRE OS MÉTODOS GG(2) E QG

Fixado o limite 0,1 para a probabilidade de erro de classificação, temos que os valores de $\hat{P}_{EC}$ do método QG nunca ultrapassam 0,05. Contudo, os valores de $\hat{P}_{NC}$ nunca são inferiores a 0,333. Em comparação, o método GG(2) tem valores da probabilidade total de classificação correta, estimada por par-tição, bem maiores, e menores valores de $\hat{P}_{NC}$. O efeito do ta-

TABELA 5.2.2. VALORES DA PROBABILIDADE TOTAL DE ERRO DE CLASSIFICAÇÃO E DE NÃO CLASSIFICAÇÃO, PARA OS MÉTODOS GG(1), K, GG(2) E QG, SUPONDO DISTRIBUIÇÃO NORMAL BIVARIADA

Método	(a)						(b)						(c)*					
	64			200			729			64			200			729**		
	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}$	$\hat{P}_{EC}$	$\hat{P}_{NC}^{***}$
K	0,036	0,290	0,002	0,508	0,002	0,686	0,062	0,876	0,012	0,966	0,002	0,996	0,024	0,492	0,004	0,826	0,004	0,846
GG(1)																		
(i) amostra da população π_1	0,152	0,032	0,098	0,024	0,078	0,076	0,242	0,126	0,194	0,256	0,092	0,400	0,174	0,000	0,128	0,056	0,102	0,102
(ii) amostra da população π_2	0,098	0,000	0,072	0,066	0,032	0,166	0,254	0,144	0,202	0,200	0,698	0,382	0,184	0,000	0,122	0,108	0,084	0,160
GG(2)	0,054	0,206	0,074	0,128	0,082	0,100	0,100	0,540	0,132	0,426	0,082	0,614	0,138	0,276	0,150	0,028	0,156	0,124
QG	0,040	0,336	0,020	0,428	0,028	0,394	0,042	0,800	0,018	0,828	0,010	0,872	0,028	0,522	0,024	0,544	0,016	0,124

* - casos analisados
 ** - tamanho da amostra (n)
 *** - $\hat{P}_{EC}$ é a probabilidade total de erro de classificação estimada por partição
 $\hat{P}_{NC}$ é a probabilidade de não classificação estimada por partição

manho da amostra não está claro; parece não haver uma tendência em ambos os métodos, quando n cresce.

5.3. OUTROS RESULTADOS EXPERIMENTAIS

A partir dos resultados experimentais encontrados na literatura pesquisada, observamos que poucos são os que tratam do método baseado nos v vizinhos mais próximos, no que se refere ao valor de v a ser adotado e quanto à sua comparação com outros métodos de discriminação baseados em regiões de tolerância. Seguindo a linha adotada por Gessaman e Gessaman (1972), decidimos realizar alguns testes com o método não paramétrico de discriminação forçada v -VMP, em situações onde o método discriminante linear e/ou o método discriminante quadrático são os mais apropriados, servindo como critério para avaliar o desempenho do método v -VMP. Para tanto, consideramos os mesmos casos teste utilizados por Gessaman e Gessaman (1972), apresentados na seção 5.2.2. Observamos que outros testes podem ser realizados com o objetivo de estender este estudo a outras distribuições.

5.3.1. MÉTODO v -VMP: DISCUSSÃO SOBRE O VALOR DE v

Nas considerações sobre a escolha do valor de v , apresentada na seção 4.2.2 do Capítulo 4, observamos que o valor de v poderia ser escolhido como uma função de $(n_1+n_2)^{1/2}$. Como, nos casos por nós analisados, as amostras, geradas pelos pares (a), (b) e (c) de distribuições normais, tinham tamanhos $n = n_1 = n_2 = 64, 200$ e 729 , utilizamos os seguintes valores de

v:

- para $n = 64$, $v = 3, 9, 11, 23$ e 33 ;
- para $n = 200$, $v = 3, 5, 15, 21, 41, 61, 81$ e 101 , e
- para $n = 729$, $v = 5, 7, 27, 39, 77, 115, 153, 191, 229, 267, 305, 343$ e 381 .

Observemos que além de adotarmos, arbitrariamente, alguns valores pequenos para v , adotamos, também, os valores $n^{1/2}$, $(2n)^{1/2}$, $(2n)^{1/2}$, etc., até atingirmos um valor de v próximo a $n/2$. Dessa forma, garantimos que o valor de v fosse sempre menor que o valor de n (vide as condições (4.2.1) impostas sobre v , apresentadas na seção 4.2.1 do Capítulo 4). Além disso, adotamos somente valores ímpares para v , a fim de evitar problemas com empates (vide seção 4.2.2 do Capítulo 4).

Um programa de computação sobre o método v-VMP (vide Apêndice E) foi processado no CCE-USP, num total de dez rodadas, para cada conjunto de valores de v , p e n e cada par (a), (b) e (c), de distribuições normais, gerando em cada processamento pares de amostras diferentes. O fato de não processarmos o programa sobre o método v-VMP mais do que dez vezes, para cada caso, se deve ao fato de cada rodada consumir um tempo de processamento muito alto (para $n = 729$, o tempo é de aproximadamente 15 minutos para cada rodada!)

A Tabela 5.3.1 apresenta, em função de v , valores estimados, por partição, da probabilidade total de erro de classificação ($\hat{P}_{EC}$), para o método v-VMP, construído a partir de amostras de tamanhos 64, 200 e 729, geradas pelos pares (a), (b) e (c) de distribuições normais. Convém observar que os va-

TABELA 5.3.1 VALORES DA PROBABILIDADE TOTAL DE ERRO DE CLASSIFICAÇÃO, ESTIMADA POR PARTIÇÃO, PARA O MÉTODO v-VMP, SUPONDO DISTRIBUIÇÃO NORMAL BIVARIADA

n = 64			
v	(a)	(b)	(c) *
3	0,110	0,393	0,195
9	0,107	0,371	0,175
11	0,103	0,367	0,168
23	0,106	0,353	0,171
33	0,109	0,374	0,173

n = 200			
v	(a)	(b)	(c)
3	0,117	0,385	0,186
5	0,108	0,372	0,180
15	0,103	0,371	0,164
21	0,098	0,355	0,162
41	0,101	0,352	0,166
61	0,099	0,354	0,169
81	0,102	0,359	0,172
101	0,106	0,362	0,176

n = 729			
v	(a)	(b)	(c)
5	0,111	0,365	0,175
7	0,110	0,358	0,170
27	0,101	0,333	0,150
39	0,097	0,328	0,149
77	0,097	0,329	0,149
115	0,097	0,328	0,153
153	0,099	0,334	0,153
191	0,098	0,328	0,156
229	0,098	0,330	0,157
267	0,099	0,328	0,160
305	0,098	0,329	0,161
343	0,099	0,329	0,164
381	0,098	0,330	0,165

* casos analisados.

lores de $\hat{P}_{EC}$ são a média aritmética dos valores obtidos em cada uma das dez rodadas do programa v-VMP.

Os resultados da Tabela 5.3.1 indicam que o desempenho do método v-VMP não é influenciado, significativamente, por nenhum dos valores de v adotados, pois considerando separadamente cada um dos casos (a), (b) e (c) analisados e cada valor de n , observamos que a diferença entre os valores estimados, por partição, da probabilidade total do erro de classificação nunca superou 4% do menor valor.

5.3.2. COMPARAÇÃO ENTRE OS MÉTODOS DE DISCRIMINAÇÃO LINEAR AMOSTRAL, QUADRÁTICO AMOSTRAL, v VIZINHOS MAIS PRÓXIMOS E DE GESSAMAN E GESSAMAN (1972) - DISCRIMINAÇÃO FORÇADA

Consideramos os mesmos casos teste apresentados na seção 5.2.2.

A fim de realizarmos uma comparação adequada entre os métodos de discriminação linear amostral, quadrático amostral, v vizinhos mais próximos e de Gessaman e Gessaman (1972), considerando discriminação forçada, calculamos, em primeiro lugar, os valores das probabilidades totais do erro de classificação para os métodos ótimos de discriminação apropriados aos casos (a), (b) e (c). Como visto na seção 5.2.2, o método ótimo de discriminação apropriado aos casos (a) e (b) analisados é o quadrático, enquanto para o caso (c) é o linear. Na seção 2 do Capítulo 1 vimos que o cálculo da probabilidade exata do erro de classificação, P_Q , para o método quadrático é muito trabalhoso e que uma alternativa ao cálculo dessa probabilidade

de seria estimá-la por simulação (vide seção 6 do Capítulo 1). Um programa de computação sobre o método ótimo quadrático de discriminação (vide Apêndice E) foi processado no CCE-USP, num total de dez rodadas, para cada um dos casos (a) e (b); a cada rodada, o desempenho do método quadrático foi medido, calculando-se a média aritmética dos valores estimados, por simulação, das probabilidades totais do erro de classificação, $\hat{P}_Q$, através de 1.400 observações adicionais (diferentes a cada processamento), 700 de π_1 e 700 de π_2 . As médias dos valores de $\hat{P}_Q$, para os casos (a) e (b), encontram-se na Tabela 5.3.2. O valor da probabilidade total exata do erro de classificação, P_L , para o método ótimo de discriminação linear apropriado ao caso (c) foi obtido através da expressão (2.5) apresentada na seção 2 do Capítulo 1 e encontra-se na Tabela 5.3.2.

TABELA 5.3.2 VALORES DA PROBABILIDADE TOTAL DO ERRO DE CLASSIFICAÇÃO, PARA OS MÉTODOS LINEAR E QUADRÁTICO, SUPONDO DISTRIBUIÇÃO NORMAL BIVARIADA

	(a)	(b)	(c) *
Probabilidade total do erro de classificação	0,097	0,326	0,149

* casos analisados

A Tabela 5.3.3 fornece as estimativas, por partição, das probabilidades totais do erro de classificação para os métodos linear amostral, quadrático amostral, v vizinhos mais próximos e de Gessaman e Gessaman (1972), considerando discriminação forçada. Tais métodos foram construídos através de amostras de tamanhos 64, 200 e 729, geradas pelos pares (a), (b) e (c) de

distribuições normais.

Observemos que:

- i) para o método de Gessaman e Gessaman, os valores foram extraídos do artigo de Gessaman e Gessaman (1972), sendo que os autores não fazem referência à forma como foram obtidos.
- ii) um programa para os métodos de discriminação linear amostral e quadrático amostral (vide Apêndice E) foi processado no CCE-USP num total de dez rodadas para cada conjunto de valores de p_1 e n e para cada um dos casos (a), (b) e (c), gerando-se em cada processamento pares de amostras diferentes. O programa de computação forneceu estimativas, por partição, das probabilidades totais do erro de classificação para os métodos linear amostral e quadrático amostral, obtidas através da classificação de 500 observações adicionais, 250 de π_1 e 250 de π_2 . Os valores apresentados na Tabela 5.3.3 representam a média aritmética dessas estimativas.
- iii) para o método v vizinhos mais próximos, consideramos os valores mínimos e máximos, apresentados na Tabela 5.3.1, para cada um dos casos analisados.

Os resultados da Tabela 5.3.3 indicam que no caso (a), onde as matrizes de covariância são diferentes, o método de discriminação linear amostral é o que apresenta pior desempenho, como era esperado. O método de discriminação quadrático amostral é o que apresenta melhor desempenho, como também era es-

TABELA 5.3.3. VALORES DA PROBABILIDADE TOTAL DE ERRO DE CLASSIFICAÇÃO, ESTIMADA POR PARTIÇÃO, PARA OS MÉTODOS LINEAR AMOSTRAL, QUADRÁTICO AMOSTRAL, VVI-ZINHOS MAIS PRÓXIMOS E DE GESSERAN E GESSERAN (1972). SUPONDO DISTRIBUIÇÃO NORMAL BIVARIADA

Método	(a)		(b)		(c)*	
	64	200	64	200	64	200
Método Amostral	0,131	0,131	0,502	0,499	0,161	0,150
Quadrático Amostral	0,058	0,097	0,331	0,327	0,160	0,152
V-VI?	de 0,103 a 0,110 de 0,098 a 0,117 de 0,097 a 0,111		de 0,353 a 0,393 de 0,352 a 0,385 de 0,328 a 0,365		de 0,168 a 0,195 de 0,162 a 0,166 de 0,149 a 0,175	
Gesseran e Gesseran	0,098	0,096	0,304	0,292	0,178	0,158
						0,142

* - casos analisados

** - tamanho da amostra (n)

perado, pois é o apropriado a este caso. Os métodos não paramétricos de discriminação v vizinhos mais próximos e de Gessaman e Gessaman apresentam bons desempenhos, com probabilidades próximas a do método ótimo (0,097). Tais desempenhos são superiores ao do método linear amostral sugerindo que, no caso de distribuições normais com matrizes de covariância distintas, os métodos não paramétricos de discriminação v vizinhos mais próximos e de Gessaman e Gessaman representam uma boa alternativa para o método quadrático amostral, especialmente no uso de micro-computadores para casos com maior dimensionalidade, onde o método quadrático amostral exige inversão de matrizes de elevada ordem. Considerando o caso (b), chegamos às mesmas conclusões do caso (a), com as seguintes ressalvas:

- i) a probabilidade total do erro de classificação estimada, por simulação, para o método ótimo quadrático de discriminação é, agora, igual a 0,326.
- ii) a igualdade entre as médias μ_1 e μ_2 afeta o desempenho dos métodos de discriminação analisados, de modo que as probabilidades totais do erro de classificação, estimadas, por partição, são altas.

Quanto ao caso (c), onde as matrizes de covariância são iguais, observamos que os métodos de discriminação linear amostral e quadrático amostral têm desempenhos semelhantes, pois, nesta situação, o método linear amostral é um caso particular do método quadrático amostral. Além disso, estes métodos de discriminação são os de melhor desempenho por serem apropriados a este caso. Os métodos não paramétricos de discriminação v vizinhos mais próximos e de Gessaman e Gessaman apresentam

um bom desempenho, se levarmos em consideração o valor da probabilidade (total) exata do erro de classificação para o método ótimo de discriminação linear que é igual a 0,149 (vide Tabela 5.3.2).

Cumpre observar, também, que, para quaisquer um dos casos, (a), (b) ou (c), analisados, a diferença entre o desempenho do método v vizinhos mais próximos e de Gessaman e Gessaman é muito pequena.

Concluimos, portanto, que ao considerarmos as distribuições normais dos casos (a), (b) e (c), não observamos diferenças significativas entre as probabilidades totais estimadas, por partição, do erro de classificação para os métodos de discriminação analisados. Neste caso, existe uma vantagem ao utilizarmos os métodos não paramétricos de discriminação v vizinhos mais próximos e de Gessaman e Gessaman: facilidade de construção e aplicação. Entre os dois métodos, preferimos o v vizinhos mais próximos por ser consistente com o método (ótimo) da razão de verossimilhança.

APÊNDICE A

O teorema apresentado neste apêndice foi formulado por Fix e Hodges (1951) e por nós demonstrado.

TEOREMA. Seja $\hat{f}_i(\underline{x})$ um estimador de $f_i(\underline{x})$, $i = 1, 2$, que depende dos tamanhos n_1 e n_2 das amostras retiradas das populações π_1 e π_2 , respectivamente.

Se $\hat{f}_i(\underline{x})$ for consistente com $f_i(\underline{x})$, para todo $\underline{x}$, então o método de discriminação baseado em $\hat{f}_i(\underline{x})$ é consistente com o método da razão da verossimilhança.

DEMONSTRAÇÃO. Sejam

$$c = \frac{p_2}{p_1}$$

e $\epsilon > 0$, números fixados.

Usando a consistência de $\hat{f}_i(\underline{x})$, encontremos $\delta > 0$ tão pequeno, M_1 e M_2 , números inteiros tão grandes, de modo que quando $n_1 > M_1$ e $n_2 > M_2$, tenhamos

$$P[|\hat{f}_1(\underline{x}) - f_1(\underline{x})| > \delta] < \frac{\epsilon}{2}$$

e

$$P[|c\hat{f}_2(\underline{x}) - cf_2(\underline{x})| > \delta] < \frac{\epsilon}{2}$$

Denotemos por $L(c)$, o método da razão de verossimilhança

e por $L^*(c, \hat{f}_1, \hat{f}_2)$, o método de discriminação onde $\hat{f}_i$ é um estimador consistente de f_i . Temos:

$$\begin{aligned} P[|L(c) - L^*(c, \hat{f}_1, \hat{f}_2)| > \delta] &= \\ &= P\left\{ |[f_1(\underline{x}) - cf_2(\underline{x})] - [\hat{f}_1(\underline{x}) - c\hat{f}_2(\underline{x})]| > \delta \right\} \leq \\ &\leq P[|f_1(\underline{x}) - \hat{f}_1(\underline{x})| > \delta] + \\ &+ P[|cf_2(\underline{x}) - c\hat{f}_2(\underline{x})| > \delta] \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Logo, $P[|L(c) - L^*(c, \hat{f}_1, \hat{f}_2)| > \delta] < \epsilon$, isto é, a diferença entre $L(c)$ e $L^*(c, \hat{f}_1, \hat{f}_2)$ cresce com probabilidade menor que ϵ . Em outras palavras, temos que $L^*(c, \hat{f}_1, \hat{f}_2)$ é consistente com $L(c)$.

APÊNDICE B

O teorema que formularemos e demonstraremos a seguir, foi baseado no apresentado por Gessaman (1970).

TEOREMA. Seja $\{\underline{x}_1, \dots, \underline{x}_n\}$ uma amostra da variável aleatória p -dimensional $\underline{X}$, cuja função de distribuição denotaremos por F . Suponhamos que F seja contínua e que a correspondente função densidade de probabilidade seja denotada por f . Seja $\underline{x}$, um ponto pertencente a S_p onde $f(\underline{x})$ é positiva.

Através da amostra de $\underline{X}$, construamos regiões de tolerância segundo o método de subpartição de regiões de tolerância, de modo que cada uma contenha, aproximadamente, m observações amostrais, sendo m uma função de n que satisfaz a seguinte condição: m cresce sempre que n cresce, mas de modo mais lento que n .

Denotemos por $S_{\underline{x}}$, a região de tolerância que contém $\underline{x}$ e por $V_{S_{\underline{x}}}$, o volume de $S_{\underline{x}}$.

Contemos o número de observações amostrais pertencentes a $S_{\underline{x}}$ e denotemos este número por N .

Um estimador consistente de $f(\underline{x})$ é dado por

$$\hat{f}(\underline{x}) = \frac{N}{n} \cdot \frac{1}{V_{S_{\underline{x}}}}.$$

DEMONSTRAÇÃO. Pela aplicação do método de subpartição de regiões de tolerância, temos que ao fixarmos o número L^q de regiões de tolerância construídas, o valor de m dependerá de n . Suponhamos que estamos nesta situação e que

$$\lim_{n \rightarrow \infty} m = \infty \quad \text{e} \quad \lim_{n \rightarrow \infty} \frac{m}{n} = 0. \quad (1)$$

Se n é grande, podemos dizer que $\underline{x}$ está localizado no interior de uma região de tolerância limitada, com grande probabilidade; nesta demonstração vamos supor que n seja suficientemente grande de modo que esta condição esteja satisfeita.

Seja $S = \{R, \text{ onde } R \text{ é uma região de tolerância e } \underline{x} \text{ é um ponto interior de } R\}$; para cada R pertencente a S , denotemos por V_R e $\|e(R)\|$, o volume de R e o maior lado de R , respectivamente. Sendo f contínua em $\underline{x}$, para cada $\epsilon > 0$, existe um $\delta > 0$, tal que se $\|e(R)\| < \delta$ para $R \in S$, então

$$| [P(R)/V_R] - f(\underline{x}) | < \epsilon. \quad (2)$$

Ao utilizarmos a primeira coordenada de $\underline{x}$ (x_1) como função de corte, obtemos uma região de tolerância $S_{\underline{x}}^1$ que é a união de pelo menos $(n^{q-1} \cdot m)^{1/q}$ blocos estatisticamente equivalentes. A probabilidade de $S_{\underline{x}}^1$ é uma variável aleatória com distribuição beta de parâmetros (aproximadamente) $(n^{q-1} \cdot m)^{1/q}$ e $n - (n^{q-1} \cdot m)^{1/q} + 1$. Como $(n^{q-1} \cdot m)^{1/q}$ satisfaz a condição (1), temos que $P(S_{\underline{x}}^1)$ converge em probabilidade a zero. Mas, sob as suposições em $\underline{x}$, $P(S_{\underline{x}}^1)$ converge a zero em probabilidade, so-

mente se o comprimento de $S_{\underline{x}}^1$ sob x_1 convergir em probabilidade a zero. Com grande probabilidade, o comprimento de $S_{\underline{x}}^1$ sob a primeira coordenada de $\underline{x}$, isto é, o comprimento de x_1 em $S_{\underline{x}}^1$ é menor que δ para n suficientemente grande.

Supondo $q = p$, ao utilizarmos a p -ésima coordenada de $\underline{x}$ (x_p) como função de corte, temos que a probabilidade de $S_{\underline{x}}$, uma união de N blocos estatisticamente equivalentes, converge em probabilidade a zero. Como no caso de $S_{\underline{x}}^1$, isto ocorre somente se o comprimento de x_p em $S_{\underline{x}}$ convergir a zero. Mas, para n suficientemente grande, o comprimento de $S_{\underline{x}}$ sob $x_1, \dots, x_p$ é menor que δ com grande probabilidade, isto é, se $\|e(S_{\underline{x}})\|$ denotar o maior lado de $S_{\underline{x}}$, então

$\|e(S_{\underline{x}})\|$ converge a zero, em probabilidade (3).

Através de (2) e (3) temos que $P(S_{\underline{x}} | V_{S_{\underline{x}}})$ converge, em probabilidade a $f(\underline{x})$ ou

$$\{[nP(S_{\underline{x}})]/N\} / \{[nV_{S_{\underline{x}}}] / N\}$$

converge a $f(\underline{x})$, em probabilidade (4).

Através de uma simples aplicação da desigualdade de Chebyshev, temos que o numerador $[nP(S_{\underline{x}})]/N$ converge em probabilidade a 1 (5).

Então, o denominador $[nV_{S_{\underline{x}}}] / N$ converge em probabilidade a $1/f(\underline{x})$. (6)

Mas (6) é equivalente a mostrar que $\hat{f}(\underline{x})$ é consistente com $f(\underline{x})$.

APÊNDICE C

O teorema que segue foi elaborado e demonstrado por Fix e Hodges (1951).

TEOREMA. Denotemos por X_i , $i=1,2$, a variável aleatória X em π_i , $i=1,2$. Seja f_{iX} , $i=1,2$, a função densidade de probabilidade de $d(X_i, x)$ e suponhamos que f_{iX} seja positiva e contínua no ponto zero.

Seja v um número inteiro e positivo, função dos tamanhos n_1 e n_2 das amostras retiradas respectivamente, de π_1 e π_2 , satisfazendo às seguintes condições:

a) $1 \leq v \leq n_1 + n_2$

b) quando n_i , $i=1,2$, cresce, v cresce, e $\frac{v}{n_i}$, $i=1,2$, tende a zero.

Combinemos as amostras de π_1 e π_2 e coloquemos em ordem crescente as distâncias $d(x_{ij}, x)$, $i=1,2$; $j=1, \dots, n_i$. Definamos

U : v -ésima menor distância até x obtida das amostras de π_1 e π_2 combinadas.

N_i , $i=1,2$,: número de observações amostrais de π_i cujas distâncias até x são menores de iguais a v .

Então

$$\hat{f}_{i\bar{x}}(0) = \frac{N_i \cdot 1}{n_i \bar{U}}$$

é um estimador consistente de $f_{i\bar{x}}$, $i=1,2$, no ponto zero.

DEMONSTRAÇÃO. Seja $\epsilon > 0$ e $\delta > 0$. Definamos v_1 e v_2 através de $v_1 + v_2 = v$ e de

$$\frac{v_1}{v_2} = \frac{n_1 f_{1\bar{x}}(0)}{n_2 f_{2\bar{x}}(0)}.$$

Definamos

$$z = \frac{v_1}{n_1 f_{1\bar{x}}(0) (1+\delta)^2} \quad \text{e} \quad w = \frac{v_1}{n_1 f_{1\bar{x}}(0) (1+\delta)^2}.$$

Observemos que

$$z = \frac{v_2}{n_2 f_{2\bar{x}}(0) (1+\delta)^2} \quad \text{e} \quad w = \frac{v_2}{n_2 f_{2\bar{x}}(0) (1-\delta)^2}.$$

Definamos N_1^Z igual ao número de observações amostrais de π_1 cujas distâncias até $\bar{x}$ são menores que z , N_1^W igual ao número de observações amostrais de π_1 cujas distâncias até $\bar{x}$ são menores que w , N_2^Z igual ao número de observações amostrais de π_2 cujas distâncias até $\bar{x}$ são menores que z e N_2^W igual ao número de observações amostrais de π_2 cujas distâncias até $\bar{x}$ são menores que w .

Usando o fato de que $f_{i\bar{x}}$ é positiva e contínua em zero, encontremos $q > 0$ tão pequeno que quando $0 \leq x \leq q$, tenhamos

$$\left[\frac{f_{i\bar{x}}(x)}{f_{i\bar{x}}(0)} - 1 \right] < \delta, \quad i=1,2.$$

Encontremos n_{11} e n_{21} tais que quando $n_1 > n_{11}$ e $n_2 > n_{21}$, tenhamos $w < q$. Observemos que

$$E(N_1^Z) = n_1 \int_0^z f_{1\bar{x}}(x) dx$$

e então

$$n_1 f_{1\bar{x}}(0) z(1-\delta) < E(N_1^Z) < n_1 f_{1\bar{x}}(0) z(1+\delta).$$

Analogamente, observemos que

$$n_1 f_{1\bar{x}}(0) w(1-\delta) < E(N_1^W) < n_1 f_{1\bar{x}}(0) w(1+\delta)$$

$$n_2 f_{2\bar{x}}(0) z(1-\delta) < E(N_2^Z) < n_2 f_{2\bar{x}}(0) z(1+\delta)$$

$$n_2 f_{2\bar{x}}(0) w(1-\delta) < E(N_2^W) < n_2 f_{2\bar{x}}(0) w(1+\delta).$$

Assim,

$$E(N_1^Z) < \frac{v_1}{1+\delta},$$

$$E(N_1^W) > \frac{v_1}{1-\delta},$$

$$E(N_2^Z) < \frac{v_2}{1+\delta}$$

$$E(N_2^W) > \frac{v_2}{1-\delta}.$$

As variáveis aleatórias envolvidas são binomiais, cujos valores esperados tendem ao infinito, mas mais devagar que o

número de tentativas, quando n_1 e n_2 tendem ao infinito. Entretanto, se encontrarmos n_{12} e n_{22} suficientemente grandes, podemos assegurar que

$$P(N_1^Z < v_1) > 1 - \epsilon,$$

$$P(N_2^Z < v_2) > 1 - \epsilon,$$

$$P(N_1^W > v_1) > 1 - \epsilon$$

e

$$P(N_2^W > v_2) > 1 - \epsilon,$$

quando $n_1 > n_{12}$ e $n_2 > n_{22}$.

Combinando e usando o fato de que U excederá z se $N_1^Z + N_2^Z < v$, temos que

$$P(U > z) > 1 - 2\epsilon$$

e

$$P(U < w) > 1 - 2\epsilon.$$

O evento $U > z$ implica no evento de que todas as observações amostrais de π_1 , cujas distâncias até $\underline{x}$ são menores que z , estão entre as primeiras v observações das amostras combinadas de π_1 e π_2 . Logo, $N_1^Z \leq N_1$. Entretanto,

$$P(N_1^Z \leq N_1) \geq P(U > z) > 1 - 2\epsilon.$$

Analogamente,

$$P(N_1^W \geq N_1) > 1-2\epsilon.$$

Restringindo $n_1 > n_{13}$ e $n_2 > n_{23}$, podemos assegurar

$$P(N_1^Z > n_1 f_{1\bar{x}}(0) z (1-\delta)^2) > 1-\epsilon$$

e

$$P(N_1^W < n_1 f_{1\bar{x}}(0) w (1+\delta)^2) > 1-\epsilon.$$

Combinando, temos:

$$P\left\{\frac{N_1}{n_1} < f_{1\bar{x}}(0) w (1+\delta)^2\right\} > 1-3\epsilon$$

e

$$P\left\{\frac{N_1}{n_1} > f_{1\bar{x}}(0) z (1-\delta)^2\right\} > 1-3\epsilon.$$

Logo,

$$P[f_{1\bar{x}}(0) \frac{z}{w} (1-\delta)^2 < \frac{N_1}{n_1} \frac{1}{U} < f_{1\bar{x}}(0) \bar{z} (1+\delta)^2] > 1-10\epsilon.$$

Desde que

$$\frac{z}{w} = \frac{(1-\delta)^2}{(1+\delta)^2},$$

a conclusão de que $\frac{N_1}{n_1} \frac{1}{U}$ converge em probabilidade a $f_{1\bar{x}}(0)$ é imediata.

Um argumento análogo mostra que $\frac{N_2}{n_2} \frac{1}{U}$ converge em probabilidade a $f_{2\bar{x}}(0)$.

APÊNDICE D

D1. CÁLCULO DA PROBABILIDADE EXATA DE ERRO DE CLASSIFICAÇÃO DO MÉTODO DE DISCRIMINAÇÃO DE PATRICK E FISCHER (1970)

Consideremos o método de discriminação de Patrick e Fischer (1970), apresentado na seção 3.1.2 do Capítulo 3. A regra discriminatória deste método é dada por:

$$\begin{aligned} \text{Se } p_1 \frac{N_1}{n_1} \frac{1}{V_{S_{1x}}} > p_2 \frac{N_2}{n_2} \frac{1}{V_{S_{2x}}}, \text{ classificamos o elemento E em } \pi_1 \\ \text{Se } p_1 \frac{N_1}{n_1} \frac{1}{V_{S_{1x}}} < p_2 \frac{N_2}{n_2} \frac{1}{V_{S_{2x}}}, \text{ classificamos o elemento E em } \pi_2, \end{aligned} \quad (1)$$

onde N_i é o número de observações amostrais de π_i encontradas na região de tolerância S_{ix} , cujo volume denotamos por $V_{S_{ix}}$, $i=1,2$.

Ao particionarmos o espaço amostral S_p , através da amostra de π_i e segundo o método de subpartição de regiões de tolerância, construímos L_i^q regiões de tolerância, cada uma contendo, aproximadamente, m_i observações amostrais de π_i , $i=1,2$.

Logo, por construção $N_i \approx m_i$, $i=1,2$.

Denotemos por:

$P_{PF}(\pi_i)$, $i=1,2$, a probabilidade do elemento E ser

classificado em π_i , segundo o método de Patrick e Fischer, e por

$$P_{PF}(\pi_i | \pi_j), \quad i, j = 1, 2; i \neq j,$$

a probabilidade exata de erro de classificação do método de Patrick e Fischer.

Seja U_i , $i = 1, 2$, a probabilidade de $S_{i\bar{x}}$, $i = 1, 2$. A fim de obtermos $P_{PF}(\pi_i | \pi_j)$, calculemos primeiramente $P_{PF}(\pi_2)$ e $P_{PF}(\pi_1)$, onde

$$P_{PF}(\pi_2) = P \left[p_1 \frac{m_1}{n_1} - \frac{1}{S_{1\bar{x}}} < p_2 \frac{m_2}{n_2} - \frac{1}{S_{2\bar{x}}} \right] \quad (2)$$

e

$$P_{PF}(\pi_1) = 1 - P_{PF}(\pi_2).$$

Para determinarmos esta probabilidade é necessário que conheçamos a distribuição de probabilidade de $V_{S_{1\bar{x}}}$ e $V_{S_{2\bar{x}}}$. Para isto, suporemos que as regiões de tolerância $S_{1\bar{x}}$ e $S_{2\bar{x}}$ sejam suficientemente pequenas de modo que seja razoável estabelecer que $U_i = f_1(\bar{x}) \cdot V_{S_{i\bar{x}}}$, $i = 1, 2$.

Então,

$$P_{PF}(\pi_2) = P \left[\frac{n_1 U_1}{p_1 m_1 f_1(\bar{x})} > \frac{n_2 U_2}{p_2 m_2 f_2(\bar{x})} \right] \quad (3)$$

Sabendo que U_i , $i = 1, 2$, tem distribuição beta, independentemente de $f_i(\bar{x})$, $i = 1, 2$, e definindo

$$T_i = \frac{U_i}{\theta_i} = \frac{n_i}{p_i m_i f_i(\underline{x})} U_i, \quad i=1,2,$$

onde

(4)

$$\theta_i = \frac{p_i m_i f_i(\underline{x})}{n_i}, \quad i=1,2,$$

temos que esta nova variável aleatória tem distribuição beta, com a seguinte função densidade de probabilidade:

$$g_i(u) = \frac{n_i!}{(m_i-1)!(n_i-m_i)!} \theta_i (u\theta_i)^{m_i-1} (1-u\theta_i)^{n_i-m_i}; \quad 0 < u < 1/\theta_i; \quad \theta_i > 0 \quad (5)$$

Logo,

$$P_{PF}(\pi_2) = P[T_1 > T_2] = \begin{cases} \int_0^{1/\theta_1} \left\{ \int_0^s g_1(s) g_2(t) dt \right\} ds & \text{se } \theta_1 > \theta_2 \\ 1 - \int_0^{1/\theta_2} \left\{ \int_0^t g_1(s) g_2(t) ds \right\} dt & \text{se } \theta_1 < \theta_2 \end{cases} \quad (6)$$

As integrais acima, desenvolvidas por Patrick e Fischer (1970), resultam em

$$P_{PF}(\pi_2) = \begin{cases} \frac{n_1! n_2!}{(m_1-1)!(m_2-1)!} \sum_{j=0}^{n_2-m_2} \frac{(-1)^j \left(\frac{\theta_2}{\theta_1}\right)^{m_2+j} (m_2+m_1+j-1)!}{j! (n_2-m_2-j)! (m_2+j) (m_2+n_1+j)!} & \text{se } \theta_1 > \theta_2 \\ 1 - \frac{n_1! n_2!}{(m_1-1)!(m_2-1)!} \sum_{j=0}^{n_1-m_1} \frac{(-1)^j \left(\frac{\theta_1}{\theta_2}\right)^{m_1+j} (m_1+m_2+j-1)!}{j! (n_1-m_1-j)! (m_1+j) (m_1+n_2+j)!} & \text{se } \theta_1 < \theta_2 \end{cases} \quad (7)$$

Seja D_1 , contido em S_p , o conjunto de vetores $\underline{x}$ para os quais temos $\theta_1 > \theta_2$. Analogamente, seja $D_2 = S_p - D_1$, contido em S_p

o conjunto dos valores de $\underline{x}$ para os quais $\theta_1 < \theta_2$.

$$P_{PF}(\pi_i | \pi_j), \quad i, j = 1, 2, i \neq j$$

é calculada da seguinte forma:

$$P_{PF}(\pi_i | \pi_j) = \int_{\theta_i > \theta_j} P_{PF}(\pi_i) \cdot f_j(\underline{x}) d\underline{x} + \int_{\theta_i < \theta_j} P_{PF}(\pi_i) \cdot f_j(\underline{x}) d\underline{x} \quad (8)$$

No caso de termos $n_1 = n_2 = n$ e $m_1 = m_2 = m$, Patrick e Fischer (1970) calcularam a expressão de $P_{PF}(\pi_2)$ para n tendendo ao infinito.

Temos, neste caso:

$$P_{PF}(\pi_2) = \begin{cases} \frac{1}{\left[1 + \frac{p_2 f_2(\underline{x})}{p_1 f_1(\underline{x})}\right]^{2m-1}} \sum_{j=m}^{2m-1} \binom{2m-1}{m} \left(\frac{p_2 f_2(\underline{x})}{p_1 f_1(\underline{x})}\right)^j & \text{se } \theta_1 > \theta_2 \\ 1 - \frac{1}{\left[1 + \frac{p_1 f_1(\underline{x})}{p_2 f_2(\underline{x})}\right]^{2m-1}} \sum_{j=m}^{2m-1} \binom{2m-1}{m} \left(\frac{p_1 f_1(\underline{x})}{p_2 f_2(\underline{x})}\right)^j & \text{se } \theta_1 < \theta_2. \end{cases} \quad (9)$$

Uma vez calculada $P_{PF}(\pi_2)$, procede-se ao cálculo de $P_{PF}(\pi_i | \pi_j)$ através da expressão dada em (8).

D.2. CÁLCULO DA PROBABILIDADE EXATA DE ERRO DE CLASSIFICAÇÃO DO MÉTODO DE DISCRIMINAÇÃO DE COVER E HART (1967)

De acordo com a seção 4.2.2 do Capítulo 4, o método v-VMP de Cover e Hart (1967) é um caso particular do método de discriminação baseado em regiões de tolerância de Patrick e Fischer,

na situação em que

$$\frac{p_1}{n_1} = \frac{p_2}{n_2} \quad e \quad m_1 = m_2 = m = \left[\frac{v+1}{2} \right] \quad (10)$$

Considerando as relações dadas em (10) na dedução da expressão da probabilidade exata de erro de classificação do método de Patrick e Fischer (1970), apresentada na seção D.1 deste Apêndice, obtemos a seguinte expressão para a probabilidade exata de erro de classificação do método de Cover e Hart:

$$P_{CH}(\pi_i | \pi_j) = \int_{f_i(\underline{x}) > f_j(\underline{x})} P_{CH}(\pi_i) f_j(\underline{x}) d\underline{x} + \int_{f_i(\underline{x}) < f_j(\underline{x})} P_{CH}(\pi_i) f_j(\underline{x}) d\underline{x},$$

$i, j = 1, 2; i \neq j,$ (11)

onde:

$$P_{CH}(\pi_2) = \begin{cases} \frac{n_1! n_2!}{((m-1)!)^2} \sum_{j=0}^{n_2-m} \frac{(-1)^j \left[\frac{f_2(\underline{x})}{f_1(\underline{x})} \right]^{m+j} (2m+j-1)!}{j! (n_2-m-j)! (m+j)! (m+n_1+j)!} & \text{se } f_1(\underline{x}) > f_2(\underline{x}) \\ 1 - \frac{n_1! n_2!}{((m-1)!)^2} \sum_{j=0}^{n_1-m} \frac{(-1)^j \left[\frac{f_1(\underline{x})}{f_2(\underline{x})} \right]^{m+j} (2m+j-1)!}{j! (n_1-m-j)! (m+j)! (m+n_2+j)!} & \text{se } f_1(\underline{x}) < f_2(\underline{x}) \end{cases}$$

(12)

e

$$P_{CH}(\pi_1) = 1 - P_{CH}(\pi_2).$$

Quando $n_1 = n_2 = n$, a expressão de $P_{CH}(\pi_2)$ para n tendendo ao infinito é dada por:

$$P_{CH}(\pi_2) = \begin{cases} \frac{1}{\left(1 + \frac{f_2(\underline{x})}{f_1(\underline{x})}\right)^{2m-1}} \sum_{j=m}^{2m-1} \binom{2m-1}{m} \left(\frac{f_2(\underline{x})}{f_1(\underline{x})}\right)^j & \text{se } f_1(\underline{x}) > f_2(\underline{x}) \\ 1 - \frac{1}{\left(1 + \frac{f_1(\underline{x})}{f_2(\underline{x})}\right)^{2m-1}} \sum_{j=m}^{2m-1} \binom{2m-1}{m} \left(\frac{f_1(\underline{x})}{f_2(\underline{x})}\right)^j & \text{se } f_1(\underline{x}) < f_2(\underline{x}). \end{cases} \quad (13)$$

APENDICE E

```

C
C
C
C *****
C  PROGRAMA DE APLICACAO DO METODO MAC PARAMETRICO DE DISCRIMINACAO
C      BASEADO NOS V VIZINHOS MAIS PROXIMOS
C *****
C
C
C $SET AUTOBIND
$BIAD=FROM 818/IMSL/1,818/IMSL/2,819/IMSL/3,818/IMSL/4

COMMON/SIGMA/ SIGMAX(10),SIGMAX(10),XMEDIA(4),YMEDIA(4)
COMMON/AMATE1/X(750,2),Y(750,2),DEST(40,40),NPAC(20),WK(4)
COMMON/COMP/D(4),E(4),ICARA,NVN,NVM,NPCN
COMMON/ELEM/N,M,INM,NPCO,DSEED
DATA SIM/"SIM"/
DOUBLE PRECISION DSEED
*****
LEITURA DOS DADOS
*****
DESCRICAO DOS DADOS
*****
X= POPULACAO 1
Y= POPULACAO 2
ICARA= DIMENSAO DOS VECTORES QUE IDENTIFICAM OS PONTOS AMOSTRAIS
      DAS POPULACOES X E Y
N= NUMERO DE PONTOS AMOSTRAIS DA POPULACAO X PARA CONSTRUCAO DA
  REGRA DISCRIMINANTE
M= MESMA DEFINICAO DE N PORÉM PARA A POPULACAO Y
NPCO= NUMERO DE VALORES DE V
NPAC= VETOR COM OS VALORES DE V
*****
READ(5,1)ICARA,N,M,NPCO

1 FORMAT(4I5)
KIND=1
CALL NORMAL(KIND)
READ(5,4)(NPAC(I),I=1,NPCO)
4 FORMAT(10I5)
*****
IMPRESSAO DOS DADOS LIDOS
*****
WRITE(6,6)N,M

6 FORMAT(///,T40,"IMPRESSAO DOS DADOS",/,T40,"*****"
  -,"///,T25," MATRIZ X = COMPRIMENTO = ",I3,///,T25,
  -" MATRIZ Y = COMPRIMENTO = ",I3,///,T24,"POSICAO",T
  -38," MATRIZ X "T56," MATRIZ Y ")
INAI=N
IF(N.GT.M)INAI=M
IF(ICARA.EQ.3)GO TO 15
IF(ICARA.EQ.4)GO TO 16
WRITE(6,7)(I,X(I,1),X(I,2),Y(I,1),Y(I,2),I=1,INAI)

```

```

7  FORMAT(T25,I3,I35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
  IF(N.EQ.M)GO TO 20
  IF(I+1.NE.N)GO TO 21
  WRITE(6,71)(I,X(I,1),X(I,2),I=M+1,N)
71  FORMAT(T25,I3,I35,F10.5,T45,F10.5)
  GO TO 20
21  WRITE(6,72)(I,Y(I,1),Y(I,2),I=N+1,M)
72  FORMAT(T25,I3,I55,F10.5,T65,F10.5)
  GO TO 20
15  WRITE(6,30)(I,X(I,1),X(I,2),X(I,3),Y(I,1),Y(I,2),Y(I,3),I=1,IMAI)
30  FORMAT(T25,I3,I35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5)
  IF(N.EQ.M)GO TO 20
  IF(IMAI.NE.M)GO TO 32
  WRITE(6,33)(I,X(I,1),X(I,2),X(I,3),I=M+1,N)
33  FORMAT(T25,I3,I35,F10.5,T45,F10.5,T55,F10.5)
  GO TO 20
32  WRITE(6,34)(I,Y(I,1),Y(I,2),Y(I,3),I=N+1,M)
34  FORMAT(T25,I3,I55,F10.5,T65,F10.5,T75,F10.5)
  GO TO 20
16  WRITE(6,60)(I,X(I,1),X(I,2),X(I,3),X(I,4),Y(I,1),Y(I,2),
  Y(I,3),Y(I,4),I=1,IM,I)
60  FORMAT(T25,I3,I35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5,
  T85,F10.5,T95,F10.5,T105,F10.5)
  IF(N.EQ.M)GO TO 20
  IF(IMAI.NE.M)GO TO 42
  WRITE(6,61)(I,X(I,1),X(I,2),X(I,3),X(I,4),I=M+1,N)
61  FORMAT(T25,I3,I35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
  GO TO 20
42  WRITE(6,62)(I,Y(I,1),Y(I,2),Y(I,3),Y(I,4),I=N+1,M)
62  FORMAT(T25,I3,I55,F10.5,T65,F10.5,T75,F10.5,T85,F10.5,T95,F10.5)
20  CONTINUE
  INM=N+M
50  CALL CONT#
51  CONTINUE
100 FORMAT(8F10.0)
  CLOSE(6,DISP=KEEP)
  STOP
  END

```

=====

```

C *****
C
C  SUBROUTINE PARA CALCULO DAS DISTANCIAS EUCLIDIANAS
C *****
C

```

```

SUBROUTINE EUCLID(CALC)
COMMON/COMP/D(4),E(4),ICARA,NVN,NVM,NPCN
P=0.0
DO 1 I=1,ICARA
P=P+(D(I)-E(I))**2.0
1 CONTINUE
CALC=SQRT(P)
RETURN
END

```

```

*****
SUBROTINA PARA ESTIMACAO DAS FUNCOES DENSIDADE DE
PROBABILIDADE, CONSTRUCAO DA REGRA DISCRIMINATORIA,
CLASSIFICACAO E CALCULO DA ESTIMATIVA DA PROBABILIDADE
TOTAL DE ERRO DE CLASSIFICACAO.
*****

```

```

SUBROUTINE CONTA
DIMENSION F(3), IR(1500), A(70,251), E(70,251), C(1500)
DIMENSION PXX(20), PYY(20)
COMMON/ABE/IFIS(20), PON(750,2), SCMX(20), SCMY(20)
COMMON/COMP/DC(4), E(4), ICARA, NVN, NVH, NPN
COMMON/AMATEI/X(750,2), Y(750,2), DIST(40,40), NPAC(20), KK(4)
COMMON/SIGM/SX(10), SY(10), XMEDIA(4), YMEDIA(4)
COMMON/LEMP/H, M, INH, NPCC, DSEED
DOUBLE PRECISION DSEED
DATA F/' ' X', ' Y', 'EMPATE'/

```

```

*****
LEITURA DE DADOS

```

```

P11= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO X
P12= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO Y
NVN= NUMERO DE PONTOS / SEREM CLASSIFICADOS, DA
POPULACAO X
NVH= NUMERO DE PONTOS / SEREM CLASSIFICADOS, DA
POPULACAO Y

```

```

*****

```

```

KK=1
READ(5,1) P11, P12, NVH, NVN
1 FORMAT(2F10.8, 2I5)
KIND=2
CALL NORMAL(KIND)
110 CONTINUE
WRITE(6,65)
IF(ICARA.EQ.3) GO TO 50
IF(ICARA.EQ.4) GO TO 51
WRITE(6,60)(I, FON(I,1), FON(I,2), I=1, NPN)
60 FORMAT(T25, I3, T35, F10.5, T45, F10.5)
GO TO 70
50 WRITE(6,61)(I, FON(I,1), FON(I,2), FON(I,3), I=1, NPN)
61 FORMAT(T25, I3, T35, F10.5, T45, F10.5, T55, F10.5)
GO TO 70
51 WRITE(6,62)(I, FON(I,1), FON(I,2), FON(I,3), FON(I,4), I=1, NPN)
62 FORMAT(T25, I3, T35, F10.5, T45, F10.5, T55, F10.5, T65, F10.5)
65 FORMAT(1H1, T40, "ELEMENTOS EM CLASSIFICACAO", /, T40, 26(" "))
1, /, T24, "ELEMENTO", T40, " COORDENADAS")
70 CONTINUE

```

```

*****
CALCULO E ORDENACAO DAS DISTANCIAS ENTRE O PONTO A SER
CLASSIFICADO E OS VETORES AMOSTRAIS DAS POPULACOES X E Y
*****

```

```

C      WRITL(6,35)
      DO 700 J=1,NPON
      DO 220 I=1,ICARA
220    D(1)=FON(J,I)
      DO 230 II=1,N
      DO 240 JJ=1,ICARA
240    E(JJ)=X(II,JJ)
      CALL EUCLID(RESUL)
230    C(II)=RESUL
      DO 250 III=1,M
      DO 260 JJJ=1,ICARA
260    L(JJJ)=Y(III,JJJ)
      CALL EUCLID(RESUL)
      IIP=III+N
      C(IIP)=RESUL
250    CONTINUE
C      WRITE(6,90)
      LA=N+M
      DO 730 I=1,LA
730    IR(I)=I
      CALL VSFIF(C,LA,IF)
      11 CONTINUE
      DO 14 I3=1,NPCC
      IX=0
      IY=0
      NKA=NPAC(I3)
      DO 15 I4=1,NKA
      IF(IR(I4)-N)21,21,22
      21 IX=IX+1
      GO TO 15
      22 IY=IY+1
      15 CONTINUE
C
C      *****
C      FSTIMACAO DAS FUNCOES DENSIDADE DE PROBABILIDADE
C      *****
C
      XX=IX*PI1/N
      YY=IY*PI2/M
      IF(XX=YY)23,2,25
23    KIMPRE=2
      LPIS=IPIS(I3)
      IPIS(I3)=LPIS+1
      SOMAY(I3)=SOMAY(I3)+1
      GO TO 26
25    KIMPRE=1
      SOMAX(I3)=SOMAX(I3)+1
      GO TO 26
24    CONTINUE
C 24 WRITL(6,20)F(3),NKA,J,F(1),IX,IY,XX,YY
C 20 FORMAT(3,A6,T12,I3,T27,I3,T47,A6,T68,I3,T85,I3,T95,F8.5,
C      -T110,F8.5)
      GO TO 25
26    CONTINUE
C 26 WRITL(6,30)NKA,J,F(KIMPRE),IX,IY,XX,YY
C 30 FORMAT(3,I3,T16,I3,T32,A6,T52,I3,T65,I3,T87,F8.6,T114,F8.6)
C 35 FORMAT(1H1,T40,"CLASSIFICACAO NAO PARAMETRICA"/,
C      -T40,"*****"/,
C      -T10,"K",T14,"ELEMENTO",
C      -T26,"CLASSIFICACAO NA PGF",T49,"FRQ EM X",T62,

```

```

C      -"FREQ EM Y",T81,"DENSID/DE ESTIMADA EM X",T107,
C      -"DENSIDADE ESTIMADA EM Y")
C 900 FORMAT(/)
      14 CONTINUE
      10 CONTINUE
      700 CONTINUE
      GO TO 920
      940 KK=2
      KIND=3
      CALL NORMAL(KIND)
      GO TO 110
      920 CONTINUE
      WRITE(6,925)F(KK),F(KK)
      DO 930 I=1,NPCG
      ISUMA=SOMAX(I)
      SMX=ISUMA
      JSOMA=SOMAX(I)
      SMY=JSOMA
      IF(KK.EQ.1)GO TO 500
      PYY(I)=SMY/NVM
      WRITE(6,935)NPAC(I),ISOMA,JSOMA,PYY(I)
      GO TO 930
      500 PXX(I)=SMY/NVM
      WRITE(6,935)NPAC(I),ISOMA,JSOMA,PXX(I)
      930 CONTINUE
      DO 1001 II=1,NFCG
      SOMAX(II)=0
      1001 SOMAX(II)=0
      IF(KK.EQ.1)GO TO 940
      WRITE(6,910)
      510 FORMAT(1H,10X,"PROBABILIDADE TOTAL DE ERRO DE CLASSIFICACAO",
      1/,10X,30(" "),///,10X,"VALOR K",5X,"PROBABILIDADE")
      DO 515 I=1,NPCG
      PTOTAL=PXX(I)*FI1+PYY(I)*FI2
      WRITE(6,520)I,PTOTAL
      520 FORMAT(10X,13,5X,F7.5)
      515 CONTINUE
      RETURN
      925 FORMAT(1H,T35,"FREQUENCIA DE CLASSIFICACAO PARA PONTOS DA POP",
      *46," E PROB. DE ERRO",/,T35,52(" ")/,
      *3X,"VALOR DE K",5X,"POPULACAO X",11X,"POPULACAO Y",7X,"PROB. DE",
      *" ERRO PARA A POPULACAO",46,/)
      935 FORMAT(6X,13,12X,15,15X,15,15X,F7.4)
      END

```

```

C      >
C      *****
C      GERADOR DE VETORES DA DISTRIBUICAO NORMAL MULTIVARIADA
C      *****
C

```

```

SUBROUTINE GERAN(NR,K,SIGMA,RVEC,KKVEC,CSEED)
INTEGER IF,IER
REAL SIGMA(10),RVEC(750,2),KKVEC(4)
DOUBLE PRECISION CSEED
IR=750
CALL GGNSP(CSEED,NR,K,SIGMA,IR,RVEC,KKVEC,IER)

```

RETURN
END

=====

```
C
C *****
C SUBROTINA QUE CERA AS MOSTRAS PARA CONSTRUCAO DA REGRA
C DISCRIMINATORIA E OS PONTOS A SEREM CLASSIFICADOS
C POPULACOES X E Y
C *****
C
```

```

SUBROUTINE NORMAL(KIND)
COMMON/SIGMA/ SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
COMMON/AM/IRI/ X(750,2),Y(750,2),DIST(40,40),NPAC(20),WK(4)
COMMON/COMP/ D(4),E(4),ICARA,NVN,NVM,NPON
COMMON/ELEM/N,M,INM,NPCC,DSEED
COMMON/AS/IPIS(20),PON(750,2),SGMAX(20),SGMAY(20)
DIMENSION SMAX(10),SMAY(10)
DOUBLE PRECISION DSEED
IF(KIND.EQ.1)GO TO 1
IF(KIND.EQ.2)GO TO 2
WK(1)=0.0
CALL GERANC(NVM,ICARA,SMAY,PCN,WK,DSEED)
NPON=NVM
DO 255 JJ=1,ICARA
DO 270 II=1,NPON
270 PCN(II,JJ)=PON(II,JJ)+YMEDIA(JJ)
255 CONTINUE
GO TO 3
1 READ(5,1210)DSEED
1210 FORMAT(D10.0)

```

```
C
C *****
C LEITURA DE DADOS
C
C XMEDIA= VETOR DE MEDIAS PARA A DISTRIBUICAO NORMAL
C          ASSOCIADA A POPULACAO X
C YMEDIA= IDEM PARA A POPULACAO Y
C SIGMAX= MATRIZ DE COVARIANCIA PARA A DISTRIBUICAO
C          NORMAL ASSOCIADA A POPULACAO X
C SIGMAY= IDEM PARA A POPULACAO Y
C DSEED= SEMENTE PARA O GERADOR DE VETORES DA
C        DISTRIBUICAO NORMAL MULTIVARIADA
C *****
C
```

```

READ(5,100)(SIGMAX(I),I=1,10)
100 FORMAT(BF10.0)
READ(5,100)(SIGMAY(I),I=1,10)
DO 10 IEA=1,10
SMAX(IEA)=SIGMAX(IEA)
SMAY(IEA)=SIGMAY(IEA)
10 WRITE(6,/)SMAX(IEA),SMAY(IEA)
WK(1)=0.0
CALL GERANC(N,ICARA,SIGMAX,X,WK,DSEED)
WK(1)=0.0
CALL GERANC(N,ICARA,SIGMAY,Y,WK,DSEED)
READ(5,100)(XMEDIA(I),I=1,ICARA), (YMEDIA(J),J=1,ICARA)

```



```
DO 130 JJ=1,ICARA
DO 140 II=1,N
140 X(II,JJ)=X(II,JJ)+XMEDIA(JJ)
DO 150 KK=1,M
150 Y(KK,JJ)=Y(KK,JJ)+YMEDIA(JJ)
130 CONTINUE
GO TO 3
2 WK(1)=0.0
CALL GERAN(NVN,ICARA,SPFX,PCN,WK,CSEED)
NPN=NPN
DO 5 JJJ=1,ICARA
DO 6 II=1,NPN
6 PDI(II,JJJ)=PDI(II,JJJ)+XMEDIA(JJ)
5 CONTINUE
3 CONTINUE
RETURN
END
```

=====

```

C
C*****
C  PROGRAMA DE APLICACAO DE METODO CLINCO DE DISCRIMINACAO QUADRATICO
C*****
C
C 1SET AUTOBINC
C1E1E=FROM B1B/1MSL/1,E1E/1MSL/2,B1B/1MSL/3,E1E/1MSL/4

COMMON/SIGM// SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
COMMON/SIG/SIGX(4,4),SIGY(4,4)

WARNING:"SIGX" COPY EXPECTED

COMMON/AMATEI/XI7E0,2),Y(7E0,2),DIST(40,40),NPAC(20),NK(4)
COMMON/COMP/DC(4),L(4),ICAR,NVA,NV,NPCO
COMMON/ELEM/I,P,INH,NPCO,DSEED
DIMENSION TEMFX(4),TEMFY(4),XM(4),YM(4),VCVX(4,4),VCVY(4,4),FC(3)
DIMENSION B(4),BFC(4),FCN1(4),FCN2(4),WKAREA(10),VX(10),VY(10)
DIMENSION VVX(4,4),XMA(4),TEMPAC(4)
DATA SIN/"SIN"/
DATA F/"X","Y","TEMPATE"/
DOUBLE PRECISION DSEED
*****
C
C  LEITURA DOS DADOS
C*****
C  DESCRICAO DOS DADOS
C*****
C
C X= POPULACAO 1
C Y= POPULACAO 2
C ICAR= DIMENSAO DOS VETORES QUE IDENTIFICAM OS PONTOS AMOSTRAIS
C      DAS POPULACOES X E Y
C N= NUMERO DE PONTOS AMOSTRAIS DA POPULACAO X PARA CONSTRUCAO DA
C    REGRA DISCRIMINANTE
C M= MESMA DEFINICAO DE N PORÉM PARA A POPULACAO Y
C NPCO= NUMERO DE VALORES DE V
C NPAC= VETOR COM OS VALORES DE V
C P1= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO X
C P2= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO Y
C NVA= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
C      POPULACAO X
C NVN= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
C      POPULACAO Y
C*****
C READ(5,1) ICAR,N,P,NPCO

WARNING:REDUNDANT COMMAS WILL BE IGNORED

1 FORMAT(4I5)
KIND=1
CALL NORMAL(KIND)
READ(5,4)(NFAC(I),I=1,NPCO)

```

```

4  FORMAT(10I5)
   IMM=N*M
   READ(5,10)PI1,PI2,NVN,NVH
10  FORMAT(2F10.8,2I5)
   VAL=ALOG(PI2/PI1)
   IX=750
   NBR(1)=ICARA
   NBR(2)=N
   NBR(3)=N
   NBR(4)=1
   NBR(5)=C
   NBR(6)=0
C*****
C      CONSTRUÇÃO DA FLORA DISCRIMINATÓRIA
C*****
C      CALL BECOVM(X,IX,NBR,TEMPX,XM,V),IER)
   NOR=ICARA
   NBE=4
C      CALL VCVT SF(VX,NOR,VCX,NBE)
   NBR(2)=M
   NBR(3)=M
C      CALL BECOVM(Y,IX,NBR,TEMPY,YM,VY,IER)
C      CALL VCVT SF(VY,NOR,VCY,NBE)
C      WRITE(6,400)((TEMPX(I),TEMPY(I)),X(I),Y(I),I=1,2)
C 400  FORMAT(4F10.3)
C      WRITE(6,410)((VCX(I),VCY(I)),I=1,2),J=1,2)
C 410  FORMAT(1X,"MATRIZ DE VARIANCIA E COVARIANCIA",/,2F20.8)
   D1=0.0
C      CALL LINV3F(CTGX,B,1,ICARA,4,D1,D2,WKAREA,IER)
C      WRITE(6,410)((VCY(I),J),I=1,2),J=1,2)
   DETY=D1*(2.0**D2)
   WRITE(6,420)D1,D2,DETY
C
C 420  FORMAT(1X,"COMPONENTES",3F20.8)
   D1=0.0
C      CALL LINV3F(CTGX,B,1,ICARA,4,D1,D2,WKAREA,IER)
C      WRITE(6,410)((VCX(I),J),I=1,2),J=1,2)
   DETX=D1*(2.0**D2)
   WRITE(6,420)D1,D2,DETX
   AAA=ALOG(CASE(DETY)/ABS(DETX))
   WRITE(6,/)AAA
   KIND=2
   CALL NORMAL(KIND)
   IX=750
   NBR(1)=ICARA
   NBR(2)=NVN
   NBR(3)=NVN
   NBR(4)=1
   NBR(5)=C
   NBR(6)=C
   CALL BECOVM(X,IX,NBR,TEMPX,XM,VX,IER)
   NOR=ICARA
   NBE=4
   CALL VCVT SF(VX,NOR,VVX,NBE)
   WRITE(6,450)((TEMPX(I),X(I),I=1,2)
C 450  FORMAT(2F10.8)
   WRITE(6,410)((VVX(I),J),I=1,2),J=1,2)
C 460  FORMAT(1X,"MATRIZ DE VARIANCIA E COVARIANCIA",/,2F20.8)
   KK=1
C 270  WRITE(6,300)

```

```

300 FORMAT(1H1,/,/,T40,"CLASSIFICACAO QUADRATICA",/,T40,24(" "),/,
1T14,"ELEMENTO",T26,"E",POP,T42,"COORDENADAS",
2T57,"CLASSIFICADO NA POP",T61,"DENSIDADE ESTIMADA")
DO 200 J=1,NVN
DO 210 J=1,ICARA
PCN1(J)=X(1,J)-XMEOT1(J)
210 PCN2(J)=X(1,J)-YMEOT1(J)
DO 220 J1=1,ICARA
DO 230 J2=1,ICARA
666=666+PCN2(J2)*SIGY(J2,J1)
230 666=666+PCN1(J2)*SIGX(J2,J1)
CCC=CCC+666*PCN2(J1)
CCC=CCC+666*PCN1(J1)
EEE=EEE+DDD*PCN1(J1)
DDD=0.0
EEE=0.0
220 CONTINUE
TCS=0.5*(AAA+CCC-EEE)
CCC=0.0
EEE=0.0
C*****
C      CALCULO DA ESTIMATIVA DE PEGABEILIDADE TOTAL DE ERRO
C      DE CLASSIFICACAO
C*****
IF(TLS.EC.VAL)GO TO 250
IMPRE=2
ISOMAY=ISOMAX+1
GO TO 260
250 IMPRE=1
ISOMAX=ISOMAX+1
250 WRITE(6,310)(I,FKK),X(1,1),X(1,2),F(1,PRE),TES
310 FORMAT(16,13,124,A6,5X,F10.5,2),F10.5,T62,16,T84,F12.5)
200 CONTINUE
IF(FKK.EC.2)ISOMAY=ISOMAX
SOMAY=ISOMAY
SOMA=SOMAY/NVN
ISOMAY=NVN-ISOMAX
WRITE(6,320)(FKK),NVN,ISOMAX,ISOMAY,SOMA
320 FORMAT(1H1,T40,"CLASSIFICACAO PARA PONTOS DA POP:",A6,
1" E PROC. DE ERRO",/,T40,56(" "),T10,"TOTAL DE PONTOS",
2T30,"CLASSIF. NA POP. X",T50,"CLASSIF. NA POP. Y",T80,
3"PROC. DE ERRO",/,T15,13,T38,13,T58,13,T82,F10.6)
IF(FKK.EC.2)GO TO 230
KIND=3
CALL NCSMAL(KIND)
NBR(1)=ICARA
NBR(2)=NVN
NBR(3)=NVN
NBR(4)=1
NBR(5)=0
NBR(6)=0
CALL BLCOVM(X,10,NBR,TEMPO,MA,VX,IER)
NCR=ICARA
NBR=4
CALL VCNTSF(VX,NBR,VVVX,NBR)
WRITE(6,470)(TEMPX(I),XM(I),I=1,2)
470 FORMAT(2F10.8)
WRITE(6,480)((VVVX(I,J)),I=1,2,J=1,2)
480 FORMAT(1X,"MATRIZ DE VARIANCIA E COVARIANCIA",/,F20.8)
PTOT=SOMA*PI1
NVNP=NVN

```

```

NVA=NVW
KK=2
ISGMAY=C
ISGPAY=C
CO TO 270
280 NVA=NVW
  FTCT=FTCT+PI2*SCM1
  WRITE(6,330)FTCT
330 FORPAT(1:1,EX,"FRESCABILIDADE TOTAL DE ERRO DE CLASSIFICACAO IGUAL"
  ,FIC.6)
  CLOSE(6,DISTF=KEEP)
  STOP
  END

```

```

=====
C
C *****
C GERADOR DE VECTORES DA DISTRIBUICAO NORMAL MULTIVARIADA
C *****
SUBROUTINE GERANCON(K,SIGMA,RVEC,KKVEC,DSEED)
  INTEGER IR,IER
  REAL SIGMA(10),RVEC(750,2),KKVEC(4)
  DOUBLE PRECISION DSEED
  IF=750
  CALL CONSMOTSEED,KK,K,SIGMA,IR,RVEC,KKVEC,IER)
  RETURN
  END

```

```

=====
C
C *****
C SUBROTINA QUE LERA AS AMOSTRAS PARA CONSTRUCAO DA REGRA
C DISCRIMINATORIA E OS FONTOS A SEREM CLASSIFICADOS
C POPULACOES X E Y
C *****

```

```

SUBROUTINE NORMAL(KIND)
  COMMON/SIG// SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
  COMMON/SIG/SIGX(4,4),SIGY(4,4)

```

```

KAFNAG="SIGX" COMMON EXPECTED

```

```

COMMON/AMATEI/X(750,2),Y(750,2),DIST(10,40),NPAC(20),KK(4)
COMMON/COMPFB/D(4),L(4),ICAF,NV,NV,NPON
COMMON/AJUC/NPGPA(40),NPGSI(40)
COMMON/AB/IFIS(20),PCN(400,4),SCHAX(20),SCHAY(20)
DIMENSION SMAX(10),SMAY(10)
DOUBLE PRECISION DSEED
IF(KIND.EQ.1)CO TO 1
IF(KIND.EQ.2)CO TO 2
KK(1)=0.0
CALL GERANCONVM,ICARA,SMAY,X,KK,DSEED)
APGN=NVW
CO 255 JJ=1,ICAF
CO 270 II=1,APON

```

```
270 X(I,JJ)=X(I,JJ)+YMEDIA(JJ)
255 CONTINUE
CO TO 3
```

```
C
C*****
C  LEITURA DE DADOS
C
C  XMEDIA= VETOR DE MEDIAS PARA A DISTRIBUICAO NORMAL
C           ASSOCIADA A POPULACAO X
C  YMEDIA= IDEN PARA A POPULACAO Y
C  SIGMAX= MATRIZ DE COVARIANCIA PARA A DISTRIBUICAO
C           NORMAL ASSOCIADA A POPULACAO X
C  SIGMAY= IDEN PARA A POPULACAO Y
C  CSEED= SEMENTE PARA O GERADOR DE VETORES DA DISTRIBUICAO
C         NORMAL MULTIVARIADA
C*****
1  READ(5,1210)CSEED
1210 FORMAT(110.0)
120  READ(5,100)(SIGMAX(I),I=1,10)
100  FORMAT(8F10.0)
100  READ(5,100)(SIGMAY(I),I=1,10)
NCR=ICARA
NBB=4
CALL VCNTSF(SIGMAX,NCR,SIGX,NBB)
CALL VCNTSF(SIGMAY,NCR,SIGY,NBB)
DO 10 IEA=1,10
SMAX(IEA)=SIGMAX(IEA)
SMAY(IEA)=SIGMAY(IEA)
10  WRITE(6,/)SMAX(IEA),SMAY(IEA)
100  READ(5,100)(XMEDIA(I),I=1,ICARA),(YMEDIA(J),J=1,ICARA)
CO TO 3
2  WK(1)=0.0
CALL CEFANC(NVN,ICARA,SMAX,X,WK,CSEED)
NFCN=NVN
DO 5 JJ=1,ICARA
CO 5 II=1,NFCN
6  X(II,JJ)=X(II,JJ)+XMEDIA(JJ)
5  CONTINUE
3  CONTINUE
RETURN
END
```

=====

```

C
C
C *****
C  PROGRAMA DE APLICACAO DO METODO DISCRIMINANTE LINEAR AMOSTRAL
C *****
C
C 1SET AUTOSTND
C 2EINC=FROM 019/INSL/1,01E/INSL/2,01B/INSL/3,01B/INSL/4

COMMON/SIGMA/ SIGMA-X(10),SIGMA-Y(10),XMEDIA(4),YMEDIA(4)
COMMON/ANATEL/XI750(2),Y(750(2),DIST(40,40),NPAC(20),WK(4)
COMMON/COMP/PC(4),E(4),ICAR!,NVX,NVY,NPON
COMMON/ELEM/N,M,IMM,NPCD,DSEED
DIMENSION TEMPX(4),TEMPY(4),XM(4),YM(4),VCVX(4,4),VCVY(4,4),FC(3)
DIMENSION B(4),NBF(6),FCN1(4),PCN2(4),WKARZ(10),VXC(10),VY(10)
DIMENSION DECA(4)
DATA SIM/"SIM"/
DATA F/"X","Y","TEMPATE"/
DOUBLE PRECISION DSEED
*****
C  LEITURA DOS DADOS
C *****
C  READ(5,*) ICARA,N,M,NCLASS,NPCD

1  FORMAT(5I5)
   KIND=1
   CALL NORMAL(KIND)
   READ(5,4)(NPAC(I),I=1,NPCD)
4  FORMAT(10I5)
   *****
   C  IMPRESSAO DOS DADOS LIDOS
   C *****
   C  DESCRICAO DOS DADOS
   C *****
   C
   C  X= POPULACAO 1
   C  Y= POPULACAO 2
   C  ICARA= DIMENSAO DOS VETORES QUE IDENTIFICAM OS PONTOS AMOSTRAIS
   C          DAS POPULACOES X E Y
   C  N= NUMERO DE PONTOS AMOSTRAIS DA POPULACAO X PARA CONSTRUCAO DA
   C      REGRA DISCRIMINANTE
   C  M= MESMA DEFINICAO DE N PORÉM PARA A POPULACAO Y
   C  NPCD= NUMERO DE VALORES DE V
   C  NPAC= VETOR COM OS VALORES DE V
   C *****
   C  WRITE(6,6)N,M
   C
6  FORMAT(///,T40,"IMPRESSAO DOS DADOS",/,T40,"*****"
   -"//////,T25," MATRIZ X - COMPLEMENTO = ",I3,///,T25,
   -" MATRIZ Y - COMPLEMENTO = ",I3,////////,T24,"POSICAO",T
   -30," MATRIZ X ",T56," MATRIZ Y ")
   IMAT=N

```

```

IF(N.CI.M)I=1
IF(ICARA.EQ.3)GO TO 15
IF(ICARA.EQ.4)GO TO 18
WRITE(6,7)(I,X(I,1),X(I,2),Y(I,1),Y(I,2),I=1,IM41)
7 FORMAT(T25,13,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
IF(N.EQ.M)GO TO 20
IF(IM41.NE.M)GO TO 21
WRITE(6,71)(I,X(I,1),X(I,2),I=M+1,N)
71 FORMAT(T25,13,T35,F10.5,T45,F10.5)
GO TO 20
21 WRITE(6,72)(I,Y(I,1),Y(I,2),I=N+1,M)
72 FORMAT(T25,13,T55,F10.5,T65,F10.5)
GO TO 20
15 WRITE(6,30)(I,X(I,1),X(I,2),X(I,3),Y(I,1),Y(I,2),Y(I,3),I=1,IM41)
30 FORMAT(T25,13,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5,
1,T85,F10.5)
IF(N.EQ.M)GO TO 20
IF(IM41.NE.M)GO TO 32
WRITE(6,33)(I,X(I,1),X(I,2),X(I,3),I=M+1,N)
33 FORMAT(T25,13,T35,F10.5,T45,F10.5,T55,F10.5)
GO TO 20
32 WRITE(6,34)(I,Y(I,1),Y(I,2),Y(I,3),I=N+1,M)
34 FORMAT(T25,13,T35,F10.5,T55,F10.5,T75,F10.5)
GO TO 20
16 WRITE(6,60)(I,X(I,1),X(I,2),X(I,3),X(I,4),Y(I,1),Y(I,2),
1Y(I,3),Y(I,4),I=1,IM41)
60 FORMAT(T25,13,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5,
1,T85,F10.5,T95,F10.5,T105,F10.5)
IF(N.EQ.M)GO TO 20
IF(IM41.NE.M)GO TO 42
WRITE(6,61)(I,X(I,1),X(I,2),X(I,3),X(I,4),I=M+1,N)
61 FORMAT(T25,13,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
GO TO 20
42 WRITE(6,62)(I,Y(I,1),Y(I,2),Y(I,3),Y(I,4),I=N+1,M)
62 FORMAT(T25,13,T55,F10.5,T65,F10.5,T75,F10.5,T85,F10.5,T95,F10.5)
20 CONTINUE
INT=A+M

```

```

*****
LEITURA DE DADOS

```

```

PI1= PROBABILIDADE A FRIGIR PARA PONTOS DA POPULACAO X
PI2= PROBABILIDADE A FRIGIR PARA PONTOS DA POPULACAO Y
NVX= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
POPULACAO X
NVY= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
POPULACAO Y

```

```

*****

```

```

10 READ(5,10)PI1,PI2,NVX,NVY
FORMAT(2F10.5,2I5)
VAL=ALOG(PI2/PI1)
WRITE(6,/)PI1,PI2,NVX,NVY,VAL
IX=750
NBR(1)= ICARA
NBR(2)=N
NBR(3)=N
NBR(4)=1
NBR(5)=6
NBR(6)=0

```



```

C*****
C   CONSTRUCAO DA REGRA DISCRIMINATORIA
C*****
CALL BECOVM(X,IX,NBR,TEMPX,XH,VX,IER)
NBR=ICARA
NBR=4
CALL VCNTSF(VX,NBR,VCVX,NBR)
NBR(2)=N
NBR(3)=N
CALL BECOVM(Y,IX,NBR,TEMPY,YH,VY,IER)
CALL VCNTSF(VY,NBR,VCVY,NBR)
WRITE(6,400)
400 FORMAT(1H1,T30,"IMPRESSAO DE RESULTADOS PARCIAIS")
WRITE(6,401)
401 FORMAT(///,T20,"COMPONENTES",T35,"MEDIA X",T50,"MEDIA Y")
WRITE(6,402)(I,XH(I),YH(I),I=1,ICARA)
402 FORMAT(/,T22,I2,T35,F10.7,T50,F10.7)
WRITE(6,403)
403 FORMAT(///,T20,"COMPONENTE",T35,"MATRIZ COV VETOR X",T70,
1"MATRIZ COV VETOR Y")
WRITE(6,404)(I,VCVX(I,1),VCVX(I,2),VCVY(I,1),VCVY(I,2),I=1,2)
404 FORMAT(/,T22,I2,T35,F10.7,T47,F10.7,T70,F10.7,T83,F10.7)
DO 408 I=1,2
DO 409 J=1,2
CALCS=VCVX(I,J)*VCVY(J,J)
CALCS=2/CALCS/2.0
VCVY(I,J)=CALCS
409 CONTINUE
408 CONTINUE
WRITE(6,405)
405 FORMAT(///,T20,"COMPONENTES",T35,"MATRIZ S - COMPOSICAO DAS COV")
WRITE(6,406)(I,VCVY(I,1),VCVY(I,2),I=1,2)
406 FORMAT(/,T22,I2,T35,F10.7,T47,F10.7)
CALL LINV3F(VCVY,N,1,ICARA,4,D1,D2,WKAREA,IER)
WRITE(6,407)
407 FORMAT(///,T20,"COMPONENTE",T35,"MATRIZ INVERSA DE S")
WRITE(6,408)(I,VCVY(I,1),VCVY(I,2),I=1,2)
DO 410 J=1,ICARA
PON1(J)=(XH(J)+YH(J))/2.0
PON2(J)=XH(J)-YH(J)
410 CONTINUE
KIND=2
CALL NORMAL(KIND)
KK=1
270 WRITE(6,300)
300 FORMAT(1H1,///,T40,"CLASSIFICACAO QUADRATICA",/,T40,24(" "),/,
1T14,"ELEMENTO",T26,"DA POP",T42,"COORDENADAS",
2T57,"CLASSIFICAO NA POP",T81,"DENSIDADE ESTIMADA")
DO 200 I=1,NVN
DO 220 J1=1,ICARA
DO 230 J2=1,ICARA
EPE=X(I,J2)+PON1(J2)
CCC=CCC+BBB*VCVY(J2,J1)
230 CONTINUE
DECA(J1)=CCC
CCC=0.0
220 CONTINUE
DO 240 J3=1,ICARA
TES=TES+DEC/(J3)*PON2(J3)
240 CONTINUE

```

```

C*****
C  CALCULO DA ESTIMATIVA DA PROBABILIDADE TOTAL
C  DE ERRO DE CLASSIFICACAO
C*****
      IF (TES.CC.VAL) GO TO 250
      IMPRE=2
      ISOMAY= ISOMAX*1
      GO TO 260
250  IMPRE=1
      ISOMAX= ISOMAX*1
260  WRITE(6,310) F(KK),X(I,1),X(I,2),F(IMPRE),YES
310  FORMAT(16,13,124,A5,5X,F10.5,2),F10.5,T62,A6,T84,F12.5)
      TES=0.0
200  CONTINUE
      IF (KK.EC.2) ISOMAY=ISOMAX
      SOMAY=ISOMAY
      SOMA=SOMAY/NVN
      ISOMAY=NVN-ISOMAX
      WRITE(6,320) F(KK),NVN,ISOMAX,ISOMAY,SOMA
320  FORMAT(1H1,T40,"CLASSIFICACAO PARA PONTOS DA POP. ",A6,
1" E PROB. DE ERRO",/,T40,56(" "),/,T10,"TOTAL DE PONTOS",
2T30,"CLASSIF. NA POP. X",T50,"CLASSIF. NA POP. Y",T80,
3"PROB. DE ERRO",/,T10,13,T35,13,T55,13,T82,F10.6)
      IF (KK.EC.2) GO TO 230
      KINO=3
      CALL NORMAL(KINO)
      PTOT=SOMA*PI1
      NVNP=NVN
      NVN=NVN-1
      KK=2
      ISOMAX=0
      ISOMAY=0
      GO TO 270
280  NVN=NVNP
      PTOT=PTOT+PI2*SOMA
      WRITE(6,327) PTOT
330  FORMAT(1H1,2X,"PROBABILIDADE TOTAL DE ERRO DE CLASSIFICACAO IGUAL"
1",F10.5)
      CLOSE(6,DISP=KEEP)
      STOP
      END

```

```

=====
C
C *****
C  GERADOR DE VECTORES DE DISTRIBUICAO NORMAL MULTIVARIADA
C  *****
C

```

```

SUBROUTINE GERAN(NR,K,SIGMA,RVEC,KKVEC,DSEED)
INTEGER IR,ILR
REAL SIGMA(10),FVEC(750,2),KKVEC(4)
DOUBLE PRECISION DSEED
IR=750
CALL CGNRM(DSEED,NR,K,SIGMA,IR,FVEC,KKVEC,IER)
RETURN
END

```

```

=====
C
C *****
C SUBROTINA QUE LEE AS AMOSTRAS PARA CONSTRUÇÃO DA REGRA
C DISCRIMINATORIA E OS PONTOS A SEREM CLASSIFICADOS
C POPULAÇÕES X E Y
C *****
C
SUBROUTINE NORMAL(KIND)
COMMON/SIGMA/ SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
COMMON/AMATEI/X(750,2),Y(750,2),DIST(40,40),NPAC(20),WK(4)
COMMON/COMPR/D(4),IC(4),ICARA,NVM,NVY,NPON
COMMON/PLFM/N,M,INP,NFCD,DSEED
COMMON/AB/IFIS(20),FON(400,4),SCMAX(20),SCMAY(20)
DIMENSION SMAX(10),SMAY(10)
DOUBLE PRECISION DSEED
IF(KIND).EQ.1 GO TO 1
IF(KIND).EQ.2 GO TO 2
WK(1)=0.0
CALL CLERAN(NVM,ICARA,SMAY,Y,WK,DSEED)
NPOS=NVM
DO 255 JJ=1,ICARA
DO 270 II=1,NPON
270 X(II,JJ)=X(II,JJ)+YMEDIA(JJ)
255 CONTINUE
GO TO 3
C
C *****
C LEITURA DE DADOS
C
C XMEDIA= VETOR DE MEDIAS PARA A DISTRIBUIÇÃO NORMAL
C ASSOCIADA A POPULAÇÃO X
C YMEDIA= IDEM PARA A POPULAÇÃO Y
C SIGMAX= MATRIZ DE COVARIÂNCIA PARA A DISTRIBUIÇÃO
C NORMAL ASSOCIADA A POPULAÇÃO X
C SIGMAY= IDEM PARA A POPULAÇÃO Y
C DSEED= SEMENTE PARA O GERADOR DE VETORES DA
C DISTRIBUIÇÃO NORMAL MULTIVARIADA
C *****
C
1 READ(5,1010)DSEED
1210 FORMAT(10.0)
READ(5,100)(SIGMAX(I),I=1,10)
100 FORMAT(8F10.0)
READ(5,100)(SIGMAY(I),I=1,10)
DO 10 IEA=1,10
SMAX(IEA)=SIGMAX(IEA)
SMAY(IEA)=SIGMAY(IEA)
10 WRITE(6,/)SMAX(IEA),SMAY(IEA)
WK(1)=0.0
CALL CLERAN(N,ICARA,SIGMAX,X,WK,DSEED)
WK(1)=0.0
CALL CLERAN(N,ICARA,SIGMAY,Y,WK,DSEED)
READ(5,100)(XMEDIA(I),I=1,ICARA),YMEDIA(J),J=1,ICARA)
DO 130 JJ=1,ICARA
DO 140 II=1,N
140 X(II,JJ)=X(II,JJ)+XMEDIA(JJ)

```

```
DO 150 KK=1,M
150 Y(KK,JJ)=Y(KK,JJ)+YMEDIA(CJJ)
130 CONTINUE
CO TO 3
2 WK(1)=0.0
CALL GEFAN(NVN,ICRA,SMAX,X,WK,DSEED)
NPDN=NVN
DO 5 JJ=1,ICRA
DO 5 II=1,NPDN
6 X(II,JJ)=X(II,JJ)+XMEDIA(CJJ)
5 CONTINUE
3 CONTINUE
RETURN
END
```

=====

```

C
C
C
C *****
C  PROGRAMA DE APLICACAO DO METODO DISCRIMINANTE QUADRATICO AMOSTRAL
C *****
C
C
C  SET AUTOBINC
C  EINC=FROM BIB/INSL/1,BIE/INSL/2,BIS/INSL/3,BIB/INSL/4

COMMON/SIGMA/ SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
COMMON/AMATF1/X(750,2),Y(750,2),DIST(40,40),NPAC(20),WK(4)
COMMON/COMP/D(4),E(4),ICARA,NVN,NVM,NPON
COMMON/ELEM/A,M,INM,NPCO,DSELC
DIMENSION TEMFX(4),TEMFY(4),XM(4),YM(4),VCVX(4,4),VCVY(4,4),FC(3)
DIMENSION B(4),KBF(5),FCH1(4),FCH2(4),WKARE(10),VX(10),VY(10)
DIMENSION VAVX(4,4),XMA(4),TEMF(4)
DATA SIM/"SIN"/
DATA F/" X"," Y","EMPATE"/
DOUBLE PRECISION DSELC
*****
C
C  LEITURA DOS DADOS
C *****
C
C  READ(5,1)ICARA,N,M,NPCO

1 FORMAT(4I5)
  KINQ=1
  CALL NORMAL(KINQ)
  READ(5,4)(NPAC(I),I=1,NPCO)
4 FORMAT(10I5)
*****
C
C  IMPRESSAO DOS DADOS LIDOS
C *****
C
C  DESCRICAO DOS DADOS
C *****
C
C  X= POPULACAO 1
C  Y= POPULACAO 2
C  ICARA= DIMENSAO DOS VETORES QUE IDENTIFICAM OS PONTOS AMOSTRAIS
C          DAS POPULACOES X E Y
C  N= NUMERO DE PONTOS AMOSTRAIS DA POPULACAO X PARA CONSTRUCAO DA
C      REGRA DISCRIMINANTE
C  M= MESMA DEFINICAO DE N POREM PARA A POPULACAO Y
C  NPCO= NUMERO DE VALORES DE V
C  NPAC= VETOR COM OS VALORES DE V
C *****
C  WRITE(6,6)N,M
*****
C
C  5 FORMAT(///,T40,"IMPRESSAO DOS DADOS",/,T40,"*****"
C      -"/,/,T25," MATRIZ X - COMPRIENTO = ",I3,/,T25,
C      -"/,/,T25," MATRIZ Y - COMPRIENTO = ",I3,/,T24,"POSICAO",T
C      -3E," MATRIZ X ",I5E," MATRIZ Y ")
C  IMAT=N

```

```

IF(N.GT.M)IMAI=M
IF(ICAR4.EQ.3)GO TO 15
IF(ICAR4.EQ.4)GO TO 16
WRITE(6,7)(I,X(I,1),X(I,2),Y(I,1),Y(I,2),I=1,IMAI)
7 FORMAT(T25,I3,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
IF(N.EQ.M)GO TO 20
IF(IMAI.NE.M)GO TO 21
WRITE(6,71)(I,X(I,1),X(I,2),I=M+1,N)
71 FORMAT(T25,I3,T35,F10.5,T45,F10.5)
GO TO 20
21 WRITE(6,72)(I,Y(I,1),Y(I,2),I=N+1,M)
72 FORMAT(T25,I3,T55,F10.5,T65,F10.5)
GO TO 20
15 WRITE(6,30)(I,X(I,1),X(I,2),X(I,3),Y(I,1),Y(I,2),Y(I,3),I=1,IMAI)
30 FORMAT(T25,I3,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5,
1,T85,F10.5)
IF(N.EQ.M)GO TO 20
IF(IMAI.NE.M)GO TO 32
WRITE(6,33)(I,X(I,1),X(I,2),X(I,3),I=M+1,N)
33 FORMAT(T25,I3,T35,F10.5,T45,F10.5,T55,F10.5)
GO TO 20
32 WRITE(6,34)(I,Y(I,1),Y(I,2),Y(I,3),I=N+1,M)
34 FORMAT(T25,I3,T55,F10.5,T65,F10.5,T75,F10.5)
GO TO 20
16 WRITE(6,30)(I,X(I,1),X(I,2),X(I,3),X(I,4),Y(I,1),Y(I,2),
1Y(I,3),Y(I,4),I=1,IMAI)
60 FORMAT(T25,I3,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5,T75,F10.5,
1,T85,F10.5,T95,F10.5,T105,F10.5)
IF(N.EQ.M)GO TO 20
IF(IMAI.NE.M)GO TO 42
WRITE(6,61)(I,X(I,1),X(I,2),X(I,3),X(I,4),I=M+1,N)
61 FORMAT(T25,I3,T35,F10.5,T45,F10.5,T55,F10.5,T65,F10.5)
GO TO 20
42 WRITE(6,62)(I,Y(I,1),Y(I,2),Y(I,3),Y(I,4),I=N+1,M)
62 FORMAT(T25,I3,T55,F10.5,T65,F10.5,T75,F10.5,T85,F10.5,T95,F10.5)
20 CONTINUE
INM=N*M

```

```

C *****
C LEITURA DE DADOS
C
C P11= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO X
C P12= PROBABILIDADE A PRIORI PARA PONTOS DA POPULACAO Y
C NVN= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
C POPULACAO X
C NVY= NUMERO DE PONTOS A SEREM CLASSIFICADOS, DA
C POPULACAO Y
C *****

```

```

READ(5,10)P11,P12,NVN,NVY
10 FORMAT(2F10.8,2I5)
VAL=ALOC(P12/P11)
IX=750
NBR(1)=ICARA
NBR(2)=N
NBR(3)=N
NBR(4)=1
NBR(5)=0
NBR(6)=C
CALL ELCOVM(X,IX,NBR,TEMPX,XH,V),IER)
C *****

```

```

C      CONSTRUCAO DA REGRA DISCRIMINATORIA
C*****
      NOR=ICARA
      NBI=4
      CALL VCWTSF(VX,NOR,VCVX,NB8)
      NB4(1)=M
      NB4(3)=M
      CALL BECOVM(Y,IX,NB8,TEMPY,YM,VY,IER)
      CALL VCWTSF(VY,NB8,VCVY,NB8)
      WRITE(6,400)(TEMPX(I),TEMPY(I),YM(I),YM(I),I=1,2)
400  FORMAT(4F10.8)
      WRITE(6,410)((VCVX(I,J),VCVY(I,J),I=1,2),J=1,2)
410  FORMAT(1X,"MATRIZ DE VARIANCIA E COVARIANCIA",//,2F20.8)
      D1=0.0
      CALL LINV3F(VCVY,0,1,ICARA,4,D1,D2,WKAREA,IER)
      WRITE(6,410)((VCVY(I,J),I=1,2),J=1,2)
      DETY=D1*(2.0**D2)
      WRITE(6,420)D1,D2,DETY
420  FORMAT(1X,"COMPONENTES",5F20.8)
      D1=0.0
      CALL LINV3F(VCVX,0,1,ICARA,4,D1,D2,WKAREA,IER)
      WRITE(6,410)((VCVX(I,J),I=1,2),J=1,2)
      DETX=D1*(2.0**D2)
      WRITE(6,420)D1,D2,DETX
      AAA=ALD((ABS(DETY))/BS(DETX))
      WRITE(6,/)AAA
      KINB=2
      CALL NORMAL(MINE)
      IX=750
      NB4(1)=ICARA
      NB4(2)=NVN
      NB4(3)=NVN
      NB4(4)=1
      NB4(5)=0
      NB4(6)=0
      CALL BECOVM(X,IX,NB4,TEMPA,XMA,IX,IER)
      NOR=ICARA
      NBE=4
      CALL VCWTSF(VX,NOR,VVX,NB4)
      WRITE(6,450)(TEMPX(I),XM(I),I=1,2)
450  FORMAT(2F10.8)
      WRITE(6,410)((VVX(I,J),I=1,2),J=1,2)
460  FORMAT(1X,"MATRIZ DE VARIANCIA E COVARIANCIA",//,2F20.8)
      KK=1
470  WRITE(6,300)
300  FORMAT(1H,///,T40,"CLASSIFICACAO QUADRATICA",//,T40,24(" "),//
1T14,"ELEMENTO",T25,"DA POP",T42,"COORDENADAS",
2T57,"CLASSIFICADO NA POP",T81,"DENSIDADE ESTIMADA")
      DO 200 I=1,NVN
      DO 210 J=1,ICARA
      PON1(J)=X(I,J)-XM(J)
210  PON2(J)=X(I,J)-YM(J)
      DO 220 J2=1,ICARA
      DO 230 J2=1,ICARA
      ERE=ERE+PON2(J2)*VCVY(J2,J1)
230  CCC=CCC+PON1(J2)*VCVX(J2,J1)
      EEL=EEL+DDD+PON1(J1)
      CCC=0.0
      ERE=0.0

```

```

220 CONTINUE
    TTS=0.5*(AAA+CCC-EEE)
    CCC=0.0
    EEE=0.0
C*****
C      CALCULO DA ESTIMATIVA DE PROBABILIDADE TOTAL DE ERRO
C      DE CLASSIFICACAO
C*****
    IF (TTS.GE.VAL) GO TO 250
    IMPRE=2
    ISOMAY=ISOMAX*1
    GO TO 260
250 IMPRE=1
    ISOMAX=ISOMAX*1
260 WRITE(6,300)1,F(KK),X(I,1),X(I,2),F(IMPRE),YES
310 FORMAT(13,I3,TZ4,A5,X,F10.5,2),F10.5,T62,I6,T84,F12.5)
200 CONTINUE
    IF (KK.EQ.2) ISOMAY=ISOMAX
    SOMAY=ISOMAY
    SOMA=SOMAY/NVN
    ISOMAY=NVN*ISOMAX
    WRITE(6,320)F(KK),NVN,ISOMAX,ISOMAY,SOMA
320 FORMAT(1H1,T40,"CLASSIFICACAO PARA PONTOS DE POP. ",A6,
1" E PROE. DE ERRO",/,T40,56(" "),/,T10,"TOTAL DE PONTOS",
2T30,"CLASSIF. NA POP. X",T50,"CLASSIF. NA POP. Y",T80,
3"PROB. DE ERRO",/,T15,I3,T38,I3,T58,I2,T82,F10.6)
    IF (KK.EQ.2) GO TO 230
    KINC=3
    CALL NORMAL(KINC)
    NER(1)=ICARA
    NER(2)=NVN
    NER(3)=NVN
    NER(4)=1
    NER(5)=C
    NER(6)=G
    CALL DCCJVM(X,IX,NBR,TEMPA,XMA,VX,IER)
    NOR=ICARA
    NEE=4
    CALL VCATSF(VX,NOR,VVX,NBB)
    WRITE(6,470)(TEMPX(I),XM(I),I=1,2)
470 FORMAT(2F10.8)
    WRITE(6,480)((VXVX(I,J)),I=1,2),J=1,2)
480 FORMAT(1X,"T12 DE VARIANCIA E COVARIANCIA",/,F20.8)
    PTOT=SOMA*PI2
    NVNP=NVN
    NVN=NVN
    KK=3
    ISOMAX=C
    ISOMAY=C
    GO TO 270
270 NVN=NVNP
    PTOT=PTOT+PI2*SOMA
    WRITE(6,330)PTOT
330 FORMAT(1H1,EX,"PROBABILIDADE TOTAL DE ERRO DE CLASSIFICACAO IGUAL"
1,F10.6)
    CLOSE(5,DISF=KEEP)
    STOP
    END

```



```

*****
GERADOR DE VETORES DA DISTRIBUICAO NORMAL MULTIVARIADA
*****

```

```

SUBROUTINE GERAN(NR,K,SIGMA,RVEC,KVEC,DSEED)
INTEGER IR,IER
REAL SIGMA(10),RVEC(750,2),KVEC(4)
DOUBLE PRECISION DSEED
IF=750
CALL GGNHCDSEED,NR,K,SIGMA,IR,RVEC,KVEC,IER)
RETURN
END

```

```

*****
SUBROTINA QUE DEFA AS AMOSTRAS PARA CONSTRUCAO DA REGRA
DISCRIMINATORIA E OS PONTOS A SEFEM CLASSIFICADOS
POPULACOES X E Y
*****

```

```

SUBROUTINE NORM(L(KIND))
COMMON/SIGMA/ SIGMAX(10),SIGMAY(10),XMEDIA(4),YMEDIA(4)
COMMON/AMATRI/ X(750,2),Y(750,2),DIST(40,40),NPAC(20),WK(4)
COMMON/COMP/ C(4),E(4),ICAR/,NVN,NVN,NPON
COMMON/ELEM/ I,M,INM,NFCO,DSEED
COMMON/AB/ IFIS(20),PCN(400,4),SCMAX(20),SOMAY(20)
DIMENSION SMAX(10),SMAY(10)
DOUBLE PRECISION DSEED
IF(KIND.EQ.1)GO TO 1
IF(KIND.EQ.2)GO TO 2
WK(1)=0.0
CALL GERAN(NVN,ICARA,SMAY,X,WK,DSEED)
NPON=NVN
DO 255 JJ=1,ICARA
DO 270 II=1,NPON
270 XC(1,JJ)=X(11,JJ)+YMEDIA(JJ)
255 CONTINUE
GO TO 3
1 READ(5,1210)DSEED

```

```

*****
LEITURA DE CACUS

```

```

XMEDIA= VETOR DE MEIAS PARA A DISTRIBUICAO NORMAL
ASOCIADA A POPULACAO X
YMEDIA= IDEM PARA A POPULACAO Y
SIGMAX= MATRIZ DE COVARIANCIA PARA A DISTRIBUICAO
NORMAL ASSOCIADA A POPULACAO X
SIGMAY= IDEM PARA A POPULACAO Y
DSEED= SEMENTE PARA O GERADOR DE VETORES DA
DISTRIBUICAO NORMAL MULTIVARIADA
*****

```

```

1230 FORMAT(20.0)
READ(5,100)(SIGMAX(I),I=1,10)
100 FORMAT(2F10.0)
READ(5,100)(SIGMAY(I),I=1,10)
DO 10 I=1,10
  SMAX(ISA)=SIGMAX(ISA)
  SMAY(ISA)=SIGMAY(ISA)
10 WRITE(6,/)SMAX(ISA),SMAY(ISA)
  WK(1)=0.0
  CALL GERAN(N,ICARA,SIGMAX,X,WK,DSEED)
  WK(1)=0.0
  CALL GERAN(N,ICARA,SIGMAY,Y,WK,DSEED)
  READ(5,100)(XMECIA(I),I=1,ICARA),(YMECIA(J),J=1,ICARA)
  DO 130 JJ=1,ICARA
    DO 140 II=1,N
      140 X(II,JJ)=X(II,JJ)+XMECIA(JJ)
    DO 150 KK=1,M
      150 Y(KK,JJ)=Y(KK,JJ)+YMECIA(JJ)
  130 CONTINUE
  DO 10 3
  2 WK(1)=0.0
  CALL GERAN(NV,N,ICARA,SMAX,X,WK,DSEED)
  NFGN=NVN
  DO 3 JJ=1,ICARA
    DO 5 II=1,NFGN
      5 X(II,JJ)=X(II,JJ)+XMECIA(JJ)
  3 CONTINUE
  3 CONTINUE
  RETURN
  END

```

=====

BIBLIOGRAFIA

- ANDERSON, T.W. 1966. Some nonparametric multivariate procedures based on statistically equivalent blocks. *Proceedings of an International Symposium on Multivariate Analysis*. New York, Academic Press. p.5-27.
- CACOULOS, T. 1966. On a class of admissible partitions. *Annals of Mathematical Statistics*, 37:189-195.
- COVER, T.M. & HART, P.E. 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21-27.
- DEVROYE, L.P. & WAGNER, T.J. 1979a. Distribution-free inequalities for the deleted and holdout error estimates. *IEEE Transactions on Information Theory*, 25:202-207.
- DEVROYE, L.P. & WAGNER, T.J. 1979b. Distribution-free performance bounds with the resubstitution error estimate. *IEEE Transactions on Information Theory*, 25:208-210.
- FISHER, R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179-188.
- FIX, E. & HODGES, J.L. 1951. *Discriminatory analysis, nonparametric discrimination: consistency properties*. Texas, U.S. School of Aviation Medicine. (Project 21-49-004, Report 4)
- FIX, E. & HODGES, J.L. 1952. *Discriminatory analysis, nonparametric discrimination: small sample performance*. Texas, U.S. School of Aviation Medicine. (Project 21-49-004, Report 11)
- FRASER, D.A.S. 1957. *Nonparametric methods in statistics*. New York, John Wiley.
- FRASER, D.A.S. 1951. Sequentially determined statistically equivalent blocks. *Annals of Mathematical Statistics*, 22:372-381.
- FRASER, D.A.S. & GUTTMAN, I. 1956. Tolerance regions. *Annals of Mathematical Statistics*, 27:162-179.

- FRITZ, J. 1975. Distribution-free exponential error bound for nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 21:552-557.
- FUKUNAGA, K. & HOSTETLER, L.D. 1975. K-nearest neighbor Bayes-risk estimation. *IEEE Transactions on Information Theory*, 21:285-293.
- FUKUNAGA, K. & KESSEL, D.L. 1973. Nonparametric Bayes error estimation using unclassified samples. *IEEE Transactions on Information Theory*, 19:434-440.
- GARNETT, J.M. & YAU, S.S. 1977. Nonparametric estimation of the Bayes error of feature extractors using ordered nearest neighbor sets. *IEEE Transactions on Computers*, 26:46-54.
- GESSAMAN, M.P. 1970. A consistent nonparametric multivariate density estimator based on statistically equivalent blocks. *Annals of Mathematical Statistics*, 41:1344-1346.
- GESSAMAN, M.P. & GESSAMAN, P.H. 1972. A comparison of some multivariate discrimination procedures. *Journal of the American Statistical Association*, 67:468-472.
- GILBERT, E.S. 1969. The effects of unequal variance covariance matrices on Fisher's linear discriminant function. *Biometrics*, 25:505-516.
- GUPTA, S.D. 1963. Nonparametric classification rules. *Sanhkya. Serie A*: 25-30.
- HELLMAN, M.E. 1970. The nearest neighbor classification rule with a reject option. *IEEE Transactions on Systems Sciences Cybernetics*, 6:179-185.
- KEMPERMAN, J.H.B. 1956. Generalized tolerance limits. *Annals of Mathematical Statistics*, 27:180-186.
- KENDALL, M.G. & STUART, A. 1966. *The advanced theory of Statistics*. London, Griffin. v.3.
- LACHENBRUCH, P.A. 1975. *Discriminant analysis*. New York, Hafner Press.
- LACHENBRUCH, P.A. & GOLDSTEIN, M. 1979. Discriminant analysis. *Biometrics*, 35:69-85.
- LACHENBRUCH, P.A.; SNEERINGER, C.; REVO, L. 1973. Robustness of the linear and quadratic discriminant function to certain types of non-normality.

Communications in Statistics, 1:39-57.

LOFTSGAARDEN, D.O. & QUESENBERRY, C.P. 1965. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36: 1049-1051.

MARKS, S. & DUNN, O.J. 1974. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, 69: 555-559.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. 1963. *Introduction to the theory of statistics*. New York, McGraw-Hill.

MURPHY, R.B. 1948. Non-parametric tolerance limits. *Annals of Mathematical Statistics*, 19:581-589.

PARZEN, E. 1962. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33:1065-1076.

PATRICK, E.A. & FISCHER, F.P. 1970. A generalized k-nearest neighbor rule. *Information and Control*, 16:128-152.

PENROD, C.S. & WAGNER, T.J. 1979. Risk estimation for non-parametric discrimination and estimation study: a simulation study. *IEEE Transactions on Information Theory*, 25:753-758.

QUESENBERRY, C.P. & GESSAMAN, M.P. 1968. Nonparametric discrimination using tolerance regions. *Annals of Mathematical Statistics*, 39:664-673.

ROGERS, W.H. & WAGNER, T.J. 1978. A finite sample distribution-free performance bound for local discrimination rules. *The Annals Statistics*, 6: 506-514.

SCHWARTZ, S.C. 1967. Estimation of a probability density by an orthogonal series. *Annals of Mathematical Statistics*, 38:1261-1265.

TUKEY, J.W. 1947. Non-parametric estimation II. Statistically equivalent blocks and tolerance regions-the continuous case. *Annals of Mathematical Statistics*, 18:529-539.

WAGNER, T.J. 1971. Convergence of the nearest neighbor rule. *IEEE Transactions on Information Theory*, 17:566-571.

WAGNER, T.J. 1973. Deleted estimates of the Bayes risk. *The Annals of Statistics*, 1:359-362.

- WALD, A. 1943. An extension of Wilks' method for setting tolerance limits. *Annals of Mathematical Statistics*, 14:45-55.
- WEGMAN, E.J. 1972. Nonparametric probability density estimation I: a summary of available methods. *Technometrics*, 14:533-546.
- WELCH, B.L. 1939. Note on discriminant functions. *Biometrika*, 31:218-220.
- WILKS, S.S. 1941. Determination of sample sizes for setting tolerance limits. *Annals of Mathematical Statistics*, 12:91-96.
- WILKS, S.S. 1962. *Mathematical Statistics*. New York, John Wiley.
- WILKS, S.S. 1942. Statistical prediction with special reference to the problem of tolerance limits. *Annals of Mathematical Statistics*, 13:400-409.