UNIVERSIDADE ESTADUAL DA PARAÍBA

CENTRO DE CIÊNCIAS E TECNOLOGIA

DEPARTAMENTO DE ESTATÍSTICA

DIOGO MEDEIROS NÓBREGA

ANÁLISE DISCRIMINANTE UTILIZANDO O *SOFTWARE* SPSS

Campina Grande – PB

2010

DIOGO MEDEIROS NÓBREGA

ANÁLISE DISCRIMINANTE UTILIZANDO O *SOFTWARE* SPSS

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Estatística, do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Edwirde Luiz Silva

Campina Grande – PB

2010

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

N337a Nóbrega, Diogo Medeiros.
Análise discriminante utilizando o software SPSS [manuscrito] / Diogo Medeiros Nóbrega. – 2010.
53 f.: il. color.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologias, 2010.
"Orientação: Prof. Dr. Edwirde Luiz Silva, Departamento de Estatística".

1. Estatística. 2. Análise Discriminante. 3. Funções Discriminantes. I. Título.

21. ed. CDD 519.535

DIOGO MEDEIROS NÓBREGA

ANÁLISE DISCRIMINANTE UTILIZANDO O *SOFTWARE* SPSS

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Estatística, do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Aprovada em 16 / 12 / 2010 .

BANCA EXAMINADORA

Prof. Dr. Edwirde Luiz Silva / UEPB
Orientador

Prof. Dr. Gustavo Henrique Esteves / UEPB
Examinador

Ruth Silveira do Nascimento

Prof. ª Msc. Ruth Silveira do Nascimento / UEPB
Examinadora

Agradecimentos

Considerando esta monografia como resultado de uma caminhada que não começou hoje, agradecer pode não ser tarefa fácil, nem justa. Para não correr o risco da injustiça, agradeço de antemão a todos que de alguma forma passaram pela minha vida e contribuíram para a construção de quem sou hoje.

E agradeço, particularmente, a algumas pessoas pela contribuição direta na construção deste trabalho:

À Deus, todo-poderoso, pelo dom de vida que me concedeu, por ter iluminado o meu caminho durante todos estes anos me dando forças para concluir essa preciosa etapa, por ter me oferecido a oportunidade de evoluir, crescer e conhecer todas as pessoas que eu conheci.

À minha Mãe, que se sacrificou mais do que eu nestes anos para que hoje eu estivesse onde estou. Por todo o suor derramado para que não me faltasse nada, pelo amor, carinho e por ter me apoiado incondicionalmente, acreditando em mim quando nem eu mesmo acreditava.

Ao meu Pai, por todo amor, carinho, apoio e por todos os sacrifícios feitos ao longo desta caminhada.

À minha Tia Jacira, por ter me acolhido na sua casa quando eu mais precisava, não medindo esforços para me ajudar.

Ao Professor e Orientador Edwirde Luiz por seu apoio, pela paciência na orientação e incentivo que tornaram possível a conclusão deste trabalho.

Ao Professor Juarez Fernandes, por sempre me incentivar a pensar e criar um senso crítico e por aceitar ser meu Orientador de Estágio, contribuindo assim, para meu amadurecimento profissional e pessoal.

À todos os meus colegas de classe, por compartilharem seus conhecimentos comigo e me ajudarem nas horas difíceis.

Aos meus amigos Fabrício Olegário, Tasso Roberto, Fabrycio Érico e Weverson Menezes, que dividiram apartamento comigo durante esses cinco anos de faculdade, sempre se mostrando verdadeiros amigos.

Às minhas Amigas Ariadne, Elziete e Socorro Gomes que me receberam tão bem na FIEP e por contribuírem diretamente para o meu amadurecimento pessoal e profissional.

E a todos que contribuíram de maneira direta ou indireta para a realização desse trabalho.

“A ferrovia que leva ao sucesso é construída em cima de um solo de humildade com pesados trilhos chamados erros que somente são fixados numa linha reta com maciços pregos de perseverança.”

Eduardo Siqueira Filho

Resumo

Este estudo tem como objetivo enunciar as técnicas e pressupostos envolvidos na Análise Discriminante (AD). A Análise de discriminante é uma técnica para redução de dimensão, relacionada à análise de componentes principais e correlação canônica. A partir de uma ou várias variáveis quantitativas e uma variável qualitativa (dependente), o pesquisador tem a possibilidade de elaborar previsões a respeito de a qual grupo certa observação (por exemplo, um produto, uma pessoa ou uma empresa) pertencerá, uma vez que se caracteriza como uma técnica de previsão de classificação. Para alcançar este objetivo, a análise discriminante gera funções discriminantes (combinações lineares das variáveis) que ampliam a discriminação dos grupos descritos pelas categorias de determinada variável dependente. Serão apresentadas a estimação da regra de classificação, a avaliação da qualidade de ajuste da regra de discriminação, estimação da probabilidade global de acerto e os testes relacionados com os pressupostos da análise discriminante. A presente pesquisa apresenta como aplicação da Análise Discriminante um exemplo destinado, exclusivamente, a fins didáticos, sendo os dados meramente ilustrativos. Consideraremos uma tabela com dados fictícios, que contém quatro grupos (1, 2, 3 e 4) e as variáveis X_1, X_2, X_3 e X_4, em um total de 40 observações. Na fase de aplicação, faremos uso do *software Statical Package for the Social Sciences* (SPSS® V. 17) para a elaboração da Análise Discriminante. Temos como objetivo que as variáveis explicativas X_1, X_2, X_3 e X_4 formem a função discriminante, a fim de que os casos sejam distribuídos nos grupos corretos.

Palavras-chave: Análise Discriminante, Regra de Classificação, Pressupostos, Funções Discriminantes.

Abstract

This Work of Conclusion of Course (TCC) has as objective to enunciate the involved estimated techniques and purposes in the Discriminant Analysis (DA). The Analysis of discriminant is one technique for reduction of dimension, related to the analysis of main components and canonic correlation. From one or some quantitatives variables and a qualitative variables (dependent variable), the researcher has the possibility to elaborate forecasts regarding which certain group comment (for example, a product, a person or a company) will belong, a time that if characterizes as one technique of classification forecast. To reach this objective, the discriminant analysis generates discriminants functions (linear combinations of the one variable) that they extend the discrimination of the described groups for the categories of determined changeable dependent. The esteem of the classification rule, the evaluation of the quality of adjustment of the discrimination rule, esteem of the global probability of rightness and the tests related with the estimated ones of the discriminant analysis will be presented. The present research presents as application of the Discriminant Analysis a destined example, exclusively, the didactic ends, being the illustrative data mere. We will consider a table with fictitious data that contains four groups (1, 2, 3 and 4) and the variables X1, X2, X3 and X4, in a total of 40 observations. In the application phase, we will make use of software Statical Package of the Social Sciences (SPSS® V. 17) for the elaboration of the Discriminant Analysis. We have as objective that the explicative variables X1, X2, X3 and X4 form the discriminant function, so that the cases are distributed in the correct groups.

Key words: Discriminant Analysis, classification rule, purposes, Discriminant Function

Lista De Figuras

Lista De Tabelas

SUMÁRIO

1 INTRODUÇÃO

A análise discriminante (AD) é uma “ferramenta” estatística utilizada para classificação de um determinado elemento num certo grupo de variáveis.

O objetivo do presente trabalho é demonstrar a técnica, partindo desde a verificação dos pressupostos envolvidos na análise discriminante que vão dizer se é possível aplicar a análise discriminante, iremos também medir o poder de discriminação de cada variável ou grupo de variáveis, descrever as diferenças entre os grupos, desenvolver regras para classificar novos elementos e analisar a probabilidade de erro, bem como a probabilidade global de acerto.

A AD oferece ao pesquisador a possibilidade de elaborar previsões a respeito de a qual grupo certa observação (por exemplo, um produto, uma pessoa ou uma empresa) pertencerá, uma vez que se caracteriza como uma técnica de previsão de classificação. Para alcançar este objetivo, a análise discriminante gera funções discriminantes (combinações lineares das variáveis) que ampliam a discriminação dos grupos descritos pelas categorias de determinada variável dependente. No processo de classificação consideram-se os custos decorrentes de eventuais erros de classificação, bem como as probabilidades “a priori” de que o elemento pertença a cada um dos grupos.

Na análise discriminante, a comparação do elemento amostral em relação aos grupos candidatos é, em geral, feita através da construção de uma regra matemática e de classificação, ou discriminação fundamentada na teoria das probabilidades. Para cada novo elemento amostral, a regra de classificação permitirá ao pesquisador decidir qual é a população mais provável de ter gerado seus valores numéricos nas p-características avaliadas.

O presente trabalho abordará de forma bastante prática, o uso desta técnica denominada análise discriminante, no qual utilizaremos o *software SPSS* para aplicação da técnica e dos testes necessários para sabermos se é viável, ou não, o uso da AD no nosso exemplo, partindo dos pressupostos existentes por trás da análise discriminante.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Uma Introdução à Análise Discriminante

A análise discriminante foi proposta na primeira metade do século XX por Fisher, para servir como critério mais confiável e para a classificação de novas espécies de vegetais de acordo com as características de homéricas, sendo rapidamente adotada além da taxonomia e sistemática vegetal (MAROCO, 2007). Percebe-se uma grande aplicação da AD em diversos campos do conhecimento, como biologia, antropologia, marketing, comportamento do consumidor, entre outros.

Durante muitos anos, a análise discriminante tem recebido uma grande atenção teórica de diversas áreas, como marketing, em que podem ser citados trabalhos de Frank, Massy e Morrison (1965), Crask e Perreault (1977) e Hora e Wilcox (1982). Merece destaque também a expressiva contribuição de alguns trabalhos em relação a modelagem matemática da análise discriminante, como os de Lachenbruch e Mickey (1986), Krzanowski (1975), Randles, Broffit, Ramberg e Hogg (1978), Constanza e Afifi (1979) e Fraley e Raftery (2002).

Nos estados unidos e na Europa, há, atualmente, uma vasta aplicabilidade da análise discriminante nas ciências sociais e do comportamento e, no Brasil, seu uso vem sendo ampliado em diversas áreas, devido à contribuição direta dos principais *softwares* estatísticos que apresentam esta técnica.

A análise discriminante envolve a relação entre o conjunto de variáveis independentes quantitativas e uma variável dependente qualitativa. Em muitos casos, verifica-se mais de três classificações para a variável dependente (neste caso, multicotômica), como, por exemplo, classificações de um tipo alto, médio e baixo e o bom pagador sem restrição ao crédito, bom pagador com restrição ao crédito e mau pagador com total restrição ao crédito.

Quando o pesquisador estiver interessado na discussão de somente dois grupos de variáveis dependentes, a técnica é chamada de análise discriminante simples. No entanto, o em muitos casos, há o interesse na descriminação entre mais de dois grupos, e sendo a técnica, assim, denominada de análise discriminante múltipla (a partir de agora, chamaremos de MDA – Multiple-group discriminant analysis).

Os objetivos principais desses dois tipos de análises são parecidos: (i) identificar as variáveis que melhor discriminam dois ou mais grupos; (ii) utilizar estas variáveis para desenvolver funções discriminante que representam as diferenças entre os grupos; (iii) fazer o uso das funções discriminantes para o desenvolvimento de regras de classificação de futuras observações nos grupos.

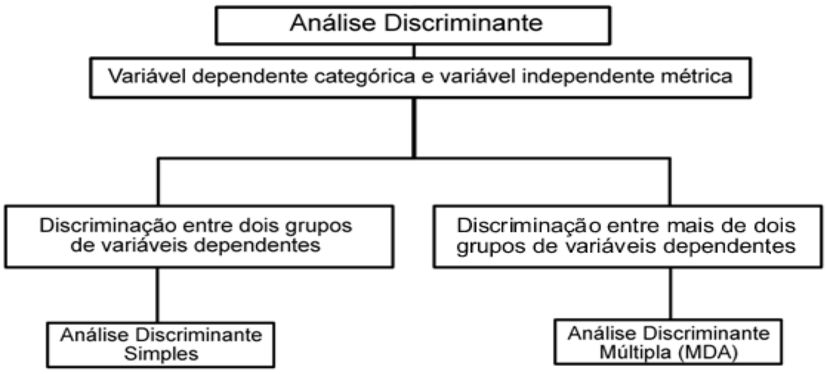
A Análise Discriminante Simples requer somente uma função de discriminação para representar todas as diferenças entre os dois grupos. Por outro lado, a MDA oferece *outputs* que conseguem explicitar as diferenças entre mais de dois grupos. Sendo assim, a MDA também tem como objetivo adicional: (iv) identificar o número mínimo de funções discriminantes que melhor proporcionam as diferenças entre os grupos (SHARMA, 1996). Merece atenção o fato de a MDA demandar que a quantidade de variáveis independentes seja maior, ou igual, ao número de grupos (categorias) da variável dependente em estudo. O pesquisador ainda perceberá que a quantidade de funções discriminantes criadas para a MDA representa o número de grupos (categorias da variável dependente) menos um.

$$\text{N° de funções discriminantes} = (g - 1)$$

Em que g representa a quantidade de categorias (grupos) da variável dependente.

A Figura 1 apresenta os tipos de análise discriminante em função da quantidade de categorias da variável dependente.

Figura 1: Tipos de análise discriminante



2.2 Análise Discriminante

A análise discriminante é uma técnica multivariada utilizada quando a variável dependente é categórica, ou seja, qualitativo (não métrica) e as variáveis independentes são quantitativas (métricas).

Definida esta proposição, um pesquisador pode estar interessada no esclarecimento das correlações que motivam as características de cada categoria na qual o objeto está posicionado, tal como estudo de uma companhia de seguros que apresentam o intuito de prever quais clientes estão mais predispostos à falência. Outro pesquisador pode desejar avaliar a percepção do departamento de vendas de uma empresa sobre o fato de um novo produto está destinada ao sucesso ou fracasso quando do seu orçamento. Um terceiro pode ainda investigar as razões por meio das quais a liberação de crédito por uma empresa do setor financeiro e em definir critérios para a estratificação de pessoas em grupos com características de bons pagadores e com acesso total ao crédito, bons pagadores com restrição ao crédito e maus pagadores e sem direito ao crédito.

Em resumo, a análise discriminante ajuda o pesquisador que deseje criar um estudo com grupos diferenciados por meio das variáveis independentes que possui.

A análise discriminante é uma técnica que pode ser utilizada para a classificação de elementos de uma amostra ou população. Para a sua aplicação, é necessário que os grupos para os quais cada elemento amostral pode ser classificado sejam predefinidos, ou seja, conhecidos a *priori* considerando-se suas características gerais. Este conhecimento permite que a elaboração de uma função matemática chamada de regra de classificação ou discriminação, que é utilizada para classificar novos elementos amostrais nos grupos já existentes. Portanto, o número de grupos é conhecido a *priori*, assim como nos métodos não-hierárquicos de agrupamentos, mas a regra de classificação é elaborada utilizando-se procedimentos que, em geral, vão além do uso de distâncias matemáticas.

Para o entendimento da técnica, suponha que se tenha n_1 elementos amostrais procedentes, com probabilidade 1, da população A e n_2 elementos amostrais procedentes,com probabilidade 1, da população B, e que em cada um dos $n_1 + n_2 = n$ elementos amostrais tenham sido medidas p-variáveis aleatórias (características). Uma análise estatística do comportamento das p-características medidas permite identificar o perfil geral de cada grupo. Deste modo, se houver um novo elemento amostral, não pertencente a nenhuma das duas amostras anteriores, e cuja origem é incerta, seria possível compará-lo de algum modo com o perfil geral dos grupos A e B e classificá-los como pertencente ao grupo cujo perfil geral fosse mais semelhante ao dele.

É evidente que todo o processo de tomada de decisões traz consigo um possível erro de decisão. O objetivo, portanto, é o de construir uma regra de classificação que minimize o número de classificações incorretas, ou seja, o erro de dizer que um elemento amostral pertence a uma população quando, na realidade, ele pertence à outra. Além disso, é importante construir uma regra que minimize o custo de classificação incorreta.

Um dos exemplos mais simples da análise discriminante é o teste de hipótese, como o que é feito rotineiramente para a média de uma população. Neste caso, observa-se uma amostra e, a partir dos resultados amostrais observados, calcula-se a estatística do teste, que é uma regra de discriminação entre a hipótese nula e a hipótese alternativa. Dependendo do valor assumido pela estatística de teste, decide-se pela veracidade, ou não, da hipótese nula. A constante que delimita a região de rejeição da hipótese nula é determinada pela fixação do nível de significância do teste, isto é, a probabilidade de rejeitarmos a hipótese nula a favor da alternativa, quando a hipótese nula é verdadeira. Além disso, busca-se o teste que minimiza a chance de não se rejeitar a hipótese nula quando esta deve, de fato, ser rejeitada, ou o chamado erro do tipo 2. Nas sessões a seguir, serão introduzidos matematicamente os fundamentos da análise discriminante para duas ou mais populações.

2.3 Modelagem da Análise Discriminante

Antes de iniciarmos a modelagem da análise discriminante propriamente dita, é pertinente esclarecemos os pressupostos inerentes a esta técnica. Os *softwares* estatísticos atualmente apresentam os cálculos dos testes referentes a esses pressupostos e, portanto, é interessante que o pesquisador avalie o resultado e compare com os níveis padrão de utilização, para avaliar se a análise discriminante em questão caracteriza-se por estar em um nível confiável e de aplicação, ou seja, se não fere alguns dos critérios estabelecidos por meio de suas premissas. No SPSS, esses testes aparecem antes da projeção da função discriminante, permitindo, assim, essa comparação.

Pode-se dizer que há dois pressupostos principais, referentes à existência de normalidade multivariada das variáveis explicativas e a presença de homogeneidade das matrizes de variância e covariância para os grupos. Em relação a primeira suposição, a combinação linear das variáveis explicativas apresentam uma distribuição normal e, caso ocorra uma violação desse pressuposto, a AD poderá causar distorções na avaliação do pesquisador, principalmente se a amostra que compõe cada grupo for pequena.

No entanto, se essa violação ocorrer somente pela existência de assimetria da distribuição, a decisão sobre a aplicação da técnica não sofrerá alteração. É importante ressaltar que, se a distribuição não for mesocúrtica, a aplicação da AD será prejudicada, sendo pior o caso em que a distribuição multivariada for platicúrtica (SHARMA, 1996).

O segundo pressuposto refere-se à existência de homogeneidade das matrizes de variância e covariância. Esse pressuposto é verificado por meio da estatística Box's M, que pode ser sensível ao tamanho da amostra e o não atendimento da hipótese de distribuição normal multivariada. Felizmente, a análise discriminante é uma técnica bastante robusta à violação destes pressupostos, desde que a dimensão do grupo seja superior ao número de variáveis em estudo e que as médias dos grupos não sejam proporcionais às suas variâncias, ou seja, caso a homogeneidade das matrizes seja violada, haverá um aumento da probabilidade de classificar observações no grupo que possuir maior dispersão.

O teste clássico, o teste da razão de verossimilhança, para testar a igualdade de matrizes de variâncias e covariâncias foi desenvolvido por Box (1950), sendo uma generalização do teste univariado de igualdade de variâncias de Bartlett (1947). As hipóteses a testar são:

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma \\ H_1 : \Sigma_l = \Sigma_k \text{ , } para\ algum\ l \neq k\ (l, k = 1, 2, \dots, g). \end{cases}$$

Sejam n a dimensão total da amostra, $v_k = n_k - 1$ os graus de liberdade associados a cada grupo, S_k a matriz de variâncias e covariâncias do grupo k, k = 1, 2,... , g, e $S = \frac{W}{n-g}$ a matriz de variâncias e covariâncias total, onde g é o número de populações, e W é a matriz de SQ_{PC} (soma de quadrados de produtos cruzados) do resíduo. O teste é definido do seguinte modo:

$$M = (n - g)ln|S| - \sum_{k=1}^{g} v_k\ ln|S| \quad (1.1)$$

Box sugeriu duas aproximações para o teste:

i. A aproximação à χ^2: É indicada quando as dimensões dos grupos são superiores a 20, o número de variáveis e de grupos inferior a seis. Esta aproximação é dada por:

$$MC \sim \chi^2_{\left(\frac{1}{2}p(p+1)(g-1)\right)}$$

onde

$$C = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\left(\sum_{k=1}^{g}\frac{1}{v_k} - \frac{1}{n-g}\right)$$

ii A aproximação à F: Indicada em todas as outras situações. Esta fórmula é dada por:

$$\frac{M(1 - \alpha_1 - \frac{v}{v_0})}{v} \sim F_{(v,v_0)}$$

sendo

$$a_1 = 1 - C\ ;$$

$$a_2 = \frac{(p-1)+(p+2)}{6(g-1)}\left(\sum_{k=1}^{g}\frac{1}{v_k^2} - \frac{1}{(n-g)^2}\right); \quad (1.2)$$

$$v = \frac{p(p+2)(g-1)}{2};$$

$$v_0 = \frac{v+2}{a_2 - a_1^2}.$$

Um nível de significância de 0, 01 ou menos é usado como ajuste para a sensibilidade da estatística. (HAIR JR. et al., 2006, p.275).

Além desses pressupostos é pertinente ressaltar que a inexistência de *outliers*, a presença de linearidade das relações e a ausência de problemas relacionados à multicolinearidade das variáveis explicativas que também são consideradas pressupostos da análise discriminante.

É essencial definimos o tamanho correto da amostra que será estudada, uma vez que esta técnica é muito sensível à proporção do tamanho da amostra em relação ao número de variáveis preditoras (HAIR, ANDERSON, TATHAM E BLACK,

2005) e, portanto, não deve haver uma grande variabilidade de dimensões entre os grupos.

É de fundamental importância que a definição do dimensionamento amostral seja estabelecida de acordo com algum critério e sempre embasada pelos conceitos de amostragem.

Apresentados os pressupostos, podemos começar a expor os passos para a composição das funções discriminantes, lembrando que isso representa um dos principais objetivos da AD, já que, a partir destas funções é que as observações serão discriminadas. Antes desta primeira etapa, lembramos que o pesquisador já deve ter estabelecido seu problema principal da pesquisa, assim como em qualquer aplicação de uma técnica multivariada, e definido as categorias de estudo com, no mínimo, dois grupos.

Tomemos como exemplo, uma loja que deseja elaborar um estudo sobre a concessão do limite de crédito aos clientes (problema principal) quando da elaboração do crediário e, para tanto, deverá discriminá-los em dois grupos, e sendo, respectivamente, o dos inadimplentes e o dos adimplentes (níveis de estudo divididos). Neste caso, estamos tratando de uma AD em nível dicotômico.

Portanto, esta etapa consiste na seleção da variável dependente (categórica) e das variáveis explicativas (métricas). Em outras palavras, a escolha de *n* variáveis discriminantes (explicativas) é feita a partir de um conjunto maior de *p* possíveis variáveis.

A análise discriminante permite um conhecimento das variáveis que mais se destacam na discriminação dos grupos. Para tanto, diversos *outputs* são gerados a partir de testes e estatísticas como o lambda de Wilks, a correlação canônica, o Qui-quadrado e o autovalor

O lambda de Wilks, que varia de 0 a 1, propicia a avaliação da existência de diferenças de médias entre os grupos para cada variável. Valores elevados desta estatística indicam ausência de diferenças entre os grupos, e sua expressão é dada por:

$$\Lambda = \frac{\mathrm{SQ_{dg}}}{\mathrm{SQT}} \quad (1.3)$$

Em que SQ_{dg} representação dos erros (dentro dos grupos) e SQT, a soma dos quadrados total.

Como a distribuição exata de Lambda não é conhecida, utiliza-se, para a existência de dois grupos, a seguinte transformação, que possui uma distribuição F com p e (N – p – 1) graus de liberdade, em que N representa o tamanho da amostra e p o número total de variáveis explicativas.

$$F = \left(\frac{1 - \Lambda}{\Lambda}\right)\left(\frac{\mathrm{N} - p - 1}{\mathrm{p}}\right) \quad (1.4)$$

Ou a seguinte transformação, para a existência de três grupos, com distribuição F com 2*p* e 2(N – *p* – 2) graus de liberdade (MAROCO, 2007):

$$F = \left(\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{\mathrm{N} - p - 2}{\mathrm{p}}\right) \quad (1.5)$$

O valor transformado de lambda de Wilks segue uma exata distribuição F somente em certos casos, apresentados no Quadro 1. Para todos os outros casos, a distribuição do valor transformado de lambda de Wilks pode ser aproximada por uma distribuição F (SHARMA, 1996).

Quadro 1: Situações em que o Lambda de Wilks apresenta distribuição F

Número de variáveis (*p*)	Número de grupos	Transformação	Graus de Liberdade
n	2	$\left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{\mathrm{N}-p-1}{\mathrm{p}}\right)$	$p, \mathrm{N}-p-1$
n	3	$\left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{\mathrm{N}-p-2}{\mathrm{p}}\right)$	$2p, 2(\mathrm{N}-p-1)$
1	Qualquer	$\left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{\mathrm{N}-\mathrm{g}}{\mathrm{g}-1}\right)$	$\mathrm{g}-1, \mathrm{N}-\mathrm{g}$
2	Qualquer	$\left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{\mathrm{N}-\mathrm{g}-1}{\mathrm{g}-1}\right)$	$2(g-1), 2(\mathrm{N}-\mathrm{g}-1)$

Fonte: Sharma (1996)

Nota: g = número de grupos.

Com a seleção das variáveis discriminantes (explicativas) para a formação dos grupos, passamos a identificação das funções discriminantes. Como já previamente discutido, a AD assemelha-se a análise de regressão em termos de objetivos e características e, desta forma, sua função geral pode ser representada por meio da seguinte equação linear:

$$Z_n = \propto + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (1.6)$$

em que:

Z: variável dependente;

α: intercepto;

X_i: variáveis explicativas;

β_i: coeficientes discriminantes para cada variável explicativa.

É importante ressaltar que esta função discriminante (1.6) é diferente da função linear de Fisher, uma vez que, enquanto a primeira é utilizada como um meio de facilitar a interpretação dos parâmetros das variáveis explicativas, a função linear discriminante de Fisher é utilizada para classificar as observações nos grupos.

Na função discriminante linear de Fisher, os valores das variáveis explicativas de uma observação são inseridos nas funções de classificação e, não conseqüentemente, um Escore de classificação é calculado para cada grupo, para aquela observação.

Dadas *p* variáveis e *g* grupos, é possível estabelecer *m = (g – 1; p)* funções discriminantes que são combinações lineares das *p* variáveis, de modo que a função linear de Fisher seja dada por:

$$Z_n = W_1X_1 + W_2X_2 + \cdots + W_nX_n \qquad (1.7)$$

Em que W_i representa o vetor de pesos das variáveis para as funções discriminantes e são estimados de modo que a variabilidade dos escores da função o discriminante seja máxima entre os grupos e mínima dentro dos grupos (MAROCO, 2007). Assim, podemos expressar o *i*-ésimo autovalor (*i*-ésima função discriminante) da seguinte forma:

$$autovalor_i = \frac{SQ_{eg}}{SQ_{dg}} \qquad (1.8)$$

em que:

SQ_{eg}: Soma dos quadrados entre os grupos;

SQ_{dg}: Soma dos quadrados dentro dos grupos.

Assim, *autovalores* altos resultam em boas funções discriminantes.

Expressa de acordo com a Equação (1.7), a função discriminante é conhecida por função discriminante linear de Fisher e, após a dedução da primeira função discriminante, os pesos W_i, das funções seguintes são obtidos sob a restrição adicional de que os escores das funções discriminantes não estejam correlacionados.

Por meio desses cálculos, é possível chegarmos à expressão da primeira função discriminante (Z_i); as outras funções dos outros grupos são encontradas pelo mesmo método. Porém, é preciso que tenhamos atenção em relação a problemas de correlação, ou seja, os escores das outras funções (Z_2, Z_3,..., Z_n) não devem se correlacionados. Assim, a expressão do autovalor da segunda função discriminante linear de Fisher pode ser escrito da seguinte forma:

$$autovalor = \frac{SQ_{eg}\ (Z_2)}{SQ_{dg}\ (Z_2)} \tag{1.9}$$

em que:

$SQ_{eg}\ (Z_2)$: soma dos quadrados entre os grupos na segunda função discriminante;

$SQ_{dg}\ (Z_2)$: soma dos quadrados dentro dos grupos na segunda função discriminante.

Por meio das funções discriminantes de Fisher, é possível estudarmos a influência que determinada variável dependente categórica sofre de um vetor de variáveis explicativas. Ademais, é possível estabelecer uma relação entre os *autovalores* e o lambda de Wilks:

$$\Lambda = \Pi \left[\frac{1}{(1+autovalor_i)}\right] \tag{1.10}$$

Como foi discutido anteriormente, para a existência de *g* grupos, há (g – 1) funções discriminantes. Assim, se tivéssemos, por exemplo, três grupos, seriam geradas duas funções discriminantes, sendo a primeira responsável pela separação de um grupo dos demais e a segunda pela separação dos dois grupos restantes.

Como nem todas as funções discriminantes podem ser estatisticamente significantes, ou seja, algumas podem representar maiores diferenças entre os grupos do que outras, é necessário a seguinte transformação que possui distribuição Qui-quadrado (χ^2) com *p*. (g – 1) graus de liberdade. A expressão a seguir é utilizada, portanto, para permitir o cálculo da estatística Qui-quadrado (χ^2) que avalia a significância estatística global de todas as funções discriminantes (SHARMA, 1996):

$$\chi^2 = -\left[n - \frac{(p+g)}{2} - 1\right].\ln\left(\Lambda_k\right) \tag{1.11}$$

Em que Λ_k é o lambda de Wilks de cada função discriminante e testa a significância das funções discriminantes, ou seja, avalia o quão bem cada função separa as observações em grupos diferentes. Se tivermos, por exemplo, duas funções discriminantes, primeiramente são testadas as duas funções em conjunto e, na seqüência, se as médias dos grupos para a segunda função discriminante são iguais. Nesta etapa, a rejeição da hipótese nula do teste Qui-quadrado (H_o: nenhuma das funções é significante para discriminar os grupos) não significa que tenhamos condições de responder quais funções discriminam significativamente os grupos, ou seja, a rejeição de H_o significa apenas que a primeira função discriminante é significante (as outras podem não ser).

Outra estatística resultante refere-se à correlação Canônica, que corresponde à razão entre a variação entre os grupos e a variação total e mede o grau de associação entre os escores discriminantes e os grupos.
Sua expressão é dada por:

$$\mathrm{C} = \sqrt{\frac{SQ_{eg}}{SQT}} \tag{1.12}$$

que resulta em:

$$\Lambda + C^2 = 1 \tag{1.13}$$

Em termos matriciais SHARMA (1996) apresenta a função discriminante como:

$$\xi = X'\gamma \tag{1.14}$$

em que *X'* (*p* x 1) é a transposta da matriz com *p* variáveis e γ representa o vetor de pesos das variáveis. A soma dos quadrados totais para os escores ξ pode ser definida como ξ' ξ = (X'γ)' (X'γ) = γ' XX' γ, sendo XX'=T a matriz da soma de quadrados e produtos cruzados totais da matriz X com *p* variáveis.

Fazendo T = B + W, em que B e W representam, respectivamente, as matrizes das somas dos quadrados entre os grupos e dentro dos grupos, a soma dos quadrados totais para a função discriminante pode agora ser escrita como ξ'ξ = γ' (B + W)γ = γ' Bγ + γ' Wγ (MAROCO, 2007).

Uma vez que γ' Bγ e γ' Bγ são, respectivamente, a soma dos quadrados entre os grupos e dentro dos grupos para a função ξ, a obtenção da função discriminante resume-se, segundo Maroco (2007), a encontrar o vetor γ, de modo que:

$$\lambda = \frac{\gamma' B \gamma}{\gamma' W \gamma} \tag{1.15}$$

seja máximo.

Sharma (1996) explica que o problema dessa maximização apresenta solução quando:

$$(W^{-1}B - \lambda I) = 0 \tag{1.16}$$

sob a restrição $|W^{-1}B - \lambda I| = 0$.

Este problema tem m = min(κ-1; *p*) soluções correspondentes aos *autovalores* da matriz $W^{-1}B$. O maior *autovalor* (λ_1) apresenta um vetor próprio (autovetor) correspondente à primeira função discriminante, o segundo maior *autovalor* (λ_2) apresenta um vetor próprio correspondente à segunda função, e assim sucessivamente, sob a restrição de que os escores das funções não estejam correlacionados (STEVENS, 2002).

Apresentamos agora as hipóteses nulas e alternativas que se referem às médias populacionais dos grupos em análise para as variáveis explicativas. Para tanto, utilizamos a estatística da distribuição F para testar as seguintes hipóteses:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_n \\ H_1: \mu_1 \neq \mu_2 \neq \cdots \neq \mu_n \end{cases}$$

em que $\mu_1, \mu_2, \ldots, \mu_n$ são as médias populacionais dos grupos 1, 2, ..., *n*, respectivamente. A hipótese nula será rejeitada se pelo menos as médias de dois grupos forem significativamente diferentes (SHARMA, 1996). Desta forma, a hipótese alternativa, quando não rejeitadas, indica que as variáveis explicativas apresentam médias diferentes entre os grupos.

Após a função discriminante ser definida, será calculado o escore discriminante da variável dependente (Z) para cada observação, ou seja, os escores serão calculados de maneira a propiciar a definição do escore crítico que determinará a forma por meio da qual iremos classificar uma observação em determinado grupo. Para grupos de mesma dimensão amostral, (tamanho), o cálculo do escore de corte (*cutoffvalue*) é:

$$f = \frac{\overline{d_1} + \overline{d_2}}{2} \tag{1.17}$$

Em que $\overline{d_1}$ e $\overline{d_2}$ representam as médias das funções discriminantes (centróides) nos grupos 1 e 2, respectivamente. Já para grupos com tamanhos diferentes, temos:

$$f = \frac{n_1\overline{d_1}+n_2\overline{d_2}}{n_1+n_2} \tag{1.18}$$

Em que $n_1 \; e \; n_2$ são os tamanhos dos grupos 1 e 2, respectivamente. Normalmente, o valor de corte é aquele que minimiza o número de classificações incorretas (SHARMA, 1996). De acordo com Maroco (2007), um caso é classificado no grupo 1 se seu escore na função discriminante for maior que *f*. No caso de mais de dois grupos, pode-se definir a zona de fronteira para cada par de grupos, conhecida por Mapa territorial. Como os centróides representam as médias das funções discriminantes em cada grupo, temos, em determinado mapa territorial, que:

2 grupos = 2 centróides

3 grupos = 3 centróides

4 grupos = 4 centróides

...

O método do cálculo do escore crítico é um dos procedimentos existentes para a classificação de futuras observações. Outros métodos também podem ser mencionados, como o da Teoria da Decisão Estatística, o da Função de Classificação e o D^2 de Mahalanobis.

Maroco (2007) descreve que a distância de cada escore ao centróide de um grupo pode ser calculada da seguinte forma:

$$D_j^2 = \left(d - \bar{d}_J\right)S_j^{-1}\left(d - \bar{d}_J\right)' \tag{1.19}$$

em que S_j^{-1} representa a matriz de variância e covariância para as funções discriminantes no grupo g e $\overline{d_j}$ o centróide deste grupo. Se uma observação pertence ao grupo *j* (*j* = 1,..., *g)* então D_g^2 possui uma distribuição Qui-quadrado com *m* (número de funções discriminantes) graus de liberdade.

2.4 Construção da Regra de Classificação

Vamos considerar o caso em que dispomos de apenas duas populações e um conjunto de observações independentes de cada população. Se a distribuição de probabilidades das características medidas dos elementos da amostra que cada população for conhecida, será possível utilizar o princípio da máxima verossimilhança (Casella; Berger, 2002) para construir uma regra de classificação que minimiza a chance de se classificar um elemento amostral incorretamente. Suponha que se saiba que para a população 1 a variável *X* tem uma distribuição normal com média μ_1 e que para a população 2, tem uma distribuição normal com média μ_2, sendo as variâncias nos dois grupos iguais a σ^2. Neste caso, para cada possível valor *x*, é possível calcular um a razão entre as duas distribuições de probabilidades, chamada de razão de verossimilhança entre as duas populações, e definida por:

$$\lambda(x) = \frac{função\ densidade\ de\ x\ na\ população\ 1}{função\ densidade\ de\ x\ na\ população\ 2} = \frac{f_1(x)}{f_2(x)}$$

que, no caso da distribuição normal, torna-se:

$$\lambda(x) = \frac{\frac{1}{\sqrt{2\pi\sigma}} exp\left\{\frac{-1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right\}}{\frac{1}{\sqrt{2\pi\sigma}} exp\left\{\frac{-1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right\}} = exp\left\{\frac{-1}{2}\left[\left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2\right]\right\} \quad (1.20)$$

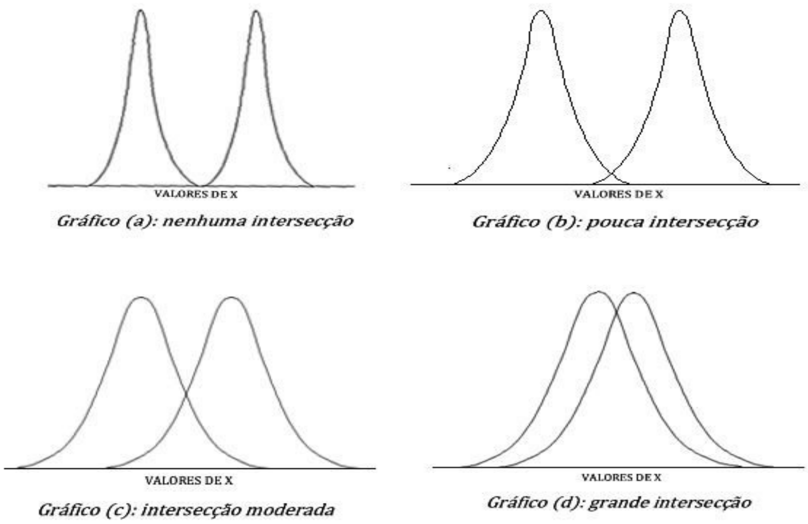
A situação de classificação em duas populações normais, quando estas têm a mesma variabilidade, está representada na equação acima (1.20). Observando-se os gráficos (a)-(d) (Figura 2), percebe-se que, se o valor x observado estiver mais próximo do valor μ_1, o elemento amostral será classificado como pertencente à população 1, enquanto que, se o valor de x estiver mais próxima do valor μ_2, ele será classificado como sendo pertencente a população 2. A qualidade da discriminação dependerá do grau de intersecção entre as duas distribuições de probabilidade. Se a variável discriminante é tal que as duas distribuições são bem separadas no espaço (Gráfico (a)) (Figura 2),, o número de classificações incorretas é aproximadamente zero. Para uma pequena intersecção (Gráfico (b)) (Figura 2),, haverá um pequeno número de erros de classificação. Por outro lado, se a área de intersecção entre as curvas é muito acentuada (Gráficos (c) e (d)) (Figura 2),, o número de erros tenderá a aumentar, chegando a valores que impossibilitam o uso da função discriminante como regra de classificação. Matematicamente, isto é decorrente dos seguintes fatos: voltando na Equação (1.20), tomando-se o logaritmo neperiano de $\lambda(x)$ e multiplicando-o por – 2, obtém-se a fórmula dada em (1.21):

$$-2ln\big(\lambda(x)\big) = \left(\frac{x-\mu_1}{\sigma}\right)^2 - \left(\frac{x-\mu_2}{\sigma}\right)^2 = \frac{1}{\sigma^2}\left[(x-\mu_1)^2 - (x-\mu_2)^2\right] \qquad (1.21)$$

que está relacionada com a diferença das distâncias euclidianas ponderadas ao quadrado de x a μ_1, sendo um fator de ponderação e igual ao inverso da variância σ^2. Quando $\lambda(x) > 1$, tem-se $-2ln\big(\lambda(x)\big) < 0$, e, neste caso, x está mais próximo de μ_1 do que de μ_2. Quando $\lambda(x) < 1$ tem-se $-2ln\big(\lambda(x)\big) > 0$ e, neste caso, x está mais próximo de μ_2 do que de μ_1. No entanto, quando $\lambda(x) = 1$, tem-se $-2ln\big(\lambda(x)\big) = 0$ e x está igualmente próximo de μ_1 e μ_2. Deste modo, a regra de classificação escrita em função de $\lambda(x)$ pode ser reescrita em função de $-2ln\big(\lambda(x)\big)$ da seguinte forma: classifique o elemento amostral na população 1 se $-2ln\big(\lambda(x)\big)$ for menor que zero e na população 2 se $-2ln\big(\lambda(x)\big)$ for maior que zero. Se $-2ln\big(\lambda(x)\big)$ for igual a zero, o elemento amostral poderá ser classificado tanto como sendo da população 1 quanto da população 2.

A função $\lambda(x)$, ou equivalentemente $-2ln\big(\lambda(x)\big)$, é chamada de função discriminante. (Pelo Método da Máxima Verossimilhança)

Figura 2: Ilustração da classificação de duas populações normais com mesma variabilidade e uma variável discriminante.



Nem sempre as populações em consideração apresentam variabilidades semelhantes, sendo importante, portanto, analisar os casos em que as variabilidades se diferem. Sejam var(X) = σ_1^2 na população 1 e var(X) = σ_2^2 na população 2. Neste caso, para um valor x fixo, a razão entre as duas funções de densidades normais será dada por:

$$\lambda(x) = \frac{\frac{1}{\sqrt{2\pi}\ \sigma_1} exp\left\{\frac{-1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\}}{\frac{1}{\sqrt{2\pi}\ \sigma_2} exp\left\{\frac{-1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\}} = \frac{\sigma_1}{\sigma_2} exp\left\{\frac{-1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\} \quad (1.22)$$

Se for considerado $-2l\,n(\lambda(x))$, tem-se:

$$-2l\,n(\lambda(x)) = -2ln\left(\frac{\sigma_1}{\sigma_2}\right) + \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \left(\frac{x-\mu_2}{\sigma_2}\right)^2\right] \quad (1.23)$$

e novamente classifica-se o elemento amostral como sendo da população 1, se $-2l\,n(\lambda(x))$ for menor que zero; como sendo da população 2, se $-2l\,n(\lambda(x))$ for maior que zero, e de qualquer uma das duas populações se $-2l\,n(\lambda(x))$ for igual a zero.

A regra apresentada até o momento para duas populações pode ser estendida para o caso em que se tem p>1 variáveis medidas em cada elemento amostral de cada população e os dados são provenientes de distribuições normais p-variadas. Suponha que, para a população 1, o vetor **X** seja normal com vetor de médias $\boldsymbol{\mu_1}$ e matriz de covariâncias $\sum_1$, e que para a população 2, X seja normal, com vetor de médias $\boldsymbol{\mu_2}$ e matriz de covariâncias $\sum_2$, onde $\boldsymbol{X = [X_1\ X_2\ ...\ X_p]'}$, a razão entre as funções densidade para duas populações em termos de logaritmo neperiano, será:

$$-2l\,n\big(\lambda(x)\big) = -2ln\left\{\frac{(2\pi)^{\frac{p}{2}}(|\sum_1|^{\frac{1}{2}})^{-1}}{(2\pi)^{\frac{p}{2}}(|\sum_2|^{\frac{1}{2}})^{-1}}\left[\frac{exp\left\{-\frac{1}{2}(x-\mu_1)'\sum_1^{-1}(x-\mu_1)\right\}}{exp\left\{-\frac{1}{2}(x-\mu_2)'\sum_2^{-1}(x-\mu_2)\right\}}\right]\right\} =$$

$$\left[(x-\mu_1)'\textstyle\sum_1^{-1}(x-\mu_1)\right] - \left[(x-\mu_2)'\textstyle\sum_2^{-1}(x-\mu_2)\right] + [\ln|\textstyle\sum_1| - \ln|\textstyle\sum_2|] \quad (1.24)$$

Assim, um elemento amostral com vetor de observações x será classificado como pertencente à população 1, se $-2l\,n\big(\lambda(x)\big)$ em (1.24) for menor que zero, e será classificado como pertencente à população 2, se $-2l\,n\big(\lambda(x)\big)$ for maior que zero. Quando $-2l\,n\big(\lambda(x)\big)$ for igual a zero, o elemento amostral poderá ser classificado em qualquer uma das populações. É interessante observar que a função discriminante em (1.23) depende das distâncias de Mahalanobis (1936) do vetor x aos vetores de médias μ_1 e μ_2, além de um fator de correção relacionando os determinantes das matrizes $\sum_1$ e $\sum_2$. A função em (1.25) é chamada de função discriminante quadrática. Quando as matrizes $\sum_1$ e $\sum_2$ são idênticas, isto é, $\sum_1 = \sum_2 = \sum$, a função $-2l\,n\big(\lambda(x)\big)$ simplifica para:

$$-2l\,n\big(\lambda(x)\big) = (x-\mu_1)'\textstyle\sum^{-1}(x-\mu_1) - (x-\mu_2)'\textstyle\sum^{-1}(x-\mu_2) \quad (1.25)$$

Que é equivalente à função conhecida como "função discriminante de Fisher" (1936) expressa como:

$$fd(x) = (\mu_1-\mu_2)'\textstyle\sum^{-1}x - \frac{1}{2}(\mu_1-\mu_2)'\textstyle\sum^{-1}(\mu_1+\mu_2) \quad (1.26)$$

Utilizando-se a fórmula (1.26), o elemento amostral que tenha um vetor de observações igual a $\boldsymbol{x}$ é classificado como pertencente à população 1 se $fd(x)$ for maior que zero ou equivalente se:

$$(\mu_1-\mu_2)'\textstyle\sum^{-1}x > \frac{1}{2}(\mu_1-\mu_2)'\textstyle\sum^{-1}(\mu_1+\mu_2) \quad (1.27)$$

e será classificado como sendo da população 2 se:

$$(\mu_1 - \mu_2)'\Sigma^{-1}x < \tfrac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \qquad (1.28)$$

É interessante observar que $(\mu_1 - \mu_2)'\Sigma^{-1}x = b'x$, onde $b' = (\mu_1 - \mu_2)'\Sigma^{-1}$, é um vetor de dimensão *1×p*. Deste modo, a função discriminante de Fisher tem a forma:

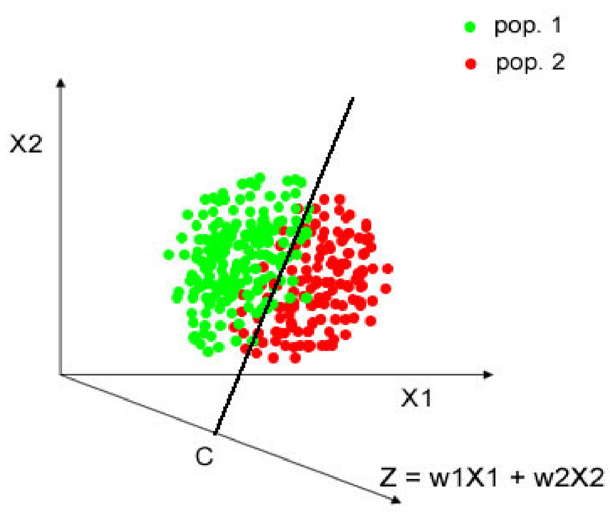
$$(\mu_1 - \mu_2)'\Sigma^{-1}x = b'x = b_1X_1 + b_2X_2 + \cdots + b_pX_p \qquad (1.29)$$

Que é uma combinação linear das variáveis originais. Dependendo do valor numérico desta combinação linear, o elemento amostral é classificado em uma ou outra população. Por sua vez, a constante (C = 1/2) que delimita a região de classificação é uma combinação linear dos seus vetores de médias das duas populações, isto é:

$$\frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) = b'\frac{(\mu_1+\mu_2)}{2} \qquad (1.30)$$

Os valores dos coeficientes do vetor b' são provenientes da matriz $(\mu_{1-}\mu_2)'\Sigma^{-1}$. A Figura 3 apresenta uma ilustração da discriminação de duas populações normais quando p=2 e a função discriminante de Fisher são utilizadas.

Figura 3: Discriminação de duas populações normais. Função discriminante de Fisher – p=2.



A regra de discriminação construída apenas pelo princípio da comparação entre as funções de densidades das duas populações, chamado de principio da máxima verossimilhança, minimiza as probabilidades dos erros decorrentes de classificações incorretas. No entanto, não leva em consideração possíveis diferenças entre os custos associados aos erros de classificação.

2.4.1 Caso De Duas Populações: Estimação Da Regra De Classificação

Na prática, não se conhece os valores verdadeiros dos vetores de médias μ_1 e μ_2 e nem das matrizes de covariâncias $\sum_1$ e $\sum_2$. Portanto, é necessário estimá-los obtendo-se, então, uma estimativa para a função discriminante $\lambda(x)$. Suponha que se tenha uma amostra aleatória de n_1 elementos de população 1 e uma amostra aleatória de n_2 elementos da população 2, sendo as duas amostras independentes entre si. Neste caso, os vetores μ_1 e μ_2 serão estimados pelos vetores de médias amostrais obtidos das amostras 1 e 2, respectivamente, isto é, $\bar{x}_1$ e $\bar{x}_2$. As matrizes $\sum_1$ e $\sum_2$ são estimadas pelas matrizes de covariâncias amostrais de cada grupo separadamente, ou seja, S_1 *e* S_2. A função discriminante estimada, de acordo com a Equação (1.24), será dada por:

$$-2l\,n\hat{\lambda}(x) = (x-\bar{x}_1)'S_1^{-1}(x-\bar{x}_1) - (x-\bar{x}_2)'S_2^{-1}(x-\bar{x}_2) + [ln|\textstyle\sum_1| - ln|\textstyle\sum_2|] \quad (1.31)$$

Caso haja indicação de que as matrizes $\sum_1$ e $\sum_2$ sejam semelhantes, a matriz comum $\sum_{pxp}$ será estimada pela matriz combinada S_{pxp} definida por:

$$S_{pxp} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$$

e a função discriminante a ser utilizada será a de Fisher, definida em (1.26), isto é,

$$\hat{f}d(x) = (\bar{x}_1 - \bar{x}_2)'S^{-1}x - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 + \bar{x}_2) \quad (1.32)$$

Se, no entanto, as matrizes de covariâncias forem diferentes entre si, a função discriminante a ser utilizada será a função quadrática definida em (1.31), que considera os sistemas de variabilidades das duas populações separadamente. Existem testes de hipótese que podem ser usados para decidir se as matrizes $\sum_1$ e

Σ_2 são iguais ou diferentes (Johnson; Wichern, 2002). Quando se tem apenas uma variável (p=1), o teste é equivalente a verificar se as variâncias das populações são iguais ou diferentes (Johnson; Bhattacharyya, 1997). No entanto, uma alternativa mais prática é aquela em que os dois modelos, linear de Fisher e quadrático, são ajustados aos dados e analisados, ficando-se, no final, com o que resultar em menores proporções de erros de classificação. Caso ambos dêem resultados semelhantes e satisfatórios, opta-se pelo modelo linear de Fisher, uma vez que a matriz de covariâncias teórica estará sendo estimada com um numero maior de observações.

2.5 Avaliação Da Qualidade De Ajuste Da Regra De Discriminação: Caso De Duas Populações

Após a construção da função discriminante, é necessário avaliar a sua qualidade. Para cada elemento amostral da população 1 e 2, calcula-se o escore numérico da função discriminante construída, e a analise destes escores permitirá que se faça uma avaliação da qualidade da função em termos de erros de classificação e capacidade de discriminação. Se a função é adequada, espera-se que os escores de uma população sejam bem diferenciados dos escores de outra população. Deste modo, um primeiro teste que pode ser feito é o de comparação das médias dos escores das duas populações. No caso de duas populações normais multivariadas independentes, esta comparação pode ser feita através da estatística de teste de Hotelling (Jobson, 1992) para dois grupos independentes, definida por:

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)}\, T^2 \tag{1.33}$$

onde,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\, (\bar{y}_1 - \bar{y}_2) \tag{1.34}$$

sendo $\bar{y}_1$ e $\bar{y}_2$ os escores da função discriminante quando aplicada aos vetores de médias amostrais das populações 1 e 2, respectivamente. Sob a hipótese de normalidade multivariada, pode ser mostrado (Anderson, 2003) que a estatística (1.33) tem distribuição F com p e $(n_1 + n_2 - p - 1)$ graus de liberdade. Assim, para um nível de significância α fixo, 0 < α < 1 a hipótese de igualdade de medias dos escores das duas populações será rejeitada se o valor observado da estatística F em (1.33) for maior ou igual ao respectivo valor crítico tabelado. Caso o teste de Hotelling não indique diferença significante entre as medias dos escores das duas populações, a função discriminante construída deve ser reformulada com o objetivo de buscar variáveis com maior poder de discriminação dos grupos.

2.6 Estimação Das Probabilidades De Classificações Incorretas

No caso de duas populações, existem dois tipos de erros que precisam ser avaliados:

Erro1: O elemento amostral pertence à população 1, mas a regra de classificação o classifica como sendo proveniente da população 2;

Erro2: O elemento amostral pertence à população 2, mas a regra de classificação o classifica como sendo proveniente da população 1;

As probabilidades de ocorrência destes erros são denotadas respectivamente por:

Prob (Erro 1) = p (2 / 1) *e Prob (Erro 2) = p* (1 / 2)

Onde p (2 / 1) é a probabilidade de classificar um indivíduo no grupo 2, dado que ele na verdade pertence ao grupo 1 e p (1 / 2) é a probabilidade de classificar um indivíduo no grupo 1, dado que ele na verdade pertence ao grupo 2.

Quanto menor forem estas probabilidades, melhor será a função de discriminação. Existem vários procedimentos para se estimar as probabilidades de classificações incorretas. A seguir, será apresentado um método não paramétrico que está disponível na maioria dos *softwares* estatísticos usuais.

2.6.1 Método da Resubstituição

Neste método, os escores de cada elemento amostral observado das populações 1 e 2 são calculados, sendo a regra de discriminação utilizada para classificar os ***n = n1 + n2*** elementos da amostra conjunta. Quando a função discriminante é de boa qualidade, espera-se que ela apresente uma grande porcentagem de acerto na classificação dos elementos amostrais em relação à população a que de fato pertencem. Portanto, neste método, os mesmos elementos amostrais participam da estimação da regra de classificação e da estimação dos erros de classificação. As freqüências de classificações corretas e incorretas podem ser resumidas como na Tabela 1.

Tabela 1: Tabela genérica das freqüências dos erros de classificação

		População classificada pela regra		
		1	**2**	**Total**
População de origem	**1**	n_{11}	n_{12}	$\boldsymbol{n_1}$
	2	n_{21}	n_{22}	$\boldsymbol{n_2}$

Sendo n_{ij} o número de elementos pertencentes à população de origem i e que são classificados pela função discriminante como pertencentes à população j, i, j = 1, 2. Quando i = j, tem-se o número de classificações corretas, e quando $i \neq j$, tem-se o numero de classificações incorretas. Com base nestes dados, as estimativas das probabilidades de ocorrência de erros 1 e 2 são dadas respectivamente por:

$$\hat{p}(2\,/\,1) = \frac{n_{12}}{n_1} \quad e \quad \hat{p}(1\,/\,2) = \frac{n_{21}}{n_2} \tag{1.35}$$

sendo n_{12} o numero de elementos da população 1 classificados incorretamente pela regra como provenientes da população 2, e n_{21} o numero de elementos da população 2 classificados incorretamente pela regra como provenientes da população 1.

Este procedimento de estimação do erro aparente de classificação (APER) é consistente, mas viciado (Johnson; Wichern, 2002) e tende a subestimar os verdadeiros valores de p(2 / 1) e p(1 / 2) para elementos que não pertencem à amostra conjunta utilizada para a construção da regra de discriminação, isto é, novos elementos amostrais. No entanto, pode servir como uma etapa inicial de avaliação, pois se os valores $\hat{p}$(2 / 1) e $\hat{p}$(1 / 2) forem muito elevados, é sinal de que a regra de discriminação deve ser reformulada. O vício deste procedimento tende a zero quando os tamanhos amostrais n_1 e n_2 são grandes.

2.7 Estimação Da Probabilidade Global De Acerto

Em todos os procedimentos de estimação tratados na seção 2.2, a probabilidade global de acerto da função discriminante é estimada por:

$$\hat{p}(acerto) = \frac{n_{11} + n_{22}}{n_1 + n_2} \tag{1.36}$$

Para a análise da qualidade da função discriminante, é recomendável estimar as probabilidades de ocorrência dos erros dos tipos 1 e 2 de classificações incorretas separadamente, e não apenas a probabilidade de acerto global da regra. É possível, por exemplo, ter uma função discriminante com alta probabilidade de acerto global, mas apresentando uma probabilidade alta em algum dos erros parciais dos tipos 1 ou 2.

3 APLICAÇÃO

3.1 Análise Discriminante: Um Exemplo Prático

Como exemplo, consideremos uma tabela com dados fictícios, que contém quatro grupos (1, 2, 3, 4) e quatro variáveis distintas, sendo elas X_1, X_2, X_3 e X_4, em um total de 40 observações.

Este exemplo prático é destinado, exclusivamente, a fins didáticos, sendo os dados meramente ilustrativos. X_1, X_2, X_3 e X_4 representam variáveis explicativas que poderão formar a função discriminante, a fim de que os casos sejam distribuídos nos grupos corretos.

Na fase de aplicação, faremos uso do *software* SPSS para a elaboração da Análise Discriminante.

3.2 Análise dos Resultados

A Tabela 2 apresenta as estatísticas descritivas univariadas, para cada um dos grupos da variável categórica, que mostra as médias, os desvios padrão e o número de observações em cada grupo, com 40 observações no total.

Tabela 2: Estatísticas Descritivas das Variáveis para cada Grupo

Grupo		*Média*	*Desvio Padrão*	*N Válido*	
				Observado	*Ponderado*
1	X1	5,5818	4,77816	10	10,000
	X2	-,4259	1,67768	10	10,000
	X3	-,0001	1,23690	10	10,000
	X4	2,5778	2,38911	10	10,000
2	X1	-,8392	2,14467	10	10,000
	X2	-,2838	1,31105	10	10,000
	X3	,6097	,66589	10	10,000
	X4	,0482	,94261	10	10,000
3	X1	,1202	,79945	10	10,000
	X2	3,3259	3,42617	10	10,000
	X3	-,1480	,92932	10	10,000
	X4	6,8649	7,11363	10	10,000
4	X1	-,2116	,73688	10	10,000
	X2	1,4499	2,04525	10	10,000
	X3	,8303	1,08380	10	10,000
	X4	2,4065	4,49183	10	10,000
Total	X1	1,2686	3,62449	40	40,000
	X2	1,0165	2,66518	40	40,000
	X3	,3230	1,04712	40	40,000
	X4	2,9743	4,90466	40	40,000

Ao se executar a análise discriminante*,* antes de qualquer coisa, é preciso analisar as variáveis independentes, verificando a existência, ou não, de algum tipo de diferença entre as médias. Segundo Maroco (2007, p. 351), o Lambda de Wilks testa a hipótese de que as médias dos grupos são iguais, onde entre as variáveis, pelo menos em um grupo as médias são diferentes; neste caso, o objetivo é rejeitar a hipótese nula. Dessa forma, as hipóteses testadas são as seguintes:

$$\begin{cases} H_0 = \text{média dos grupos são iguais; e} \\ H_1 = \text{média dos grupos diferem.} \end{cases}$$

O resultado desse teste está na Tabela 3, onde se percebe que apenas a variável **X_3** não rejeita a hipótese nula (H_0), pois o p-valor (sig.) é maior que o α (nível de significância) de 0,05. Com isso, essa variável não passou no pressuposto da igualdade das matrizes de variância e covariância, ou seja, não é significante na diferenciação entre os grupos. Por outro lado, as variáveis **X_1** e **X_2** são os que têm maior poder de discriminação, ou seja, são as variáveis que melhor discriminam os grupos, se comparadas às demais variáveis, por apresentar o p-valor F mais próximo de zero (o p-valor F expressa as diferenças entre as médias, onde os valores que se aproximam de zero indicam médias mais distintas entre si).

Tabela 3: Teste de Igualdade de Médias dos Grupos

	Lambda de Wilks	*F*	*df1*	*df2*	*Sig.*
X1	*,503*	*11,875*	*3*	*36*	*,000*
X2	*,665*	*6,057*	*3*	*36*	*,002*
X3	*,844*	*2,213*	*3*	*36*	*,103*
X4	*,742*	*4,166*	*3*	*36*	*,012*

Como pode ser observado na Tabela 4, o maior nível de correlação positiva ocorreu entre as variáveis **X_2** e **X_4**. A variável **X_1** também apresentou uma forte correlação positiva com a variável **X_4**. No decorrer da análise, veremos quais dessas variáveis serão boas discriminantes para a composição dos grupos.

Tabela 4: Matrizes de Covariância e de Correlação para todos os grupos

		X1	X2	X3	X4
Covariância	X1	*7,153*	*,496*	*-,485*	*2,774*
	X2	*,496*	*5,114*	*,041*	*8,127*
	X3	*-,485*	*,041*	*1,003*	*1,281*
	X4	*2,774*	*8,127*	*1,281*	*19,344*
Correlação	X1	*1,000*	*,082*	*-,181*	*,236*
	X2	*,082*	*1,000*	*,018*	*,817*
	X3	*-,181*	*,018*	*1,000*	*,291*
	X4	*,236*	*,817*	*,291*	*1,000*

a. A matriz de covariância tem 36 graus de liberdade.

Pode-se verificar que as variáveis X_2 e X_4 apresentaram a maior correlação. A menor correleção foi X_2 e X_3 com um valor de 0,018.

A Tabela 5 é a das matrizes de covariância dentro de cada grupo, pela qual o pesquisador pode ter uma idéia de homogeneidade de covariância, que é um dos pressupostos para a aplicação da análise discriminante. No entanto, para ter mais certeza, após a matriz de covariância, faremos uso do teste M de Box (*Box's M*) para avaliar com mais propriedade essa pressuposição.

Tabela 5: Matrizes de Covariância para cada um dos Grupos.

Grupo		X1	X2	X3	X4
1	X1	*22,831*	*1,841*	*-1,379*	*10,957*
	X2	*1,841*	*2,815*	*-1,775*	*,553*
	X3	*-1,379*	*-1,775*	*1,530*	*-,047*
	X4	*10,957*	*,553*	*-,047*	*5,708*
2	X1	*4,600*	*,427*	*-,594*	*1,920*
	X2	*,427*	*1,719*	*-,754*	*,319*
	X3	*-,594*	*-,754*	*,443*	*-,231*
	X4	*1,920*	*,319*	*-,231*	*,889*
3	X1	*,639*	*-,844*	*-,039*	*-1,882*
	X2	*-,844*	*11,739*	*1,121*	*24,087*
	X3	*-,039*	*1,121*	*,864*	*2,564*
	X4	*-1,882*	*24,087*	*2,564*	*50,604*
4	X1	*,543*	*,559*	*,074*	*,100*
	X2	*,559*	*4,183*	*1,571*	*7,551*
	X3	*,074*	*1,571*	*1,175*	*2,839*
	X4	*,100*	*7,551*	*2,839*	*20,177*
Total	X1	*13,137*	*-1,232*	*-,958*	*2,712*
	X2	*-1,232*	*7,103*	*-,161*	*10,865*
	X3	*-,958*	*-,161*	*1,096*	*,457*
	X4	*2,712*	*10,865*	*,457*	*24,056*

A matriz de variância/covariância de todas as variáveis nos 4 grupos foi:

$$S = \begin{pmatrix} 13{,}137 & -1{,}232 & -0{,}958 & 2{,}712 \\ & 7{,}103 & -0{,}161 & 10{,}865 \\ & & 1{,}096 & 0{,}457 \\ & & & 24{,}056 \end{pmatrix}$$

O teste conhecido como M de Box (*Box's M*), nos permite avaliar uma das pressuposições da análise discriminante, que é a homogeneidade das matrizes de covariância, em cada um dos grupos, para cada uma das variáveis de estudo. Se ao realizar o teste, seu p-valor (sig.) for maior que o nível de significância, então a igualdade das matrizes encontra sustentação, se for menor a suposição é violada. Logo, o objetivo é não rejeitar a hipótese que as matrizes são homogêneas. Queremos testar as seguintes hipóteses:

$$\begin{cases} H_0 = \text{matrizes são homogêneas; e} \\ H_1 = \text{matrizes não são homogêneas} \end{cases}$$

Tabela 6: Resultado do Teste Box's M

Box's M		*15,671*
F	Approx.	*1,183*
	df1	*10*
	df2	*1549,004*
	Sig.	*,298*

De acordo com a Tabela 6, o p-valor (Sig.) do teste é de 0,298, que é maior do que o nível de significância α que é de 0,05. Com isso, não temos evidências estatísticas para duvidar de que as matrizes sejam realmente homogêneas. Com isso, aceitamos H_0, passando assim no pressuposto de homogeneidade das matrizes de variância e covariância. Sendo assim, podemos dar continuidade à análise dos dados. Vale lembrar também que a estatística *Box's M* é diretamente influenciada pelo tamanho amostral e pelas diferenças de tamanho que algumas amostras podem apresentar.

Tabela 7: Autovalores (*Eigenvalues*)

Função	Autovalor	% de Variância	% Acumulado	Correlação Canônica
1	*1,234*[a]	*59,6*	*59,6*	*,743*
2	*,693*[a]	*33,5*	*93,1*	*,640*
3	*,142*[a]	*6,9*	*100,0*	*,353*

Os Autovalores são uma medida relativa de quão diferentes os grupos são na função discriminante, ou seja, quanto mais afastados de 1 forem os autovalores, maiores serão as variações entre os grupos explicadas pela função discriminante. Pela saída do SPSS, vemos que a primeira função discriminante apresenta um percentual de 59,6% [1,234/(1,234+0,693+0,142)], ou seja, esta função é a que mais contribui para demonstrar as diferenças entre os grupos. Já a segunda função apresenta um poder discriminante que explica 33,5% [0,693/(1,234+0,693+0,142)] da variância entre os grupos, seguido da terceira função, que explica apenas 6,9% [0,142/(1,234+0,693+0,142)].

O próximo teste de hipótese é o de Lambda de Wilks, que segundo Maroco (2003, p.344) serve para testar a significância das funções discriminantes e é calculado a partir do determinante da matriz da soma dos quadrados e produtos cruzados dentro dos grupos; e do determinante da matriz da soma dos quadrados e produtos cruzados total. As hipóteses a serem testadas são as seguintes:

$$\begin{cases} H_0 = \text{média populacional dos dois grupos são iguais; e} \\ H_1 = \text{média populacional dos dois grupos são diferentes.} \end{cases}$$

O objetivo do teste é não aceitar a H_0, pois as médias devem ser significativamente diferentes para melhor discriminar os grupos. O resultado desse teste está na Tabela 8.

Tabela 8: Lambda de Wilks e Qui-quadrado

Teste de Funções	Lambda de Wilks	Qui-quadrado	df	Sig.
1 até 3	,231	51,220	12	,000
2 até 3	,517	23,089	6	,001
3	,875	4,661	2	,097

Na primeira linha, são testadas as três funções em conjunto, podendo-se concluir que pelo menos a primeira função discriminante é altamente significativa. A linha seguinte testa a segunda função junto com a terceira, podendo-se concluir que pelo menos a segunda função discriminante é significativa. A terceira linha é referente à terceira função discriminante, na qual o *p-value* (sig.) é maior do que o α (nível de significância) de 0,05. Com isso, se aceita H_0, concluindo que a terceira função discriminante não é significativa. O lambda de Wilks mostra também que a primeira função é a que tem um melhor poder de discriminação, já que seu valor é o que mais se aproxima de zero. Os valores de Qui-quadrado apresentados na Tabela 8 testam a significância das funções discriminante, ou seja, o quão bem cada função separa as observações em grupos (testa se as médias são diferentes entre os grupos).

A Tabela 9 apresenta os coeficientes não padronizados das funções discriminantes para cada uma das variáveis explicativas.

Tabela 9: Coeficientes das Funções Discriminantes

	Função		
	1	2	3
X1	*,318*	*,085*	*,261*
X2	*-,311*	*,256*	*,697*
X3	*-,047*	*,923*	*,658*
X4	*,062*	*-,330*	*-,261*
(Constante)	*-,256*	*,314*	*-,476*

Por meio da Tabela 9, podemos escrever cada função da seguinte forma:

$$Z_1 = -0{,}256 + 0{,}318X_1 - 0{,}311X_2 - 0{,}047X_3 + 0{,}062X_4$$
$$Z_2 = 0{,}314 + 0{,}085X_1 + 0{,}256X_2 + 0{,}923X_3 - 0{,}330X_4$$
$$Z_3 = -0{,}476 + 0{,}261X_1 + 0{,}697X_2 + 0{,}658X_3 - 0{,}261X_4$$

Os valores dos coeficientes padronizados das funções discriminantes são apresentados na Tabela 10 a seguir:

Tabela 10: Coeficientes Padronizados das Funções Discriminantes

	Função		
	1	2	3
X1	,850	,226	,699
X2	-,703	,580	1,576
X3	-,047	,924	,659
X4	,272	-1,449	-1,148

Segundo Maroco (2007), esses coeficientes, que também são chamados de pesos discriminantes, podem ser utilizados para avaliar a importância relativa de cada variável explicativa para a função discriminante, sendo que a interpretação destas funções a partir deles deve ser feita com alguma preocupação caso haja problemas de colinearidade.

Assim, a matriz de estrutura apresentada a seguir, na Tabela 11, auxilia na interpretação da contribuição que cada variável forneceu para cada função discriminante, uma vez que apresenta as correlações entre as variáveis explicativas e as funções discriminantes canônicas padronizadas.

Tabela 11: Matriz de Estrutura

	Função		
	1	2	3
X1	*,865* *	*-,235*	*,438*
X4	*-,116*	*-,653* *	*,496*
X3	*-,135*	*,472* *	*,227*
X2	*-,412*	*-,569*	*,707* *

As variáveis cujos valores apresentam-se com asteriscos são as mais relevantes para a determinação de cada função discriminante, uma vez que oferecem maiores correlações com essas funções.

A partir dos coeficientes não padronizados das funções discriminantes, é possível definir a posição de cada um dos centróides dos grupos em um mapa territorial. Estas coordenadas encontram-se a seguir, na Tabela 12:

Tabela 12: Centróides dos Grupos

Grupo	Função		
	1	2	3
1	*1,810*	*-,173*	*,013*
2	*-,460*	*,717*	*-,504*
3	*-,820*	*-1,222*	*-,016*
4	*-,530*	*,677*	*,508*

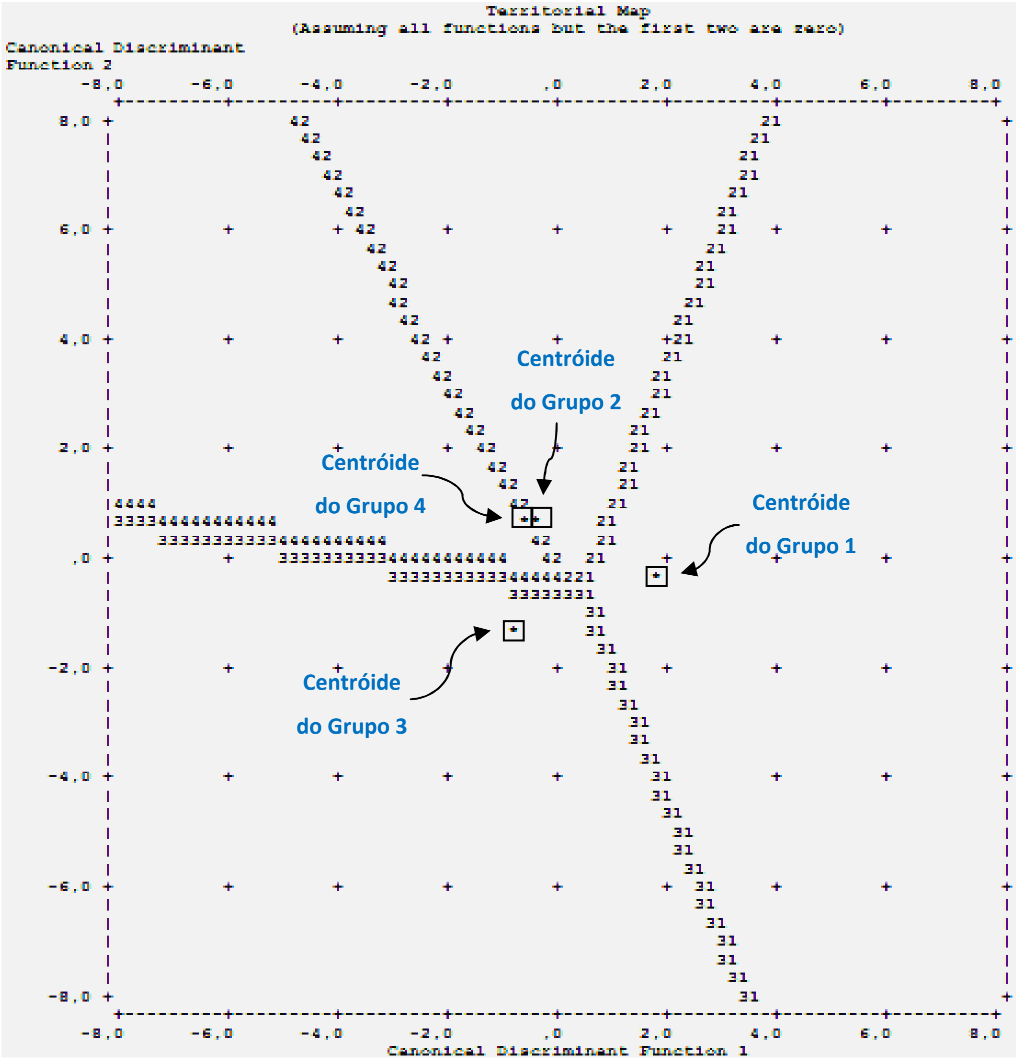
A Tabela 13 apresenta um sumário das probabilidades calculadas a *priori* a partir da amostra para a obtenção do ponto de corte crítico (MAROCO, 2007).

Tabela 13: Probabilidades Calculadas a *Priori*

Grupo	Priori	Casos Usados na Analise	
		Observado	Ponderado
1	,250	10	10,000
2	,250	10	10,000
3	,250	10	10,000
4	,250	10	10,000
Total	1,000	40	40,000

O mapa territorial deste caso é apresentado na Figura 4 a seguir, com destaque para os centróides de cada grupo.

Figura 4: Mapa Territorial da Análise Discriminante



O mapa territorial nos dá a idéia do comportamento dos grupos e nos mostra as áreas delimitadas por cada um dos grupos, bem como os centróides, que é o parâmetro que nos mostrará onde um novo elemento amostral será alocado de acordo com os valores observados das variáveis X1, X2, X3 e X4. O Mapa Territorial nos dá uma idéia também de como as funções discriminantes separam os 4 grupos. Em um exemplo prático, ao se substituir os coeficientes observados (X1, X2, X3 e X4) nas funções discriminantes e seu valor comparado com os centróides de cada grupo, o novo indivíduo será alocado no grupo que tiver o valor do centróide mais aproximado do valor estimado da função discriminante. Como vemos pelo mapa territorial, os centróides dos grupos 4 e 2 são bem parecidos e se um valor estimado pela função discriminte se aproxima do centróide 4, conseqüentemente, também se aproximará do centróide 2, podendo ser alocado tanto no grupo 2 quanto no grupo 4, gerando assim, um erro de classificação.

Por fim, a Tabela 14 apresenta os resultados da classificação.

Tabela 14: Resultados da Classificação

			Previsão de Membros de Grupos				
		Grupo	1	2	3	4	Total
Original	Contagem	1	*8*	*1*	*1*	*0*	*10*
		2	*1*	*7*	*0*	*2*	*10*
		3	*0*	*1*	*8*	*1*	*10*
		4	*0*	*4*	*1*	*5*	*10*
	%	1	*80,0*	*10,0*	*10,0*	*,0*	*100,0*
		2	*10,0*	*70,0*	*,0*	*20,0*	*100,0*
		3	*,0*	*10,0*	*80,0*	*10,0*	*100,0*
		4	*,0*	*40,0*	*10,0*	*50,0*	*100,0*

A Tabela 14 é muito informativa acerca do sucesso (ou não) das funções discriminantes. No nosso caso, a análise evidenciou que 70% [(8 + 7 + 8 + 5) / 40] dos casos foram classificados corretamente, enquanto que 30% [(1+ 6 + 2 + 3) / 40] dos casos foram classificados de maneira incorreta.

4 CONCLUSÃO

Diante do estudo realizado, pode-se concluir que no contexto do nosso exemplo, considerando-se os resultados gerais da análise discriminante, depois de verificado que os pressupostos foram de fato atendidos, que as variáveis selecionadas têm significância discriminante e que a própria função discriminante é significante, pela saída do SPSS vemos que a primeira função discriminante dada por $Z_{1,}$ apresenta um percentual de 59,6% de explicação total da variância, ou seja, esta função é a que mais contribuiu para demonstrar as diferenças entre os grupos.

Vimos também, pela matriz de estrutura (11) que a variável que mais contribuiu para a determinação da primeira função discriminante, foi a variável X_1 uma vez que ofereceu maior correlação com a mesma.

Ao fim da nossa pesquisa, os resultados mostraram que 70% dos casos foram classificados corretamente enquanto que 30% foram classificados de maneira incorreta, que nesse exemplo, pode ser considerado um bom nível de classificação.

O SPSS mostrou-se ao longo do trabalho, uma ferramenta muito poderosa no que diz respeito à aplicação das técnicas de análise discriminante, por todos os recursos que dispõe. Percebe-se, que para a aplicação das técnicas multivariadas no SPSS, a habilidade de manusear um *software* como esse, e um sólido conhecimento dos conceitos teóricos envolvidos ao longo da análise, são fundamentais para a realização de uma pesquisa dessa natureza.

Além da análise discriminante linear de Fisher, que foi usada nesse trabalho, existem ainda outras técnicas discriminantes que também podem ser usadas para o mesmo fim, como por exemplo, a análise discriminante canônica, a análise discriminante quadrática e a análise discriminante de núcleo.

Como continuidade desse estudo poderiam ser utilizadas outras técnicas de análise discriminante, bem como as já citadas, e feito uma aplicação com uma massa de dados não-fictícios, partindo-se desde a amostragem até a aplicação da técnica de AD e considerando-se um maior número de variáveis dependentes e o estudo de um número maior de grupos, o que conseqüentemente, demandaria em um número maior de funções discriminantes.

REFERÊNCIAS

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, *New York: John Wiley.*

BARTLETT, M.S. (1947), *"Multivariate Analysis*", *Journal of the Royal Statistical Society Suppl.,* Vol 9, 176-190.

BHATTACHARYYA, G. K., JOHNSON, R. *Statistical concepts and methods*. John Willey & Sons, 1977. 639p.

BHATTACHARYA, C.B., HAYAGREEVA Rao, and Mary Ann Glynn (1995), "*Understanding the Bond of Identification: An Investigation of Its Correlates Among Art Museum Members*," *Journal of Marketing*, 59 (October), 46–57.

BOX, G. E. P., 1949. *A general distribution theory for a class of likelihood criteria*. *Biometrika*, 36: 317–346.

CASELLA, G., and BERGER, R. (2002), *Statistical Inference (2nd ed*.), Paci. C Grove, CA: Duxbury Press.

CECATTO, Cristiano. *Práticas de previsão de demanda nas indústrias alimentícias brasileiras.* Disponível em: <http://www.fei.edu.br/Download%20de%20Pesquisas/FEI-PD-003-CristianoCecatto.pdf > Acesso em: 02 Set 2010.

COSTANZA M. C. and AFIFI A. A. (1979), *Comparison of stopping rules in forward stepwise discriminant analysis*, *Journal of the American Statistical Analysis*, 74, 777-785.

CRASK, M.R., and W.D. PERREAULT, Jr. (1977), "*Validation of Discriminant Analvsis in Marketing Research," Journal of Market Research 14 (February), pp.60-68.*

FÁVERO, Luiz Paulo, et al. *Análise de dados : modelagem multivariada para tomada de decisões.* Rio de Janeiro: Elsevier, 2009.

FISHER, L., and VAN NESS J.W. (1973), "*Admissible Discriminant Analysis," Journal of the American Statistical Association 68, pp. 603-607.*

FRALEY C. and RAFTERY A. 2002. *Model-Based Clustering, Discriminant Analysis, and Density Estimation*. *J. Am. Stat. Assoc., 97 (458) : 611-631.*

FRANK, R. E., MASSY, W F., & Morrison, D G. *Bias in multiple discriminant analysis. Journal of Marketing Research*, 1965, 2, 250-258.

GONÇALVES, C. A.; DIAS, A. T.; MUNIZ, R. M. *Análise discriminante das relações entre fatores estratégicos, indústria e desempenho em organizações brasileiras atuantes na indústria manufatureira.* RAC, Curitiba, v. 12, n. 2, p. 287-311, Abr./Jun. 2008.

HAIR JUNIOR, J. F. et a l. *Análise multivariada de dados*. 5. ed. Porto Alegre: Bookman, 2005.

HAIR, J. F., Anderson, R. E., TATHAM, & BLACK, W. C. (2005). *Análise multivariada de dados* (A. S. Sant'Anna & A. C. Neto, trad.). Porto Alegre: Bookman.

HORA, S. C., & WILCOX, J. B. (1982). *Estimation of error rates in several-population discriminant analysis. Journal of Marketing Research, 19, 57-61.*

JOBSON, J.D., *Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods, Springer-Verlag, 1992.*

JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis. Fifith Edition. New Jersey: Prentice Hall, 2002.*

LACHENBRUCH, P. A. and MICKEY, M. R. (1968). *Estimation of error rates in discriminant analysis, Technometrics 10: 1-11. Li, K.-C. (1991).*

KATO, H. T.; ANTUNES, M. T. P.; CORRAR, L. J. *A eficiência das informações divulgadas em "melhores & maiores" da revista exame para a previsão de desempenho de empresas.* Revista Contabilidade & Finanças - USP, São Paulo, Edição Especial, p. 41 - 50, 30 junho 2004.

KRZANOWSKI. W. J. (1975). "*Discrimination and Classification Using both binary and continuous variables.*" *Journal of the American Statistical Association, 70. 782-790.*

MAHALANOBIS, P. C. 1936. *On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India 2: 49–55.*

MAROCO, J. (2007) *Análise Estatística com a utilização do SPSS.* 3ª Ed. Silabo. Lisboa. 822 pp.

MELLO, Gilmar R.; MACEDO, F. Q.; FILHO, F. T. *Identificando o endividamento dos Estados Brasileiros: uma proposta através de análise discriminante.* Disponível em: <http://periodicos.uem.br/ojs/index.php/Enfoque/article/viewPDFInterstitial/3504/3174> Acesso em: 01 Ago 2010.

MELLO, Gilmar R.; SLOMSKI, Valmor. *A situação financeira dos Estados Brasileiros: Uma proposta utilizando análise discriminante.* RCO – Revista de Contabilidade e Organizações – FEARP/USP, v. 1, n. 1, p. 73 - 86 set./dez. 2007.

MINGOTI, S. A**.** *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005.

RANDLES, R.H., J.D. BROFITT, J.S. RAMBERG, and R.V. HOGG (1978), "*Discriminant Analysis Based on Ranks,*" *Journal of the American Statistical Association 73, pp. 379-384.*

RENCHER A.C. (2002) *Methods of Multivariate Analysis*, *2nd edn. John Wiley and Sons, New York.*

SHARMA, S. 1996. “*Proactive environmental responsiveness: Catalysts for changing organizational paradigms and organizational capabilities*.” BPS, ONE, and SIM. *Academy of Management Annual Meetings, Cincinnati, OH, August 1996.*

SARTORIO, Simone Daniela. *Aplicações de técnica de análise multivariada em experimentos agropecuários usando o software R* / Simone Daniela Sartorio. - Piracicaba, 2008. 130 p.

SILVA, Tomás da. *Notas para a disciplina de Análise Estatística III*. Disponível em: < https://woc.uc.pt/fpce/getFile.do?tipo=2&id=1684> Acesso em: 21 Ago 2010.

STEVENS, J.P. (2002), *Applied Multivariate Statistics for the Social Sciences*, *Lawrence Erlbaum Associates, Mahwah, NJ.*

APÊNDICE A

Tabela 15: Dados usados na aplicação da AD

	Grupos	X1	X2	X3	X4
1	1	14,03	0,43	-0,04	7,2
2	1	12,74	-0,9	-0,28	5,64
3	1	4,62	1,85	-1,86	1,38
4	1	4,01	2,33	-1,77	1,4
5	1	6,3	-1,68	1,63	3,95
6	1	2,95	-1,22	1,15	2,01
7	1	-2,02	-1,58	1,43	-0,37
8	1	5,36	0,55	-0,87	2,08
9	2	-2,87	-0,87	1,01	-0,86
10	1	1,85	-3,03	0,58	0
11	1	5,99	-1,01	0,01	2,5
12	2	-4,75	-1,1	1,52	-1,41
13	2	-2,02	-1,58	1,43	-0,37
14	2	-0,2	2	-0,32	0,58
15	2	-1,8	-0,77	0,7	-0,59
16	2	2,95	-1,22	1,15	2,01
17	2	-0,54	0,09	0,4	0,18
18	2	1,07	-1,23	0,41	0,33
19	2	0,11	-0,2	0,13	0,08
20	2	-0,35	2,04	-0,32	0,53
21	3	-0,28	10,07	0,39	21,56
22	3	-0,36	7,2	0,43	14,03
23	4	-0,32	1,38	1,85	4,62
24	3	-0,19	5,63	0,51	11,39
25	3	-0,32	3,95	-1,68	6,3
26	3	-0,34	0,13	-0,7	1,35
27	3	1,43	1,92	1,23	3,29
28	3	-0,87	2,21	0,73	6,02
29	3	0,58	-0,86	-0,87	-2,87
30	4	0,01	-0,74	-0,35	-2,59
31	3	1,52	2,5	-1,01	5,99
32	4	-0,27	1,41	0,63	3,94
33	4	-0,21	0,28	0,44	0,15
34	4	1,06	1,4	0,74	4,03
35	4	-0,28	-0,79	-1,03	-2,67
36	4	-0,78	1,2	2,05	3,12
37	4	1,15	0,63	0,15	0,39
38	4	1,35	3,77	2,21	0,33
39	4	0,4	5,96	1,61	12,75
40	3	0,03	0,51	-0,51	1,59