ANÁLISE MULTIVARIADA APLICADA AS CIÊNCIAS AGRÁRIAS

PÓS-GRADUAÇÃO EM AGRONOMIA CIÊNCIA DO SOLO: CPGA-CS

ANÁLISE DISCRIMINANTE

Carlos Alberto Alves Varella[1]

ÍNDICE

[1] Professor. Universidade Federal Rural do Rio de Janeiro, IT-Departamento de Engenharia, BR 465 km 7 - CEP 23890-000 – Seropédica – RJ. E-mail: varella@ufrrj.br.

INTRODUÇÃO

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos. Segundo KHATTREE & NAIK (2000) é uma técnica da estatística multivariada que estuda a separação de objetos de uma população em duas ou mais classes. A discriminação ou separação é a primeira etapa, sendo a parte exploratória da análise e consiste em se procurar características capazes de serem utilizadas para alocar objetos em diferentes grupos previamente definidos. A classificação ou alocação pode ser definida como um conjunto de regras que serão usadas para alocar novos objetos (JOHNSON & WICHERN, 1999). Contudo, a função que separa objetos pode também servir para alocar, e, o inverso, regras que alocam objetos podem ser usadas para separar. Normalmente, discriminação e classificação se sobrepõem na análise, e a distinção entre separação e alocação é confusa. Segundo REGAZZI (2000) o problema da discriminação entre dois ou mais grupos, visando posterior classificação, foi inicialmente abordado por Fisher (1936). Consiste em obter funções matemáticas capazes de classificar um indivíduo X (uma observação X) em uma de várias populações π_i, (i=1, 2, ..., g), com base em medidas de um número p de características, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente um indivíduo em uma população π_i, quando realmente pertence a população π_j, $(i \neq j)$ i,j=1, 2, ..., g.

DISCRIMINAÇÃO E CLASSIFICAÇÃO

O problema consiste em se obter uma combinação linear de características observadas que apresente maior poder de discriminação entre populações. Esta combinação linear é denominada função discriminante. Tal função tem a propriedade de minimizar as probabilidades de má classificação, quando as populações são normalmente distribuídas com média μ e variâncicia Σ conhecidas. Entretanto, tal situação não ocorre, isto é, a média e a variância das populações normalmente não são conhecidas, portanto havendo a necessidade de estimação desses parâmetros. Podemos assumir que as populações têm uma mesma matriz de covariâncias ou não. Conforme a seleção as funções discriminantes são denominadas de lineares ou quadráticas. No caso particular da função de FISHER assume-se que as matrizes de covariâncias são iguais e é dita função discriminante linear de Fisher.

REGIÕES DE ALOCAÇÃO

Regiões de alocação são conjunto de valores separados por uma fronteira definida por uma função discriminante qualquer. Essa função discriminante é obtida a partir de amostras de treinamento. Pode ter como base modelos estatísticos ou não, tais como redes neurais e lógica fuzzy. Então, uma observação pode ser alocada como sendo da população π_1 e ou da população π_2. Contudo é importante observar que no mundo real a fronteira entre regiões não está exatamente definida e sempre haverá superposição, isto é, erro de classificação. A Figura 1 ilustra regiões de alocação para o caso de duas populações.

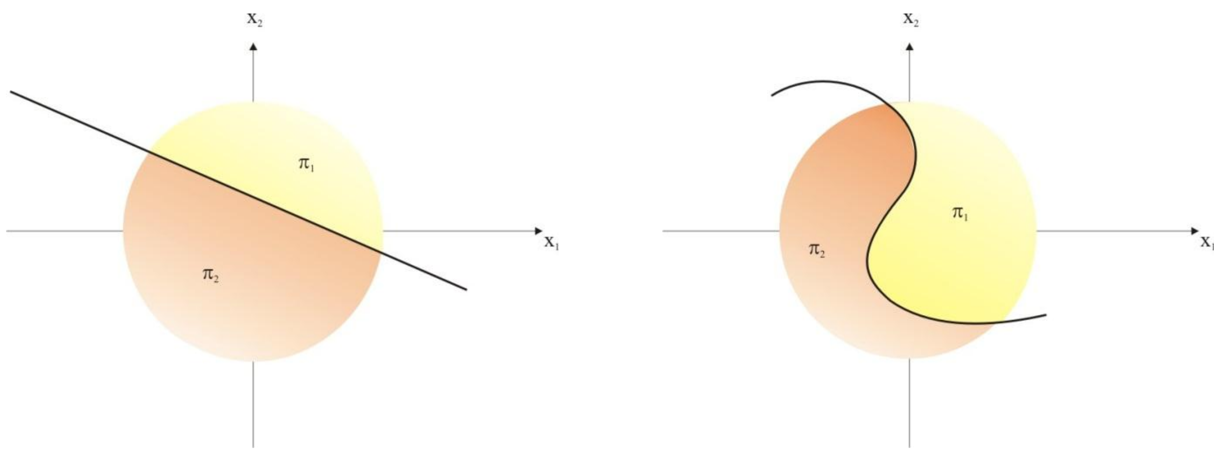

Figura 1. Regiões de alocação para o caso de duas populações.

REGRAS DE CLASSIFICAÇÃO

Uma boa classificação deve resultar em pequenos erros, isto é, deve haver pouca probabilidade de má classificação. Segundo JOHNSON & WICHERN (1999) para que isso ocorra a regra de classificação deve considerar as probabilidades a priori e os custos de má classificação. Outro fator que uma regra de classificação deve considerar é se as variâncias das populções são iguais ou não. Quando a regra de classificação assume que as variâncias das populações são iguais, as funções discriminantes são ditas lineares e quando não são funções discriminantes quadráticas. Regras de classificação também podem ser construídas com base em modelos de redes neurais ou lógica fuzzy. Segundo GONZALEZ & WOODS (1992), citado por KHOURY JR. (2004), em comparação com classificadores estatísticos, que determinam planos lineares ou quadráticos, o maior benefício da modelagem por redes neurais é sua capacidade de determinar planos não-lineares de separação de classes.

FUNÇÃO DISCRIMINANTE LINEAR DE FISHER

A função discriminante linear de Fisher é uma combinação linear de características originais a qual se caracteriza por produzir separação máxima entre duas populações.

Considerando que μ_i e Σ são parâmetros conhecidos e respectivamente, os vetores de médias e a matriz de covariâncias comum das populações π_i. Demonstra-se que a função linear do vetor aleatório X que produz separação máxima entre duas populações é dada por:

$$D(X) = L' \cdot X = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot X$$

em que,

$X = \lfloor X_1 \; X_2 \; \cdots \; X_p \rfloor$ e $\pi = [\pi_1, \pi_2]$

L = vetor discriminante;
X = vetor aleatório de características das populações;
μ = vetor de médias p-variado;
Σ = matriz comum de covariâncias das populações π_1 e π_2;

O valor da função discriminante de Fisher para uma dada observação x_o é:

$$D(x_o) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_o$$

O ponto médio entre as duas médias populacionais univariadas μ_1 e μ_1 é:

$$m = \frac{1}{2}[\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot [\mu_1 + \mu_2] \text{, ou seja}$$

$$m = \frac{1}{2}[D(\mu_1) + D(\mu_2)]$$

A regra de classificação baseada na função discriminante de Fisher é:

Alocar x_o em π_1 se $D(x_o) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_o \geq m$
Alocar x_o em π_2 se $D(x_o) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_o < m$

Assumindo-se que as populações π_1 π_2 têm a mesma matriz de covariâncias Σ podemos então estimar uma matriz comum de covariâncias S_c:

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_2$$

em que,

S_c	=	estimativa da matriz comum de covariâncias Σ;
n_1	=	número de observações da população π_1;
n_2	=	número de observações da população π_2;
S_1	=	estimativa matriz de covariâncias da população π_1;
S_2	=	estimativa matriz de covariâncias da população π_2;

A função discriminante linear amostral de Fisher é obtida substituindo-se os parâmetros μ_1, μ_2 e Σ pelas respectivas quantidades amostrais $\bar{x}_1$, $\bar{x}_2$ e S_c:

$$D(x) = \hat{L}' \cdot x = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x$$

em que,

$D(x)$	=	função discriminante linear amostral de Fisher;
$\hat{L}'$	=	estimativa do vetor disriminante;
$\bar{x}_1$	=	média amostral da população π_1;
$\bar{x}_2$	=	média amostral da população π_2.

EXEMPLO DE APLICAÇÃO

Como exemplo ilustrativo para obetenção da função discriminante linear amostral de Fisher, vamos considerar os dados de duas raças de insetos (Quadro 1), apresentados por HOEL (1966) e citado por REGAZZI (2000).

Quadro 1. Número médio de cerdas primordiais (X_1) e número médio de cerdas distais (X_2) em duas raças de insetos

Raça A		Raça B	
X_1	X_2	X_1	X_2
6,36	5,24	6,00	4,88
5,92	5,12	5,60	4,64
5,92	5,36	5,64	4,96
6,44	5,64	5,76	4,80
6,40	5,16	5,96	5,08
6,56	5,56	5,72	5,04
6,64	5,36	5,64	4,96
6,68	4,96	5,44	4,88
6,72	5,48	5,04	4,44
6,76	5,60	4,56	4,04
6,72	5,08	5,48	4,20
		5,76	4,80

Estimativa das médias das raças A e B

Com base nos dados apresentados no Quadro 1, temos:

Raça A

$$\bar{\chi}_A = \begin{bmatrix} \bar{x}_{A1} \\ \bar{x}_{A2} \end{bmatrix} = \begin{bmatrix} 6,46545 \\ 5,32364 \end{bmatrix} \quad \text{e} \quad S_A = \begin{bmatrix} 0,091287 & 0,011258 \\ 0,011258 & 0,052625 \end{bmatrix}$$

Raça B

$$\bar{\chi}_B = \begin{bmatrix} \bar{x}_{B1} \\ \bar{x}_{B2} \end{bmatrix} = \begin{bmatrix} 5,55000 \\ 4,72667 \end{bmatrix} \quad \text{e} \quad S_B = \begin{bmatrix} 0,160327 & 0,107418 \\ 0,107418 & 0,111661 \end{bmatrix}$$

Assumindo-se que $\Sigma_A = \Sigma_B = \Sigma$, a matriz comum de covariâncias amostral S_c é dada por:

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)}\right] \cdot S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)}\right] \cdot S_2$$

$$S_c = \left[\frac{11 - 1}{(11 - 1) + (12 - 1)}\right] \cdot S_1 + \left[\frac{12 - 1}{(n_1 - 1) + (n_2 - 1)}\right] \cdot S_2$$

$$S_c = \begin{bmatrix} 0{,}12745 & 0{,}06162 \\ 0{,}06162 & 0{,}08354 \end{bmatrix} \quad \text{e} \quad S_c^{-1} = \begin{bmatrix} 12{,}1960015 & -8{,}995964 \\ -8{,}995464 & 18{,}604583 \end{bmatrix}$$

Obtenção da função discriminante linear amostral de Fisher

$$D(x) = \hat{L}' \cdot x = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x$$

$$[\bar{x}_1 - \bar{x}_2] = \begin{bmatrix} 6{,}46545 & -5{,}55000 \\ 5{,}32364 & -4{,}72667 \end{bmatrix} = \begin{bmatrix} 0{,}91545 \\ 0{,}59697 \end{bmatrix}$$

$$[\bar{x}_1 - \bar{x}_2]' = [0{,}91545 \quad 0{,}59697]$$

$$\hat{L}' = [0{,}91545 \quad 0{,}59697] \cdot \begin{bmatrix} 12{,}196015 & -8{,}995964 \\ -8{,}995464 & 18{,}604583 \end{bmatrix}$$

$\hat{L}' = [5{,}794819 \quad 2{,}871023]$ é o estimador do vetor discriminante e:

$$D(x) = [5{,}794819 \quad 2{,}871023] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$D(x) = 5{,}794819x_1 + 2{,}871023x_2$$

Construção da regra para alocação de novos indivíduos

A questão é se um novo indivíduo ou uma nova observação X_o pertence a raça A (π_1) ou a raça B (π_2). Então vamos aplicar a regra de classificação com base na função discriminante de Fisher. Primeiro determinamos o ponto médio das populações $\hat{m}$:

$$\hat{m} = \frac{1}{2} \cdot \left[D(\bar{x}_A) + D(\bar{x}_B)\right]$$

$$D(\bar{x}_A) = \hat{L}' \cdot \bar{x}_A = \begin{bmatrix} 5,794819 & 2,871023 \end{bmatrix} \cdot \begin{bmatrix} 6,46545 \\ 5,32364 \end{bmatrix}$$

$$D(\bar{x}_A) = 52,750405$$

$$D(\bar{x}_B) = \hat{L}' \cdot \bar{x}_B = \begin{bmatrix} 5,794819 & 2,871023 \end{bmatrix} \cdot \begin{bmatrix} 5,55000 \\ 4,72667 \end{bmatrix}$$

$$D(\bar{x}_B) = 45,731624$$

$$\hat{m} = \frac{1}{2} \cdot (52,750405 + 45,731624) = 49,241$$

$$\hat{m} = 49,241$$

A regra de classificação é:

Alocar x_o em Raça A se $D(x_o) \geq 49,241$
Alocar x_o em Raça B se $D(x_o) < 49,241$

Tendo-se um novo indivíduo x_o que apresenta número médio de cerdas primordiais e distais de 6,21 e 5,31, respectivamente, aplicamos a regra de classificação:

$$D(\bar{x}_o) = \hat{L}' \cdot \bar{x}_o = \begin{bmatrix} 5,794819 & 2,871023 \end{bmatrix} \cdot \begin{bmatrix} 6,21 \\ 5,31 \end{bmatrix} = 51,230958$$

Como $D(\bar{x}_o) = 51,230958 > 49,241$, o indivíduo é alocado na Raça A.

FUNÇÕES DISCRIMINANTES DE ANDERSON

Sejam $\pi_1, \pi_2, ..., \pi_g$ um grupo de 'g' populações. Para se obter as funções discriminantes para esse grupo de populações são necessárias algumas pressuposições: 1) as populações apresentam algum tipo de distribuição; 2) existe uma probabilidade de ocorrência a priori para cada população no grupo; 3) existe um custo de má classificação. Antes de se construir regras de classificação, recomenda-se fazer um estudo detalhado de quais 'p' variáveis têm efeito siginificativo no fenômeno. Essa parte da análise é denominada por alguns autores como discriminação. Nesta etapa é fundamental a experiência do pesquisador para que a técnica obtenha sucesso.

Desenvolvimento do classificador

Os classificadores são desenvolvidos da necessidade de se alocar uma observação 'x' em uma dentre 'g' populações, sendo g > 2. Para desenvolver um classificador é necessário fazer algumas pressuposições para o modelo da função discriminate, neste caso as pressuposições são as seguintes:

1) As 'g' populações apresentam distribuição normal multivariada;
2) As 'p_i' probabilidades a priori de ocorrência das populações são iguais e $\sum_{i=1}^{n} p_i = 1$.
3) As populações apresentam custos iguais de má classificação.

Obtenção das funções discriminantes

Considerando-se que as 'g' populações apresentam distribuição normal multivariada a função discriminante é dada por:

$$\mathrm{D_i}(\tilde{\mathrm{x}}) = -\frac{1}{2}\ln|\Sigma_\mathrm{i}| - \frac{1}{2}[\tilde{\mathrm{x}} - \tilde{\mu}_\mathrm{i}]'\,\Sigma_\mathrm{i}^{-1}[\tilde{\mathrm{x}} - \tilde{\mu}_\mathrm{i}] + \ln(\mathrm{p_i}), \mathrm{i} = 1, 2, \cdots, \mathrm{g}$$

em que,

$\mathrm{D_i}(\tilde{\mathrm{x}})$	=	função discriminante da população 'i' do vetor aleatório $\tilde{x}$;
Σ_i	=	matriz de covariância da população 'i';
$\tilde{x}$	=	vetor aleatório de características;
$\tilde{\mu}_i$	=	vetor de médias da população 'i';
p_i	=	propabilidade de ocorrência da população 'i'.

Supondo igualdade das matrizes de covariâncias, então os componentes constantes para todo 'i' podem ser retirados e a função discriminante é:

$$D_i(\tilde{x}) = \tilde{L}_i' \cdot \tilde{x} - \frac{1}{2} \cdot \tilde{L}_i' \cdot \tilde{\mu}_i + \ln(p_i)$$

em que,

$D_i(\tilde{x})$	=	função discriminante da população 'i' do vetor aleatório $\tilde{x}$;
$\tilde{L}_i$	=	vetor aleatório discriminante da população 'i';
$\tilde{x}$	=	vetor aleatório de características;
$\tilde{\mu}_i$	=	vetor de médias da população 'i';
p_i	=	propabilidade de ocorrência da população 'i'.

sendo que,

$$\tilde{L}_i = \Sigma^{-1} \cdot \mu_i$$

Σ = matriz comum de covariâncias das 'g' populações.

A regra de classificação para alocar um indivíduo 'x' é a seguinte: classificar 'x' em π_i se e somente se

$$D_i(\tilde{x}) = \max\left(D_1(\tilde{x}),\ D_2(\tilde{x}), \ldots, D_g(\tilde{x})\right)$$

em que,

$D_i(\tilde{x})$	=	valor da função discriminante da população 'i' para o vetor de características $\tilde{x}$;
$D_1(\tilde{x})$	=	valor da função discriminante da população '1' para o vetor de características $\tilde{x}$;
$D_2(\tilde{x})$	=	valor da função discriminante da população '2' para o vetor de características $\tilde{x}$;
$D_g(\tilde{x})$	=	valor da função discriminante da população 'g' para o vetor de características $\tilde{x}$.

A regra acima pode também ser utilizada para o caso particular de g = 2.

Teste de igualdade das matrizes de covariâncias

A próxima etapa é fazer inferência sobre a igualdade das matrizes de covariâncias das populações. Se a opção for pela igualdade das matrizes 'Σ_i' então a função discriminante é dita linear de Anderson, caso contrário é dita função discriminante quadrática de Anderson. Segundo JOHNSON & WICHERN (1999), é possível aplicar um teste para a igualdade das matrizes de covariâncias das populações. Contudo, o resultado desse teste não é condição suficiente para selecionar o modelo da função discriminante (linear ou quadrático). Assim, recomenda-se a aplicação de um teste de validação para decidir qual o melhor modelo a ser adotado. Nesta etapa o processo de seleção do modelo é do tipo iterativo, onde testamos as funções discriminantes a partir de resultados obtidos na classificação dos indivíduos.

No Quadro 2 são apresentados resultados da acurácia de classificação obtidos por VARELLA (2004) na avaliação de funções discriminantes lineares e quadráticas. Observa-se que a acurácia não foi a mesma para as funções discriminantes lineares e quadráticas. Desta maneira, seleciona-se a a função que apresentar maior acurácia de classificação nas amostras de teste (25% total). Observa-se também que na maioria dos casos a função linear apresentou melhor resultado. Segundo HOFFBECK & LANDGREBE (1996), em situações em que o número de observações utilizadas para o treinamento do classificador é limitado, a estimativa de uma covariância comum para todas as populações pode resultar numa melhor classificação, devido a redução dos parâmetros a serem estimados.

Quadro 2. Acurácia de classificação de funções discriminantes lineares e quadráticas obtidas na classificação de amostras de teste

Área	Vôo	Estádio da cultura do milho	Acurácia de classificação (%)	
			Função Linear	Função Quadrática
1	1	VT	92	100
	2	R1	91	9
2	1	R1	100	62
	2	R2	60	100
3	1	V12	100	63
	2	VT	100	92

Estimativa da matriz comum de covariâncias amostral

Supondo a igualdade das matrizes de covariancias Σ_i a covariância amostral é estimada por S_c, dada por:

$$S_c = \frac{(n_1 - 1) \cdot S_1 + (n_2 - 1) \cdot S_2 + \cdots + (n_g - 1) \cdot S_g}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_g - 1)}$$

S_c = estimativa da matriz comum de covariâncias das ‘g’ populações.

Avaliação das funções discriminantes

Essa etapa consiste em se avaliar a acurácia de classificação por meio do coeficiente kappa ‘κ’, obtido a partir da matriz de erros da classificação (CONGALTON e MEAD, 1983). Para se obter a matriz de erros usamos 25% do total de observações e os restantes 75% são usados para obter as funções discriminantes (VARELLA, 2004). Recomenda-se testar os dois modelos de funções discriminantes, isto é o modelo linear e o modelo quadrático.

A matriz de erros é de dimensão g x g, em que g é o número de populações envolvidas na análise discriminante. As colunas dessa matriz apresentam as informações das observações de referência, enquanto as linhas as informações das observações classificadas. Dessa maneira, na diagonal estão o número de observações corretamente classificadas. A partir dessa matriz são calculados os erros de omissão, de comissão, a exatidão global e o coeficiente kappa.

A exatidão global é determinada pela seguinte expressão:

$$EG = \frac{N_c}{N_t} \cdot 100 \tag{1}$$

em que:

EG = exatidão global, %;
N_c = número de observações corretamente classificadas;
N_t = número total de observações.

O coeficiente kappa é estimado pela seguinte expressão:

$$\hat{K} = \frac{n\sum_{i=1}^{c} X_{ii} - \sum_{i=1}^{c} X_{i+}X_{+i}}{n^2 - \sum_{i=1}^{c} X_{i+}X_{+i}} \tag{5}$$

em que:

c = número de classes na matriz de erros;
X_{ii} = valores na linha i e na coluna i;
X_{i+} = total da linha i;
X_{+i} = total da coluna i;
n = número total de observações.

Os coeficientes kappas são comparados pelo teste Z determinado pela seguinte expressão (CONGALTON & MEAD, 1983):

$$Z = \frac{\hat{K}_1 - \hat{K}_2}{\sqrt{\hat{\sigma}_1 + \hat{\sigma}_2}} \tag{6}$$

em que:

Z	=	valor Z calculado;
$\hat{K}_1$	=	estimativa do coeficiente Kappa do classificador 1;
$\hat{K}_2$	=	estimativa do coeficiente Kappa do classificador 2;
$\hat{\sigma}_1$	=	estimativa da variância do Kappa do classificador 1;
$\hat{\sigma}_2$	=	Estimativa da variância do Kappa do classificador 2.

Se o valor Z calculado para o teste for maior que o valor Z tabelado, diz-se que o resultado foi significativo e rejeita-se a hipótese nula (Ho: $K_1 = K_2$) concluindo-se que os dois classificadores são estatisticamente diferentes. O valor de Z tabelado ao nível de 5% de probabilidade é igual a 1,96.

EXEMPLO DE APLICAÇÃO

Como exemplo ilustrativo para obetenção de função discriminantes de Anderson, vamos considerar os dados do Quadro 2.), apresentados por HOEL (1966) e citado por REGAZZI (2000).

PROGRAMAÇÃO ‘SAS’ PARA ANÁLISE DISCRIMINANTE

PROC DISCRIM: procedimento empregado no programa computacional SAS para realizar análises discriminantes. A forma geral do proc discrim é:

```
PROC DISCRIM <options>;
CLASS <class var>;
VAR <var1 var2 var3 ... var n>;
PRIORS <options>;
```

O SAS apresenta diversas opções para o procedimento PROC DISCRIM. As opções mais usuais são:

OUTSTAT = salva a função discriminante.
CROSSVALIDATE – retorna a estimativa da acurácia e erro da classificação. Usa o método da validação cruzada deixando um de fora.
CROSSLISTERR – apresenta o erro para cada observação do CROSSVALIDATE.
POOL = seleciona se as matrizes de covariâncias são iguais ou diferentes.
POOL = TEST testa se as matrizes são iguais e usa o resultado para proceder as análises subsequentes.

Se POOL=YES o SAS assume que as matrizes de covariâncias da populações são iguais. O resultado é uma **Função Discriminante Linear.**

Se POOL=NO o SAS assume que as matrizes de covariâncias das populações são diferentes. O resultado é uma **Função Discriminante Quadrática.**

CLASS – define a variável que representa a classe ou população.
VAR – define as variáveis de resposta. Neste caso é o vetor de características, isto é, as variáveis consideradas com efeito na discriminação.
PRIORS – especifica a probabilidade a priori de dada uma observação esta pertencer ao grupo. Pode ser EQUAL ou PROPORTIONAL. EQUAL assume probabilidades iguais para todas as classes. PROPORTIONAL assume probabilidades proporcionais ao número de observações de cada classe. Na ausência de informações sobre as probabilidades de ocorrência a *priori* de cada classe recomenda-se usar EQUAL.

EXEMPLO DE APLICAÇÃO

Neste exemplo, pretende-se obter funções discriminantes para desenvolver um classificador estatístico que a partir de dados de sensoriamento remoto seja capaz de discriminar cinco tipos de culturas: milho, soja, algodão, beterraba e um tipo de trevo (SAS, 2007).

METHOD=NORMAL assume uma distribuição normal multivariada para todas as classess;

POOL=YES assume a igualdade das matrizes de covariâncias para todas as classes;

PRIORS PROP assume probabilidades proporcionanais ao número de observações de cada classe.

OBTENÇÃO DA FUNÇÃO DISCRIMINANTE

/* Programa SAS para análise discriminante Linear */

```
data crops;
      title 'Discriminant Analysis of Remote Sensing Data
             on Five Crops';
      input Crop $ 4-13 x1-x4 xvalues $ 14-24;
      datalines;
   Corn      16 27 31 33
   Corn      15 23 30 30
   Corn      16 27 27 26
   Corn      18 20 25 23
   Corn      15 15 31 32
   Corn      15 32 32 15
   Corn      12 15 16 73
   Soybeans  20 23 23 25
   Soybeans  24 24 25 32
```

```
   Soybeans  21 25 23 24
   Soybeans  27 45 24 12
   Soybeans  12 13 15 42
   Soybeans  22 32 31 43
   Cotton    31 32 33 34
   Cotton    29 24 26 28
   Cotton    34 32 28 45
   Cotton    26 25 23 24
   Cotton    53 48 75 26
   Cotton    34 35 25 78
   Sugarbeets22 23 25 42
   Sugarbeets25 25 24 26
   Sugarbeets34 25 16 52
   Sugarbeets54 23 21 54
   Sugarbeets25 43 32 15
   Sugarbeets26 54  2 54
   Clover    12 45 32 54
   Clover    24 58 25 34
   Clover    87 54 61 21
   Clover    51 31 31 16
   Clover    96 48 54 62
   Clover    31 31 11 11
   Clover    56 13 13 71
   Clover    32 13 27 32
   Clover    36 26 54 32
   Clover    53 08 06 54
   Clover    32 32 62 16
   ;
   proc discrim data=crops outstat=cropstat
                method=normal pool=yes
                list crossvalidate;
      class Crop;
      priors prop;
      id xvalues;
      var x1-x4;
      title2 'Using Linear Discriminant Function';
   run;
```

/* Fim do programa */

/* Programa SAS para análise discriminante Quadrática */

```
data crops;
      title 'Discriminant Analysis of Remote Sensing Data
             on Five Crops';
      input Crop $ 4-13 x1-x4 xvalues $ 14-24;
      datalines;
   Corn      16 27 31 33
   Corn      15 23 30 30
   Corn      16 27 27 26
   Corn      18 20 25 23
   Corn      15 15 31 32
   Corn      15 32 32 15
   Corn      12 15 16 73
   Soybeans  20 23 23 25
   Soybeans  24 24 25 32
   Soybeans  21 25 23 24
   Soybeans  27 45 24 12
```

```
   Soybeans   12 13 15 42
   Soybeans   22 32 31 43
   Cotton     31 32 33 34
   Cotton     29 24 26 28
   Cotton     34 32 28 45
   Cotton     26 25 23 24
   Cotton     53 48 75 26
   Cotton     34 35 25 78
   Sugarbeets22 23 25 42
   Sugarbeets25 25 24 26
   Sugarbeets34 25 16 52
   Sugarbeets54 23 21 54
   Sugarbeets25 43 32 15
   Sugarbeets26 54  2 54
   Clover     12 45 32 54
   Clover     24 58 25 34
   Clover     87 54 61 21
   Clover     51 31 31 16
   Clover     96 48 54 62
   Clover     31 31 11 11
   Clover     56 13 13 71
   Clover     32 13 27 32
   Clover     36 26 54 32
   Clover     53 08 06 54
   Clover     32 32 62 16
   ;
   proc discrim data=crops
                method=normal pool=no
                crossvalidate;
      class Crop;
      priors prop;
      id xvalues;
      var x1-x4;
      title2 'Using Quadratic Discriminant Function';
      run;
```

/* Fim do programa */

VALIDAÇÃO NA AMOSTRA DE TESTE

```
   data test;
      input Crop $ 1-10 x1-x4 xvalues $ 11-21;
      datalines;
   Corn       16 27 31 33
   Soybeans   21 25 23 24
   Cotton     29 24 26 28
   Sugarbeets54 23 21 54
   Clover     32 32 62 16
   ;



   proc discrim data=cropstat testdata=test testout=tout
                testlist;
      class Crop;
      testid xvalues;
      var x1-x4;
      title2 'Classification of Test Data';
```

```
run;
proc print data=tout;
   title2 'Output Classification Results of Test Data';
run;
```

RESULTADOS DA ANÁLISE

A seguir os resultados obtidos no programa computacional SAS encontrados na janela output.

The DISCRIM Procedure

Observations	36	DF Total	35
Variables	4	DF Within Classes	31
Classes	5	DF Between Classes	4

Class Level Information

Crop	Variable Name	Frequency	Weight	Proportion	Prior Probability
Clover	Clover	11	11.0000	0.305556	0.305556
Corn	Corn	7	7.0000	0.194444	0.194444
Cotton	Cotton	6	6.0000	0.166667	0.166667
Soybeans	Soybeans	6	6.0000	0.166667	0.166667
Sugarbeets	Sugarbeets	6	6.0000	0.166667	0.166667

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
4	21.30189

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} (\bar{X}_i - \bar{X}_j) - 2 \ln PRIOR_j$$

Generalized Squared Distance to Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets
Clover	2.37125	7.52830	4.44969	6.16665	5.07262
Corn	6.62433	3.27522	5.46798	4.31383	6.47395
Cotton	3.23741	5.15968	3.58352	5.01819	4.87908
Soybeans	4.95438	4.00552	5.01819	3.58352	4.65998
Sugarbeets	3.86034	6.16564	4.87908	4.65998	3.58352

Linear Discriminant Function

$$\text{Constant} = -.5\ \bar{X}_j'\ \text{COV}^{-1}\ \bar{X}_j + \ln \text{PRIOR}_j \qquad \text{Coefficient Vector} = \text{COV}^{-1}\ \bar{X}_j$$

Linear Discriminant Function for Crop

Variable	Clover	Corn	Cotton	Soybeans	Sugarbeets
Constant	-10.98457	-7.72070	-11.46537	-7.28260	-9.80179
x1	0.08907	-0.04180	0.02462	0.0000369	0.04245
x2	0.17379	0.11970	0.17596	0.15896	0.20988
x3	0.11899	0.16511	0.15880	0.10622	0.06540
x4	0.15637	0.16768	0.18362	0.14133	0.16408

Discriminant Analysis of Remote Sensing Data on Five Crops 35
Using Linear Discriminant Function 07:52 Friday, March 15, 2002

The DISCRIM Procedure
Classification Results for Calibration Data: WORK.CROPS
Resubstitution Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)'\ \text{COV}^{-1}\ (X-\bar{X}_j) - 2 \ln \text{PRIOR}_j$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5\ D_j^2(X)) \ / \ \mathrm{SUM}_k \exp(-.5\ D_k^2(X))$$

Posterior Probability of Membership in Crop

xvalues	From Crop	Classified into Crop		Clover	Corn	Cotton	Soybeans	Sugarbeets
16 27 31 33	Corn	Corn		0.0894	0.4054	0.1763	0.2392	0.0897
15 23 30 30	Corn	Corn		0.0769	0.4558	0.1421	0.2530	0.0722
16 27 27 26	Corn	Corn		0.0982	0.3422	0.1365	0.3073	0.1157
18 20 25 23	Corn	Corn		0.1052	0.3634	0.1078	0.3281	0.0955
15 15 31 32	Corn	Corn		0.0588	0.5754	0.1173	0.2087	0.0398
15 32 32 15	Corn	Soybeans	*	0.0972	0.3278	0.1318	0.3420	0.1011
12 15 16 73	Corn	Corn		0.0454	0.5238	0.1849	0.1376	0.1083
20 23 23 25	Soybeans	Soybeans		0.1330	0.2804	0.1176	0.3305	0.1385
24 24 25 32	Soybeans	Soybeans		0.1768	0.2483	0.1586	0.2660	0.1502
21 25 23 24	Soybeans	Soybeans		0.1481	0.2431	0.1200	0.3318	0.1570
27 45 24 12	Soybeans	Sugarbeets	*	0.2357	0.0547	0.1016	0.2721	0.3359
12 13 15 42	Soybeans	Corn	*	0.0549	0.4749	0.0920	0.2768	0.1013
22 32 31 43	Soybeans	Cotton	*	0.1474	0.2606	0.2624	0.1848	0.1448
31 32 33 34	Cotton	Clover	*	0.2815	0.1518	0.2377	0.1767	0.1523
29 24 26 28	Cotton	Soybeans	*	0.2521	0.1842	0.1529	0.2549	0.1559
34 32 28 45	Cotton	Clover	*	0.3125	0.1023	0.2404	0.1357	0.2091
26 25 23 24	Cotton	Soybeans	*	0.2121	0.1809	0.1245	0.3045	0.1780
53 48 75 26	Cotton	Clover	*	0.4837	0.0391	0.4384	0.0223	0.0166
34 35 25 78	Cotton	Cotton		0.2256	0.0794	0.3810	0.0592	0.2548
22 23 25 42	Sugarbeets	Corn	*	0.1421	0.3066	0.1901	0.2231	0.1381
25 25 24 26	Sugarbeets	Soybeans	*	0.1969	0.2050	0.1354	0.2960	0.1667
34 25 16 52	Sugarbeets	Sugarbeets		0.2928	0.0871	0.1665	0.1479	0.3056
54 23 21 54	Sugarbeets	Clover	*	0.6215	0.0194	0.1250	0.0496	0.1845
25 43 32 15	Sugarbeets	Soybeans	*	0.2258	0.1135	0.1646	0.2770	0.2191
26 54 2 54	Sugarbeets	Sugarbeets		0.0850	0.0081	0.0521	0.0661	0.7887
12 45 32 54	Clover	Cotton	*	0.0693	0.2663	0.3394	0.1460	0.1789
24 58 25 34	Clover	Sugarbeets	*	0.1647	0.0376	0.1680	0.1452	0.4845
87 54 61 21	Clover	Clover		0.9328	0.0003	0.0478	0.0025	0.0165
51 31 31 16	Clover	Clover		0.6642	0.0205	0.0872	0.0959	0.1322
96 48 54 62	Clover	Clover		0.9215	0.0002	0.0604	0.0007	0.0173

Discriminant Analysis of Remote Sensing Data on Five Crops 36

The DISCRIM Procedure
Classification Results for Calibration Data: WORK.CROPS
Resubstitution Results using Linear Discriminant Function

Posterior Probability of Membership in Crop

xvalues	From Crop	Classified into Crop		Clover	Corn	Cotton	Soybeans	Sugarbeets
31 31 11 11	Clover	Sugarbeets	*	0.2525	0.0402	0.0473	0.3012	0.3588
56 13 13 71	Clover	Clover		0.6132	0.0212	0.1226	0.0408	0.2023
32 13 27 32	Clover	Clover		0.2669	0.2616	0.1512	0.2260	0.0943
36 26 54 32	Clover	Cotton	*	0.2650	0.2645	0.3495	0.0918	0.0292
53 08 06 54	Clover	Clover		0.5914	0.0237	0.0676	0.0781	0.2392
32 32 62 16	Clover	Cotton	*	0.2163	0.3180	0.3327	0.1125	0.0206

* Misclassified observation

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.CROPS
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j) - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	6	0	3	0	2	11
	54.55	0.00	27.27	0.00	18.18	100.00
Corn	0	6	0	1	0	7
	0.00	85.71	0.00	14.29	0.00	100.00
Cotton	3	0	1	2	0	6
	50.00	0.00	16.67	33.33	0.00	100.00
Soybeans	0	1	1	3	1	6
	0.00	16.67	16.67	50.00	16.67	100.00
Sugarbeets	1	1	0	2	2	6
	16.67	16.67	0.00	33.33	33.33	100.00
Total	10	8	5	8	5	36
	27.78	22.22	13.89	22.22	13.89	100.00
Priors	0.30556	0.19444	0.16667	0.16667	0.16667	

Error Count Estimates for Crop

	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Rate	0.4545	0.1429	0.8333	0.5000	0.6667	0.5000
Priors	0.3056	0.1944	0.1667	0.1667	0.1667	

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.CROPS
Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} (X-\bar{X}_{(X)j}) - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5\ D_j^2(X)) \ / \ \mathrm{SUM}_k \ \exp(-.5\ D_k^2(X))$$

Number of Observations and Percent Classified into Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	4 36.36	3 27.27	1 9.09	0 0.00	3 27.27	11 100.00
Corn	0 0.00	4 57.14	1 14.29	2 28.57	0 0.00	7 100.00
Cotton	3 50.00	0 0.00	0 0.00	2 33.33	1 16.67	6 100.00
Soybeans	0 0.00	1 16.67	1 16.67	3 50.00	1 16.67	6 100.00
Sugarbeets	2 33.33	1 16.67	0 0.00	2 33.33	1 16.67	6 100.00
Total	9 25.00	9 25.00	3 8.33	9 25.00	6 16.67	36 100.00
Priors	0.30556	0.19444	0.16667	0.16667	0.16667	

Error Count Estimates for Crop

	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Rate	0.6364	0.4286	1.0000	0.5000	0.8333	0.6667
Priors	0.3056	0.1944	0.1667	0.1667	0.1667	

The DISCRIM Procedure

Observations	36	DF Total	35
Variables	4	DF Within Classes	31
Classes	5	DF Between Classes	4

Class Level Information

Crop	Variable Name	Frequency	Weight	Proportion	Prior Probability
Clover	Clover	11	11.0000	0.305556	0.305556
Corn	Corn	7	7.0000	0.194444	0.194444
Cotton	Cotton	6	6.0000	0.166667	0.166667
Soybeans	Soybeans	6	6.0000	0.166667	0.166667
Sugarbeets	Sugarbeets	6	6.0000	0.166667	0.166667

Within Covariance Matrix Information

Crop	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
Clover	4	23.64618
Corn	4	11.13472
Cotton	4	13.23569
Soybeans	4	12.45263
Sugarbeets	4	17.76293

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV_j^{-1} \, (\bar{X}_i - \bar{X}_j) + \ln |COV_j| - 2 \ln PRIOR_j$$

Generalized Squared Distance to Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets
Clover	**26.01743**	**1320**	**104.18297**	**194.10546**	**31.40816**
Corn	**27.73809**	**14.40994**	**150.50763**	**38.36252**	**25.55421**
Cotton	**26.38544**	**588.86232**	**16.81921**	**52.03266**	**37.15560**
Soybeans	**27.07134**	**46.42131**	**41.01631**	**16.03615**	**23.15920**
Sugarbeets	**26.80188**	**332.11563**	**43.98280**	**107.95676**	**21.34645**

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.CROPS
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\overline{X}_j)' \, COV_j^{-1} \, (X-\overline{X}_j) + \ln |COV_j| - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	9	0	0	0	2	11
	81.82	0.00	0.00	0.00	18.18	100.00
Corn	0	7	0	0	0	7
	0.00	100.00	0.00	0.00	0.00	100.00
Cotton	0	0	6	0	0	6
	0.00	0.00	100.00	0.00	0.00	100.00
Soybeans	0	0	0	6	0	6
	0.00	0.00	0.00	100.00	0.00	100.00
Sugarbeets	0	0	1	1	4	6
	0.00	0.00	16.67	16.67	66.67	100.00
Total	9	7	7	7	6	36

	25.00	19.44	19.44	19.44	16.67	100.00
Priors	0.30556	0.19444	0.16667	0.16667	0.16667	

Error Count Estimates for Crop

	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Rate	0.1818	0.0000	0.0000	0.0000	0.3333	0.1111
Priors	0.3056	0.1944	0.1667	0.1667	0.1667	

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.CROPS
Cross-validation Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\overline{X}_{(X)j})' \, COV_{(X)j}^{-1} \, (X-\overline{X}_{(X)j}) + \ln |COV_{(X)j}| - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	9	0	0	0	2	11

	81.82	0.00	0.00	0.00	18.18	100.00
Corn	3 42.86	2 28.57	0 0.00	0 0.00	2 28.57	7 100.00
Cotton	3 50.00	0 0.00	2 33.33	0 0.00	1 16.67	6 100.00
Soybeans	3 50.00	0 0.00	0 0.00	2 33.33	1 16.67	6 100.00
Sugarbeets	3 50.00	0 0.00	1 16.67	1 16.67	1 16.67	6 100.00
Total	21 58.33	2 5.56	3 8.33	3 8.33	7 19.44	36 100.00
Priors	0.30556	0.19444	0.16667	0.16667	0.16667	

Error Count Estimates for Crop

	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Rate	0.1818	0.7143	0.6667	0.6667	0.8333	0.5556
Priors	0.3056	0.1944	0.1667	0.1667	0.1667	

The DISCRIM Procedure
Classification Results for Test Data: WORK.TEST
Classification Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Posterior Probability of Membership in Crop

xvalues	From Crop	Classified into Crop		Clover	Corn	Cotton	Soybeans	Sugarbeets
16 27 31	Corn	Corn		0.0894	0.4054	0.1763	0.2392	0.0897
29 24 26	Cotton	Soybeans	*	0.2521	0.1842	0.1529	0.2549	0.1559
32 32 62	Clover	Cotton	*	0.2163	0.3180	0.3327	0.1125	0.0206

* Misclassified observation

The DISCRIM Procedure
Classification Summary for Test Data: WORK.TEST
Classification Summary using Linear Discriminant Function

Generalized Squared Distance Function

2 _ -1 _

$$D_j(X) = (X-X_j)' \, COV \, (X-X_j)$$

Posterior Probability of Membership in Each Crop

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into Crop

From Crop	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	0 0.00	0 0.00	1 100.00	0 0.00	0 0.00	1 100.00
Corn	0 0.00	1 100.00	0 0.00	0 0.00	0 0.00	1 100.00
Cotton	0 0.00	0 0.00	0 0.00	1 100.00	0 0.00	1 100.00
Total	0 0.00	1 33.33	1 33.33	1 33.33	0 0.00	3 100.00
Priors	0.30556	0.19444	0.16667	0.16667	0.16667	

Error Count Estimates for Crop

	Clover	Corn	Cotton	Total
Rate	**1.0000**	**0.0000**	**1.0000**	**0.7083**
Priors	**0.3056**	**0.1944**	**0.1667**	**0.6667**

Obs	Crop	x1	x2	x3	x4	xvalues	Clover	Corn	Cotton	Soybeans	Sugarbeets	_INTO_
1	Corn	**16**	**27**	**31**	**33**	**16 27 31**	**0.08935**	**0.40543**	**0.17632**	**0.23918**	**0.08972**	Corn
2	Soybean	.	**21**	**25**	**23**	s **21 25 23**	.	.	.	.	.	
3	Cotton	**29**	**24**	**26**	**28**	**29 24 26**	**0.25213**	**0.18420**	**0.15294**	**0.25486**	**0.15588**	Soy
4	Sugarbe	.	**23**	**21**	**54**	ets54 **23 21**	.	.	.	.	.	
5	Clover	**32**	**32**	**62**	**16**	**32 32 62**	**0.21633**	**0.31799**	**0.33266**	**0.11246**	**0.02056**	Cot

BIBLIOGRAFIA

FISHER, R.A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v.**7**, p.179-188, 1936.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, 815 p.

KHATTREE, R. & NAIK, D.N**. Multivariate data reduction and discrimination with SAS software**. Cary, NC, USA: SAS Institute Inc., 2000. 558 p.

KHOURY JR, J.K. **Desenvolvimento e avaliação de um sistema de visão artificial para classificação de madeira serrada de eucalipto**. 2004. 101 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2004.

REGAZZI, A.J. Análise multivariada, notas de aula INF 766, Departamento de Informática da Universidade Federal de Viçosa, v.2, 2000.

VARELLA, C.A.A. **Estimativa da produtividade e do estresse nutricional da cultura do milho usando imagens** digitais. 2004. 92 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Viçosa, Viçosa, 2004.

SAS. Online doc version 8. Disponível em: http://v8doc.sas.com/sashtml/. Acesso em 14 mar. 2007.