

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
PRODUÇÃO**

**APLICAÇÃO DE UMA METODOLOGIA MULTIVARIADA  
AO CONTROLE INTERNO DE QUALIDADE DOS  
LABORATÓRIOS DE ANÁLISES CLÍNICAS**

**Dissertação submetida à Universidade Federal de Santa Catarina  
para obtenção do grau de mestre em Engenharia de Produção**

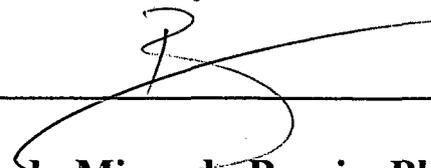
**MARCELO FLORENTINO**

**Florianópolis, Maio de 2000**

**APLICAÇÃO DE UMA METODOLOGIA MULTIVARIADA  
AO CONTROLE INTERNO DE QUALIDADE DOS  
LABORATÓRIOS DE ANÁLISES CLÍNICAS**

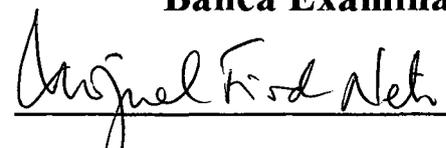
**MARCELO FLORENTINO**

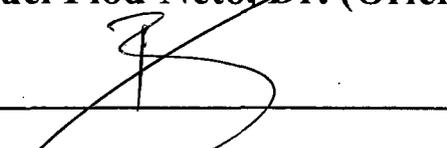
**Esta dissertação foi julgada adequada para obtenção do título de  
"Mestre em Engenharia", na especialidade Engenharia de  
Produção e aprovada em sua forma final pelo Programa de Pós-  
Graduação.**

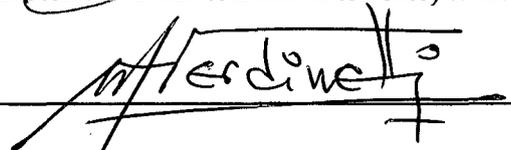
  
\_\_\_\_\_  
**Prof. Ricardo Miranda Barcia, Ph.D.**

**Coordenador do Curso**

**Banca Examinadora:**

  
\_\_\_\_\_  
**Prof. Miguel Fiod-Neto, Dr. (Orientador)**

  
\_\_\_\_\_  
**Prof. Ricardo Miranda Barcia, Ph.D.**

  
\_\_\_\_\_  
**Prof. Miguel Verdinelli, Dr.**

  
\_\_\_\_\_  
**Prof. Antônio Sérgio Coelho, Dr.**

**“(...) Dans chacun de tes gestes un aveu, un secret dans chaque attitude. Même la plus discrète ne peut mentir à tant de solitude(...)”**

**(Jean Bourrô, música Dans Notre Main)**

**(Em cada um de seus gestos um consentimento, um segredo em cada atitude. Mesmo o mais discreto não pode mentir a tanta solicitude)**

## **AGRADECIMENTOS**

**Aos meus pais, Adelaudio e Maria Terezinha, pelo carinho e amor: vocês são a razão maior de todos os meus esforços, agradeço a Deus todos os dias por tê-los;**

**ao Prof. Miguel Fiod-Neto, pela paciência, compreensão e orientação;**

**aos amigos, professores e a todos que passaram e continuam a passar pelo meu caminho, nesta jornada infundável que é a vida, obrigado pelos ensinamentos deixados;**

**ao pessoal do Laboratório de Análises Clínicas do Hospital Universitário, pela disposição em me ceder os dados trabalhados nesta dissertação;**

**aos demais professores, funcionários e colegas do Programa de Pós-Graduação em Engenharia de Produção, pela colaboração prestada;**

**à Universidade Federal de Santa Catarina e à CAPES, que proporcionaram os meios para a realização deste trabalho;**

**a essa "Força Natural" que nos rege, a qual chamamos Deus;**

**enfim, a todos aqueles que de alguma forma contribuíram para a realização deste trabalho.**

## RESUMO

O presente trabalho se propõe a desenvolver uma metodologia que permita examinar se duas “amostras controles”, artificios utilizados no controle de qualidade interno de laboratório de análises clínicas, possuem um perfil de dosagens dos seus componentes nitidamente diferentes. Para tal, aplica-se a técnica estatística multivariada de análise discriminante nos dados coletados “em campo”.

O método exige que as observações pertençam a grupos já estabelecidos *a priori* e sejam distintos, pois a análise irá construir uma combinação linear, representado por um índice  $Z_i$ , que permitirá verificar, estatisticamente, se os grupos possuem uma relação causal de igualdade ou diferenciadoras entre suas características.

A metodologia preconiza que uma combinação linear ótima será aquela que apresentar alto poder discriminante, ou seja, a que mais corretamente alocar as observações em seu grupo de origem.

Assim, pode-se descobrir o comportamento (perfil) das dosagens de cada “amostra controle” submetida ao teste, e então verificar se uma nova observação obedece às características do seu grupo de origem. Caso ela seja classificada, através da metodologia, em outro grupo diferente, conclui-se que essa observação sofreu alguma alteração no seu processo e terá que ser revista.

A técnica mostrou-se eficaz ao classificar as dosagens dos componentes corretamente, podendo ser usada na construção de um modelo matemático que fornecerá uma “ferramenta multivariada” para o controle de qualidade interno do laboratório de análises clínicas.

## ABSTRACT

The present work considers to develop a methodology that allows to examine if two " samples controls ", artifices used in the internal quality control of laboratory, possess a profile of dosagens of its nitidamente different components. For such, we apply the technique multivariated statistics of discriminant analysis in the collected data.

The method demands that the groups already are established a priori and are distinct, therefore the analysis will only go to construct a linear combination, represented for a  $Z_i$  index that will allow to solve, estatisticamente, if the groups possess a causal relation of equality or diferenciadoras between its features.

The methodology praises an excellent linear combination will be that one that to present high to be able discriminant, or either, the one that more correctly to place the comments in its group of origin

Thus, the behavior (profile) of the dosagens of each " sample can be discovered has controlled " submitted to the test, and then, to verify if a new comment obeys the features of its group of origin. In case that it is classified, through the methodology, in another different group, she concludes that this comment suffered some alteration in its process and will have that to be reviewed.

Finally, the technique correctly revealed efficient when classifying the dosagens of the components, being able to be used in the construction of a mathematical model that will supply a " multivariated tool " the internal quality control.

## SUMÁRIO

CAPÍTULO I .....	1
<i>INTRODUÇÃO</i> .....	1
1.2 <i>Objetivo do trabalho</i> .....	5
1.3 <i>Importância do Trabalho</i> .....	5
1.4 <i>Metodologia do Trabalho</i> .....	5
1.5 <i>Estrutura do Trabalho</i> .....	7
CAPÍTULO II .....	9
<i>O LABORATÓRIO DE ANÁLISES CLÍNICAS</i> .....	9
2.2. <i>Procedimentos de Controle de Qualidade</i> .....	12
2.3. <i>Fases e Etapas dos Procedimentos de Controle</i> .....	17
CAPÍTULO III.....	26
<i>MÉTODO ESTATÍSTICO MULTIVARIADO DE ANÁLISE DISCRIMINANTE</i> <i>[FUNÇÃO DISCRIMINANTE LINEAR DE FISHER]</i> .....	26
3.1. <i>Métodos estatísticos multivariados</i> .....	26
3. 2. <i>Análise Discriminante</i> .....	29
3. 3. <i>Métodos de análise da qualidade da discriminação</i> .....	36
3. 4. <i>Recursos Computacionais</i> .....	38
CAPÍTULO IV.....	39
<i>APLICAÇÃO DO MÉTODO PROPOSTO</i> .....	39
4.2. <i>Coleta, seleção e padronização da amostra</i> .....	40
4. 3. <i>Função discriminante (modelo matemático)</i> .....	41

4. 4. <i>Validade da classificação</i> .....	43
4. 5. <i>Discussão</i> .....	47
CAPÍTULO V.....	48
<i>CONCLUSÕES DO TRABALHO E SUGESTÕES PARA TRABALHOS FUTUROS</i> .....	48
5.1 <i>Sugestões para trabalhos futuros</i> .....	50
REFERÊNCIAS BIBLIOGRÁFICAS.....	52
ANEXO.....	56
<i>ANEXO I - APLICANDO O EXEMPLO NO SOFTWARE STATISTICA</i> .....	57

## FIGURAS

Figura 2. 1 Esquema das possíveis utilizações com os resultados fornecidos por laboratório de análises clínicas .....	11
Figura 2. 2 - Gráfico de Controle de Shewhart com os limites de controle e relação com a curva de distribuição normal, onde $X_m$ é o valor médio [Tibúrcio, 1995] .....	15
Figura 2. 3 – Folha de Registro.....	19
Figura 2.4 – Fluxograma para uso com as regras múltiplas [Tibúrcio, 1995].....	24
Figura 3. 1 – Classificação dos métodos multivariados, segundo o número de populações e o número de conjuntos de variáveis em consideração [Silva & Machado, 1996].....	28
Figura 3. 2 – Ilustração gráfica do método demonstrando o novo vetor $y$ que separa os conjuntos A e B [Steiner & Carnieri, 1995].....	32

## TABELAS

Tabela 4. 1 - Valores padronizados em porcentagem dos compostos dosados nas “amostras controles” utilizados para alimentar o <i>software</i> na construção da função discriminante, onde: os dados de um a vinte e um pertencem ao conjunto de observações advindas da “amostra controle” que representa o grupo 1, do dado vinte e dois ao quarenta e dois compõe o grupo 2.....	42
Tabela 4. 2 - Matriz de classificação das observações das “amostras controle”, através do método escolhido, com as respectivas percentagens de acerto da discriminação, onde nas linhas está representado o número de observações que compõem os grupos <i>a priori</i> e nas colunas está o número das observações classificadas depois da aplicação do modelo matemático.....	44
Tabela 4.3 - Tabela de classificação dos casos, onde se mostra se as variáveis dos dois grupos foram identificados cada qual no seu grupo, revelando uma perfeita discriminação dos soros controles. Os dados mal classificados, que pertencem a um grupo mas estão classificados em outro grupo, estão representados com um * [Origem: STATISTICA, 1999].....	46

# CAPÍTULO I

## INTRODUÇÃO

As atuais tendências de modernização e os conceitos da qualidade total vêm mudando o perfil das instituições prestadoras de serviço, tornando-as mais conscientes das expectativas dos seus clientes.

Dentro dessa visão, muitas mudanças vêm ocorrendo também nos Laboratórios de Análises Clínicas, com o uso cada vez mais intenso de processos automatizados, de forma que os exames sejam cada vez mais precisos, exatos, rápidos e com menor custo. Para se ter certeza de estar satisfazendo às novas expectativas do usuário, é necessária a introdução e atualização, por parte do laboratório, de indicadores e índices de desempenho para cada uma de suas atividades [Henry, 1995].

O controle de qualidade utilizado dentro dos Laboratórios segue antiga preocupação do ser humano: a de realizar tarefas que não possuam defeitos. A primeira idéia de controlar as características de um processo produtivo é devida a Walter Shewhart que, em 1931, desenvolveu e instrumentalizou um conjunto de conceitos, lançando as bases do moderno Controle Estatístico de Processos (CEP), técnica já consagrada no ambiente industrial [Berwick, 1990].

O CEP utiliza técnicas estatísticas e ferramentas da qualidade para medir e avaliar sistematicamente um processo. Caso, determinado processo, apresente situação indicativa de descontrole estatístico, ações corretivas devem ser determinadas e executadas de forma a restabelecer a estabilidade do processo. Após o restabelecimento do estado de estabilidade é possível determinar a capacidade do processo e conduzi-lo a níveis de qualidade desejáveis [Croiser, 1988].

Alguns benefícios do CEP recaem sobre a melhoria contínua da qualidade, com atuação direta sobre produtos, processos e serviços.

Durante a fase de controle por meio do CEP, o desempenho do processo é medido e avaliado. Isso possibilita o surgimento de oportunidades de melhoria ou aparecimento de problemas a serem resolvidos, tal que o processo possa ser aprimorado. Acompanhar o processo sistematicamente é importante para evitar que problemas resolvidos reapareçam, porém o investimento na manutenção dos níveis atuais não assegura sua efetividade ao longo do tempo. É preciso buscar sempre o aprimoramento. Conforme o processo de melhoria da qualidade vai avançando, o CEP vai se tornando menos usado para rotina de controle da qualidade e mais usado como ferramenta para melhoria da qualidade.

A aplicação, com exemplo bem sucedido nas indústrias [Cembrowski & Sullivan, 1996; Juran & Gryna, 1993], do CEP nos Laboratórios de Análises Clínicas ocorreu 20 anos depois de introduzidos por Shewhart. A idéia inicial partiu de Levey e Jennings [Tibúrcio, 1995], que descreveram o programa denominado Controle Interno da Qualidade Analítica (CIQA), com o objetivo de inspecionar o procedimento analítico utilizado para mensurar as amostras biológicas dos clientes e monitorar se está operando com a estabilidade, a exatidão e a precisão adequadas.

Para poder almejar a este objetivo, o programa CIQA é planejado utilizando-se um artifício chamado “amostra controle”, consistindo de material de composição humana, com características estáveis e reprodutivas dos compostos analisados na rotina do laboratório, podendo ser empregado no controle diário de acordo com as dosagens programadas.

Conforme Levey e Jennings, os gráficos de controles são preparados num processo baseado na média das medidas realizadas em uma única “amostra controle”.

Em 1952, Henry e Segalone [in Tibúrcio, 1995] modificaram o gráfico de controle de Levey e Jennings, no qual valores individuais de uma amostra controle são colocados diretamente no gráfico de controle, originando assim um procedimento de controle simplificado. Este tipo de procedimento de controle é o mais comumente usado pelos Laboratórios de Análises Clínicas.

Atualmente, para dar maior agilidade ao processo de Controle Interno, as “amostra controle” comerciais já trazem os valores médios das diferentes duplicatas de medida padronizadas e tabeladas para cada composto, facilitando a produção dos gráficos de controle.

Para se postular a excelência em controle de qualidade interno para o laboratório, conforme a literatura, é necessário focalizar algumas metas, como a utilização simultânea de no mínimo dois lotes de “amostra controle”, sendo que cada lote deve situar-se em faixas de concentração distinta, para obter-se maior número de informações sobre a estabilidade do processo e, posteriormente, auxiliando na tomada de decisão para correção dos eventuais erros no procedimento de dosagem [Tibúrcio, 1995].

Dada a grande importância do CEP para a melhoria contínua, houve, durante as últimas décadas, considerável crescimento do interesse por novas técnicas que se encaixem nesse instrumento.

Uma das mais recentes técnicas que está despontando nesse contexto, graças aos avanços da estatística aliados ao amplo crescimento da informática, é o de Reconhecimento de Padrões, ciência que compreende a identificação ou classificação das observações em categorias [Steiner & Carnieri, 1995].

Vários estudos já têm demonstrado as aplicações de Reconhecimento de Padrões, que incluem reconhecimento de caracteres, diagnóstico médico [Wolberg & Mangasarian, 1990], problemas bancários [Tam & Kiang, 1992], análise de sinais e imagens biomédicas, reconhecimento e compreensão de línguas, identificação de faces humanas [Samal & Iyengar, 1992] e impressões digitais, confiabilidade, automação industrial, socioeconomia, arqueologia, controle de qualidade [Steiner & Carnieri, 1994] e outras.

As várias técnicas usadas para resolver problemas de Reconhecimento de Padrões podem ser agrupadas em duas abordagens gerais [Fu,1982]:

- a abordagem da decisão teórica ou discriminante;
- a abordagem sintática ou estrutural.

Na abordagem da decisão teórica, um conjunto de medidas de características é extraído de padrões. Estes métodos incluem o reconhecimento de caracteres, construindo uma informação que descreve um padrão com capacidade de designar aspectos do modelo para uma classe particular, de tal forma que o mesmo seja inelegível para designação para outra classe.

A abordagem sintática, por sua vez, ao enfatizar a descrição estrutural dos padrões, tenta fazer uma analogia entre a estrutura de padrões e a sintaxe de uma linguagem. A analogia é atraente por causa da disponibilidade de lingüística matemática como fundamentação teórica [Fu,1982].

A presente dissertação trata do Reconhecimento de Padrões e de suas aplicações através do uso da abordagem da decisão teórica ou discriminatória, nos dados extraídos das dosagens das “amostras controles” do setor de bioquímica clínica de um laboratório de análises clínicas, criando padrões que poderão ser utilizados na construção de uma nova forma de inspeção da qualidade.

Para este método, além de se trabalhar com ensaio multivariado, tem-se a vantagem de que se a combinação linear for eficaz em construir um índice que diferencie o perfil das observações, esses dados serão alojados em seus respectivos grupos já estabelecidos.

Com o uso desta técnica, pode-se então verificar se uma nova observação, advinda de uma “amostra controle”, obedece às características desse grupo. Caso ela seja classificada, através da metodologia, em outro grupo que não o de origem, conclui-se que essa observação sofreu alguma alteração no seu processo e terá que ser revista.

Com essa técnica, é possível criar novo modelo matemático que fornecerá uma “ferramenta multivariada” para o controle interno de qualidade, que não seja somente o do

modelo atual, construído por um processo baseado na média univariada das dosagens da “amostra controle”.

## 1.2 Objetivo do trabalho

Visando dar melhor estrutura e a facilitar sua análise e permitir uma implementação gradativa das propostas aqui formuladas, os objetivos foram divididos em dois objetivos principais apresentados a seguir :

1) dos pontos coletados "em campo", avaliar, através de análise estatística, a existência de diferença estatística significativa entre as “amostras controle”.

Para isso, serão aplicados os procedimentos de Reconhecimento de Padrões através do uso da abordagem da decisão teórica ou discriminatória, criando um possível ambiente para construção dos padrões dos conjuntos de “amostra controle” A e B, estatisticamente;

2) após, construir o modelo matemático que constitua nova ferramenta para o controle de qualidade no serviço do laboratórios de análises clínicas.

## 1.3 Importância do Trabalho

O resultado do serviço prestado pelos Laboratórios de Análises Clínicas é comparado com controles externos e internos, procurando verificar se os procedimentos analíticos estão dentro dos limites de qualidade adequados e gerar soluções, caso a metodologia esteja fora das especificações. Dessa forma, as oportunidades de melhoria tornam-se mais visíveis e pode-se aprimorar cada vez mais a precisão dos resultados de diferentes técnicas que são empregadas, transmitindo aos usuários confiança no serviço prestado. Realmente, a aplicação da sistemática de melhoria contínua em ambientes de prestação de serviços de saúde mostrou-se bastante eficiente e eficaz [Berwick *et al.*, 1990].

O uso do resultado dos exames realizados nos Laboratórios de Análises Clínicas é um fator determinante na estruturação das atividades relacionadas à qualidade. Um resultado não

confiável pode gerar insatisfação ao cliente de tal grandeza que qualquer atributo adicional que acompanha o serviço – como instalações, cortesia, preços, facilitadores, etc. – pode não compensar.

Como o interesse em problemas de reconhecimento de padrões multivariados tem crescido consideravelmente nestas últimas décadas, como por exemplo o uso de técnicas multivariadas para calibração descrito por Beebe & Kowalski (1987), vê-se a necessidade de estudar uma metodologia estatística multivariada, para mais eficientemente poder-se fazer o reconhecimento de novos padrões, a fim de estruturar novas técnicas estatísticas para auxiliar os laboratórios.

#### 1.4 Metodologia do Trabalho

O desenvolvimento do trabalho passou por quatro momentos:

##### 1.4.1 - Revisão bibliográfica:

Inicialmente foi realizada revisão da literatura a respeito de técnicas de controle estatístico de qualidade e sobre as técnicas referentes ao reconhecimento de padrões.

##### 1.4.2 - Caracterização do universo estudado:

O estudo experimental de reconhecimento de padrão foi realizado no Laboratório de Análises Clínicas do Hospital Universitário da Universidade Federal de Santa Catarina, localizado em Florianópolis. Aqui, foi caracterizado, detalhadamente, como se dá o processo e a técnica de controle da qualidade interno para os exames laboratoriais.

1.4.3 - Utilização das técnicas selecionadas, análise dos resultados e caracterização amostral:

Neste momento, é colocada em prática a técnica de reconhecimento de padrão com abordagem da decisão teórica ou discriminante.

A característica amostral de interesse pode ser classificada como qualitativa ou quantitativa e tratada como uma variável. A variável que representa uma determinada característica populacional é chamada de qualitativa quando resultar de uma classificação por tipos ou atributos. Já as variáveis quantitativas podem ser classificadas em discretas ou contínuas, se seu domínio for finito (ou infinito enumerável) ou real, respectivamente. [Grotschel & Wakabayashi, 1989], [Mangasarian et al, 1990], [Batista, 1990].

No presente trabalho, trabalhou-se com  $n$  variáveis, sendo que destas apenas uma (tipo de “amostra controle”) é qualitativa; as demais são quantitativas.

#### 1.4.4 - Conclusões

Após o término do estudo, foi possível tirar conclusões a respeito do procedimento analisado. Identificado o método com melhor performance, o Reconhecimento de Padrões no caso do problema controle de qualidade interno fica resolvido. Dado um novo ponto, ou seja, os exames clínicos da “amostra”, aplica-se o método em questão para classificá-lo e através de comparação do resultado verifica-se se o procedimento apresenta da qualidade esperada.

#### 1.5 Estrutura do Trabalho

O trabalho divide-se em cinco capítulos, incluindo esta introdução.

O Capítulo II trata do estado da arte, a revisão de literatura. Apresenta a caracterização das principais técnicas que são utilizadas no processo de controle de qualidade interno no setor de bioquímica clínica, no caso encenado em laboratório clínico, e são mostrados os procedimentos para se fazer a coleta de dados e a análise estatística destes.

No Capítulo III é apresentado o método de Reconhecimento de Padrões pesquisado.

Dentro da abordagem teórica mencionada no item 1.1, os métodos de Reconhecimento de Padrões são aqueles capazes de fazer a análise discriminante e de classificação de um dado ponto (observação) em um de muitos grupos. Um caso especial, do qual trata o presente trabalho, é a classificação binária na qual o número de grupos é limitado a dois.

A análise discriminante [Johnson & Wichern, 1988] são técnicas estatísticas multivariadas preocupadas com a separação de conjuntos distintos de pontos e com a alocação de novos pontos aos grupos definidos previamente. Assim, os objetivos da discriminação e classificação, respectivamente, são:

- descrever graficamente (construção de um novo vetor discriminatório) ou algebricamente as características diferenciais dos pontos de duas coleções (populações) conhecidas;
- designar novos pontos para uma das duas classes. A ênfase está em derivar uma regra que possa ser usada para, otimamente, designar um novo ponto para uma das classes.

Neste trabalho, o método de Função Discriminante Linear de Fisher [Johnson & Wichern, 1988], capaz de efetuar a análise discriminante e de classificação foi pesquisado, estudado e aplicado ao exemplo real, descrito posteriormente nos capítulos III e IV.

No Capítulo IV, é feita a aplicação da metodologia descrita aqui a um caso real e é apresentada a construção de um modelo matemático.

O problema da análise dos dados coletados no setor de bioquímica clínica objetiva fazer a discriminação entre duas amostra controle, usadas nos procedimentos de controle de qualidade interno do setor.

Reconhecendo os padrões das variáveis escolhidas, pode-se com esse procedimento reclassificar um novo ponto da “amostra controle”.

Com este objetivo, é apresentada a construção de um novo modelo de programação para ampliação do CEP, com nova técnica estatística, que permita, junto com técnicas já consagradas, obter maior controle da qualidade total.

Finalmente, no Capítulo V, apresentam-se as conclusões e sugestões para futuros trabalhos.

No Anexo, tem-se como o exemplo foi aplicado no *Software Statistica*.

## CAPÍTULO II

### O LABORATÓRIO DE ANÁLISES CLÍNICAS

#### 1. Introdução

A função principal do Laboratório de Análises Clínicas é fornecer informações clínicas a seus clientes para a gerência da saúde dos pacientes. Tal informação tem valor somente se o resultado é preciso, e se seu significado é apreciado pelo médico, de modo que possa ser usado apropriadamente como guia para tomada de decisão.

Os teste bioquímicos são utilizados de forma extensiva na medicina, tanto com relação às doenças que têm um metabolismo básico óbvio (por exemplo, diabetes *mellitus*), como àquelas que provocam mudanças bioquímicas em seu desenvolvimento (por exemplo, falha renal).

Portanto, os exames liberados pelos laboratórios são adotados pelo médico como meio de diagnóstico, prognóstico, monitoramento e pesquisa [Anderson & Cockayne,1993] (Figura 2. 1).

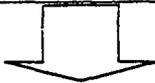
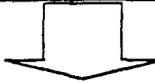
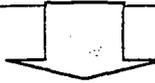
<b>DIAGNÓSTICO</b>	<b>PROGNÓSTICO</b>
 CONFIRMAÇÃO OU REJEIÇÃO DO DIAGNÓSTICO CLÍNICO	 INFORMAÇÃO A RESPEITO DO RESULTADO PROVÁVEL DA DOENÇA
<b>MONITORAMENTO</b>	<b>PESQUISA</b>
 HISTÓRIA NATURAL DA DOENÇA OU RESPOSTA AO TRATAMENTO	 DETECÇÃO DE DOENÇAS QUE AINDA NÃO SE MANIFESTARAM

Figura 2.1. Esquema das possíveis utilizações com os resultados fornecidos por laboratório de análises clínicas.

### 1.1. Diagnóstico

O diagnóstico médico geralmente está baseado na história do paciente e nos sinais clínicos encontrados no exame, mas frequentemente esses achados clínicos podem levar a um diagnóstico diferencial, isto é, há vários diagnósticos possíveis, sem esclarecer a doença causadora do mal. Por isso, as investigações bioquímicas podem ser usadas para fazer distinção entre as doenças.

As investigações podem ser selecionadas para confirmar ou refutar um diagnóstico. Por isso, é importante que o clínico aprecie o quanto é útil o teste escolhido para essas finalidades.

## 1. 2. Prognóstico

O teste usado primeiramente para o diagnóstico pode também fornecer informação prognóstica e alguns testes são usados especificamente para essa finalidade. Por exemplo, as dosagens da concentração crescente do colesterol do sangue são usadas para medir o risco de desenvolver uma doença artério-coronária.

Entretanto, tais riscos de desenvolvimento da doença são calculados através de dados epidemiológicos e não permitem fazer predição precisa para um indivíduo particular.

## 1. 3. Monitoramento

Um dos principais usos dos testes bioquímicos é o de monitorar o transcurso natural de uma doença e/ou os efeitos de algum tratamento. Para fazer isto, deve haver um composto orgânico apropriado, por exemplo, a glicose nos pacientes com diabetes *mellitus*. Os testes podem, também, ser utilizados para detectar alguma complicação no tratamento, tal como a diminuição do sódio no sangue durante tratamento com diuréticos.

Testes especiais são extensivamente usados para detecção de possível intoxicação por drogas de abuso (cocaína, solventes, etc.), e em particular monitorar a quantidade sanguínea de certa droga terapêutica, quando se estabelece seu uso em tratamentos.

## 1. 4. Pesquisa

Os testes bioquímicos podem ser adotados para pesquisar e determinar se o paciente é predisposto a uma dada doença que ainda não se manifestou. O melhor exemplo conhecido é a seleção maciça de todos os bebês recém-nascidos para o teste do pezinho (fenilcetonúria), que é realizado em muitos países. Nessa pesquisa, é descrito o perfil bioquímico dos bebês através de bateria de testes executados, podendo prever e determinar se a criança terá problemas mentais no futuro.

## 2. Procedimentos de Controle de Qualidade

A maioria dos processos na indústria não estão sob controle e, quando analisados pela primeira vez, muitos pontos fora dos limites de controle almejado são encontrados. Nos laboratórios de análises clínicas, essa observação não foge muito à regra, mas, como no ambiente industrial, as causas desses desvios podem ser descobertas e eliminadas.

Nos laboratórios, é clássico trabalhar-se com a presença de dois tipos de variações (erros) que podem ocorrer na área de bioquímica clínica:

- erro ao acaso ou aleatório (EA): erro de causa indeterminada, sendo possível apenas a sua medida através de procedimentos estatísticos; porém, esses procedimentos não viabilizam sua previsão ou correção por medidas dirigidas. Provocam aumento da imprecisão;
- erro sistemático (ES): erro devido a causa determinada, passível de cálculo, previsão e correção. Sempre contribui para a inexatidão. Este erro pode advir de:
  - a) ES experimentais – encontrados quando se dosa uma mesma substância, usando-se dois métodos diferentes;
  - b) ES provocado – são as variações introduzidas inconscientemente pelo próprio analista, por exemplo, na medição de volumes com vidraria não calibrada;
  - c) ES laboratoriais – ocorrem pela deficiência na manutenção dos aparelhos, reagentes, métodos e devidos ao ambiente do laboratório.

Com a implementação de um programa de controle de qualidade, o laboratório pretende proporcionar mecanismos de medir e comparar a precisão e a exatidão de métodos, materiais de medida e instrumentos, detectando a presença dos erros e suas consequências, a fim de indicar medidas corretivas para restabelecer a qualidade do resultado e evitar a incidência e/ou reincidência [Tibúrcio, 1995] dos erros.

As preocupações com os tratamentos desses erros se iniciaram em 1950, quando Levey e Jennings aplicaram os gráficos de controle de Shewhart nos laboratórios de análises clínicas, com o propósito fundamental de identificar lotes analíticos com erros aleatórios e/ou sistemáticos inaceitáveis, tanto do ponto de vista clínico como analítico, dotando os resultados de credibilidade [Tibúrcio, 1995].

Os gráficos de controles foram, na década de 30, aplicados nas indústrias, e consistem num procedimento que faz uso de subgrupos e limites estatísticos simples. Eles podem monitorar o processo através do valor médio, diagnosticando a sua estabilidade. Esta verificação assegura que a distribuição estatística da característica do produto é consistente com as exigências de qualidade [Berwick, 1990].

Dentro do laboratório, é descrito um gráfico controle baseado na média e diferença das duplicatas de mensuramentos realizados em uma única “amostra controle”.

Em 1952, Henry e Segalone [in Tibúrcio, 1995] modificaram o gráfico de controle de Levey-Jennings no qual valores individuais de uma amostra controle são colocados diretamente no gráfico de controle, originando assim um procedimento de controle simplificado. Este tipo de procedimento de controle é o mais comumente usado pelos laboratórios atualmente.

O amplo crescimento da informática permitiu o desenvolvimento de inúmeros programas aplicados aos computadores pessoais para realizar os procedimentos de controle.

## 2. 1. Método clássico de Shewhart [Berwick, 1990; Croiser, 1998]

O método clássico de Shewhart propõe um procedimento para a construção e operação de gráfico de controle para média de um processo caracterizado pela variável  $X$ , com média  $\bar{x}$  e o desvio padrão estimado  $s$ :

a intervalos regulares de tempo  $h$ , tomam-se amostras de  $n$  unidades produzidas:

- mede-se  $x$  para cada uma das  $n$  unidades amostradas e calcula-se a média  $\bar{x}$ ;

- grafa-se  $\bar{x}$  em gráfico seqüencial onde estão demarcados por linhas horizontais os valores de  $M$  e  $M \pm ks$ , onde:

o valor  $s$  é o desvio padrão estimado da média  $\bar{x}$ ;

$k$  é o fator multiplicador de  $s$ . Shewhart recomenda o valor de 1 a 3 para  $k$ , pois este procedimento leva à construção de limites correspondem a 68,2 %; 95,0 % e 99,7 %, respectivamente, da área sobre a curva da distribuição normal de Gauss, na Figura 2.2 podemos ver a ilustração gráfica dessa relação.

- enquanto se obtiverem valores de  $\bar{x}$  entre os limites, que são o limite inferior e superior de controle ( $\bar{x} \pm 3s$ ), assume-se que o processo está calibrado e em estado de controle estatístico.

Ao estabelecer este roteiro, Shewhart pretendia compor um esquema simples e robusto, adequado às limitações técnicas e culturais da época de 1931. A grande simplicidade do modelo clássico de Shewhart permitiu sua rápida difusão no sistema produtivo americano e, em seguida, em todo o mundo industrializado. Sua operacionalização, no chão de fábrica, é bastante simples e pressupõe poucos conhecimentos estatísticos.

### 2. 1. 1. Interpretação dos gráficos construídos de Shewhart

O modelo clássico de Shewhart não assume uma regra de decisão, onde a obtenção de um valor de  $x$  fora dos limites de controle caracteriza um processo não estável, mas é auxiliado pelos testes de seqüência conhecidos como “Run Tests”.

Se o estado de controle é alcançado no processo, adota-se o gráfico de controle para monitorar as observações atuais e futuras; em caso negativo, conduzem-se ações de melhoria até que seja atingido o nível de qualidade desejado para o processo, os limites de controle são recalculados e os pontos fora destes limites devem ser revistos constantemente, para que seja possível manter os valores dos limites sempre atualizados.

Na figura 2.2, estão os limites de controle, com o valor de  $k$  variando de 1 a 3, formando zonas entre os limites, descritas como A, B e C [Croiser, 1998; Cembrowski & Sullivan, 1996].

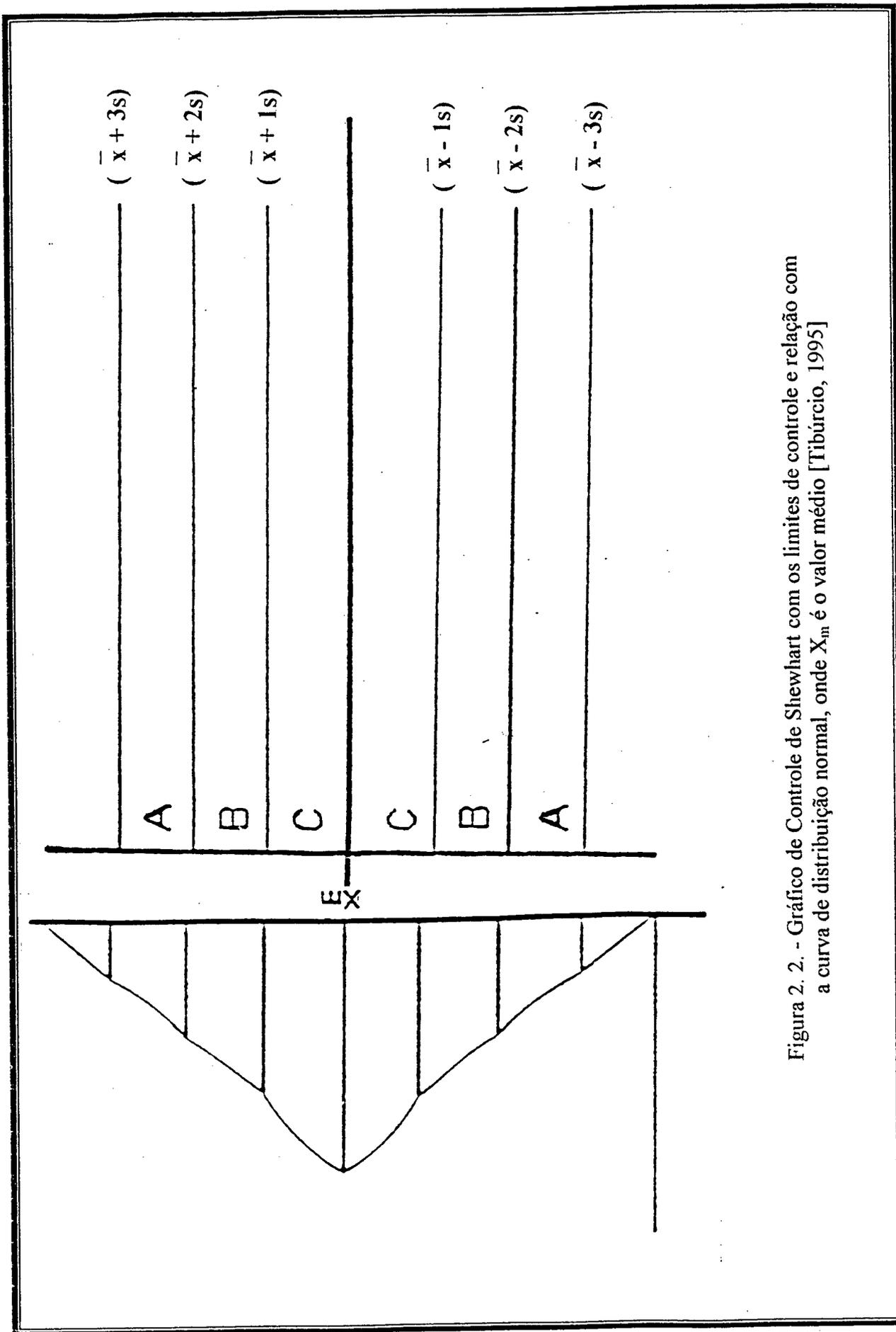


Figura 2. 2. - Gráfico de Controle de Shewhart com os limites de controle e relação com a curva de distribuição normal, onde  $X_m$  é o valor médio [Tibúrcio, 1995]

Segundo a análise do gráfico subdividido em zonas, quando ocorrer qualquer uma das seis configurações descritas abaixo o processo será considerado fora de controle e deve ser regulado:

- um ponto acima da zona A;
- nove pontos sucessivos de um único lado da zona C, ou além dela;
- seis pontos sucessivos aumentando ou diminuindo constantemente;
- dois a três pontos sucessivos na zona A;
- quatro a cinco pontos sucessivos na zona B;
- oito pontos sucessivos acima ou abaixo da linha central fora da zona C.

## 2. 1. 2. Alguns aspectos importantes na construção e utilização dos gráficos controles de Shewhart

A teoria subjacente aos gráficos de controle de S afirmam que:

- se todos os pontos estiverem dentro dos limites de controle e nenhuma configuração especial estiver evidente, podemos dizer que o processo esteve sob controle no passado, então os limites de controle experimental são apropriados para controlar a produção atual e futura;
- se algum ponto está fora do limite de controle, estes pontos devem ser removidos e os limites experimentais devem ser recalculados, para que se possa monitorar a produção atual e futura;
- para que o processo possa ser considerado sob controle estatístico, é necessário que ele esteja estável em relação à média e à variabilidade dos pontos nos gráficos controles, sem ferir, portanto, quaisquer configurações estabelecidas acima;
- se houver mudanças de maquinaria, matéria-prima, operador ou outras mudanças dessa natureza, os limites de controle devem ser reavaliados.

### 2.3. Fases e Etapas dos Procedimentos de Controle [Tonks, 1995; Friedrich et al, 1998]

Para atingir um nível de alta confiabilidade, o laboratório deve assegurar a obtenção de resultados com baixa variabilidade analítica, em tempo útil para influenciar as decisões médicas.

Com esse objetivo é descrito todo um programa de controle de qualidade interno para os laboratórios de análises clínicas, com o qual espera-se que após sua implantação o sistema permita ao mesmo prevenir, detectar, classificar e corrigir os erros.

A implantação do sistema segue como linha geral o pensamento de Shewhart, com algumas peculiaridades para os laboratórios, e se distingue em três fases.

#### 2.2.1 - Fase inicial

Esta primeira fase do procedimento de controle consiste em:

- identificar qual é a *performance* atual existente no laboratório para um determinado procedimento analítico utilizado na avaliação de um composto, qualitativa ou quantitativamente. Utilizam-se para esta determinação os resultados obtidos pelo laboratório anteriormente, ou por outro laboratório que use a mesma metodologia de dosagem.
- estabelecer os limites da qualidade analítica, utilizando para tal uma “amostra controle” de valor conhecido.

#### “Amostra Controle”

É um artifício contendo compostos biológicos com valores pré-determinados, com o propósito de estabelecer o controle interno da qualidade analítica.

A “amostra controle” é submetida à metodologia analítica, com o objetivo de monitorar a precisão do processo, ou seja, a dispersão entre os valores obtidos após sucessivas

dosagens em relação ao valor médio; e a exatidão da metodologia, que consiste na distância entre o valor obtido por uma série de dosagens e o valor real da concentração.

Se a estabilidade e a precisão do processo estão adequados, então pode-se começar a processar as amostras biológicas dos clientes.

Os gráficos controles são confeccionados a partir do registro dos valores encontrados nas dosagens de um composto na “amostra controle”.

- Etapa 1

A “amostra controle” é avaliada utilizando o mesmo procedimento das outras amostras biológicas dos pacientes. O valor obtido é registrado na Folha de Registro (Figura 2.3.), que tem dois campos: a identificação e a planilha de registro.

No campo da identificação, descreve-se o composto pesquisado, por exemplo, glicose; a metodologia que o laboratório utiliza para avaliar o composto e, ainda, a identificação da “amostra controle” usada, marca comercial e lote, bem como os dados da pessoa responsável pela execução do procedimento.

A planilha de registro compreende um campo composto pelo número de avaliações ( $n$ ) que são realizadas, a data da realização, e ao lado é anotado o valor encontrado do processo.

Esse procedimento continua até que sejam obtidos, conforme o estabelecido na literatura da área de controle estatístico de processo, pelo menos 20 valores, quando será estimada a média ( $\bar{x}$ ) do conjunto [Tonks,1995; Hainline, 1993].

## Identificação

Analito	Metodologia	Amostra controle	Responsável

## Planilha de registro

n	Data	$X_i$	$X_i - X_m$	$(X_i - X_m)^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
$\Sigma$	//////////		//////////	

Figura 2.3 – Folha de Registro

- Etapa 2

Após obter o valor médio  $\bar{x}$  da “amostra controle”, pode-se fazer o cálculo da estimativa do desvio padrão. Para tanto, pode-se usar a equação abaixo:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (2.2.1)$$

onde:

S = desvio padrão calculado;

n = número de observações;

$x_i$  = valor da dosagem, e

$\bar{x}_i$  = valor da média calculado

Essa é a maneira de calcular manualmente a estimativa do desvio padrão. Os *softwares* estatísticos podem ser utilizados, hoje, com muito mais propriedade.

- Etapa 3

Com os resultados dos cálculos das estimativas da média e do desvio padrão, estabelecem-se os “limites da qualidade” utilizando os valores ( $\bar{x} \pm 1S$ ), ( $\bar{x} \pm 2S$ ) e ( $\bar{x} \pm 3S$ ).

Estes limites correspondem a 68,2 %, 95,0 % e 99,7 %, respectivamente, da área sobre a curva da distribuição normal de Gauss.

Ora, como o erro aleatório tanto pode ser positivo como negativo em relação à média, e sua magnitude não pode ser prevista, os limites da qualidade são portanto superior (+) e inferior (-) no intervalo de 1s a 3s. Acima dessa variabilidade, os resultados são inaceitáveis, pois mostram que o procedimento está fora dos limites de qualidade estabelecida, indicando a necessidade de medidas corretivas.

Neste ponto, está encerrada a fase inicial do processo de controle interno de qualidade e terá início nova fase denominada de intermediária. Os valores dos limites da qualidade analítica foram obtidos e serão repassados para o gráfico de controle e este será o passo seguinte.

### 2. 2. 2. Fase intermediária

Nesta fase será introduzida a construção do gráfico controle. Também as regras de controles serão acionadas, para julgar o valor fornecido com a “amostra controle” e, conseqüentemente, a aceitabilidade ou não do procedimento analítico.

- Etapa 1

O gráfico de controle é um método gráfico, tal como sugerido por Shewhart [Tibúrcio, 1995]. Nos laboratórios de análises clínicas, eles são construídos com os seguintes características:

- a escala do eixo X corresponde à periodicidade de avaliação ou ao dia da dosagem, durante um mês de trabalho ou mais tempo;

- a escala do eixo Y deverá cobrir o intervalo dos valores calculados na fase inicial (média e desvio padrão), que são valores extremos encontrados da dosagem da “amostra controle”. É conveniente que a escala suporte valores menores e maiores do que os valores extremos, pois estes podem ocorrer.

No gráfico são demarcadas as linhas correspondentes à média aritmética do composto controlado e os limites superiores e inferiores para  $1s$ ,  $2s$  e  $3s$ , respectivamente.

Com essas ações, o gráfico de controle está pronto para receber o valor do composto avaliado na “amostra controle”.

- Etapa 2

Após o lançamento dos valores no gráfico, é feita a interpretação de acordo com as Regras do Controle. Os resultados das amostras biológicas dos clientes só serão dosadas e liberadas com a interpretação das Regras.

A base teórica da interpretação do gráfico de controle é a curva de distribuição normal de Gauss, como sugerido por Shewhart.

Quando a *performance* analítica é estável, existindo apenas o erro aleatório inerente, a curva de distribuição é homogênea e o valor da média é o ponto central. A instabilidade da *performance* será vista pelo deslocamento da média, quando o problema é exatidão do processo, e pela perda da homogeneidade da curva, decorrente da maior dispersão dos valores, quando o problema é originário da precisão das medidas.

Nesta parte entram as Regras do controle, que nos laboratórios de análises clínicas são conhecidas como Regras Múltiplas de Westgard. A utilização dessas regras foi iniciada em 1977 e foram pouco a pouco recebendo adesões, sendo hoje muito usadas e recomendadas pela Federação Internacional de Química Clínica e pela Organização Mundial da Saúde [WHO, 1992].

Descrição das regras múltiplas [Westgard, 1992]:

As regras utilizam a seguinte notação:  $A_L$ , na qual a letra “A” é uma abreviatura para determinada estatística ou número de ponto plotado no gráfico controle e “L” é um limite de controle ou limite de qualidade onde esse(s) ponto(s) se encontra(m).

- $1_{2s}$

Um valor excede os limites da qualidade de  $\pm 2s$  (limites aceitáveis). Historicamente, isto foi o limite de alerta no gráfico de Shewhart, e no laboratório é conhecido como limite de rejeição no gráfico de Levey-Jennings.

O ponto é interpretado como uma exigência para adicionais avaliações com as demais regras seguintes.

- $1_{3s}$   
Um valor excede a  $\pm 3s$  (limites de alerta). Este ponto é o limite de ação ou de rejeição, recomendado por Shewhart e similarmente utilizado por Levey-Jennings.  
  
Esta regra é a principal indicadora de erros aleatórios, mas pode também indicar erros sistemáticos grosseiros.
- $2_{2s}$   
Dois pontos consecutivos excedem o limite de  $+2s$  ou  $-2s$ . Quando esta regra é rompida, sua interpretação dependerá da quantidade de “amostras controle” que são usadas: se são duas de concentrações diferentes, esta regra é indicadora de erros sistemáticos ocorrendo no intervalo; se é de uma concentração, ela é indicadora de erro sistemático.
- $R_{4s}$   
É anotação utilizada para indicar que ocorreu a faixa R, isto é, a diferença entre uma medição de valor alto e outra, de valor baixo, excede a  $\pm 4s$  ou quando um ponto no controle excede a  $+2s$  e em outra medida a  $-2s$ . Há indicação de erros aleatórios.
- $4_{1s}$   
Quatro valores consecutivos excedem o mesmo limite de  $\pm 1s$  (limite aceitável); é a principal indicadora de erros sistemáticos.
- $7_{\bar{x}}$   
Sete pontos consecutivos são obtidos em um mesmo lado da média: é indicativo de erros sistemáticos.
- $7_t$   
Sete valores consecutivos mostrando tendência para crescimento ou decrescimento. Esta regra é conhecida como regra das tendências e é indicadora de erros sistemáticos intensos.

Através dessas regras múltiplas de Westgard, criou-se um fluxograma a fim de auxiliar e agilizar a interpretação das regras de controle, fornecendo também a ordem de uso. O nomograma está representado na Figura 2.4.

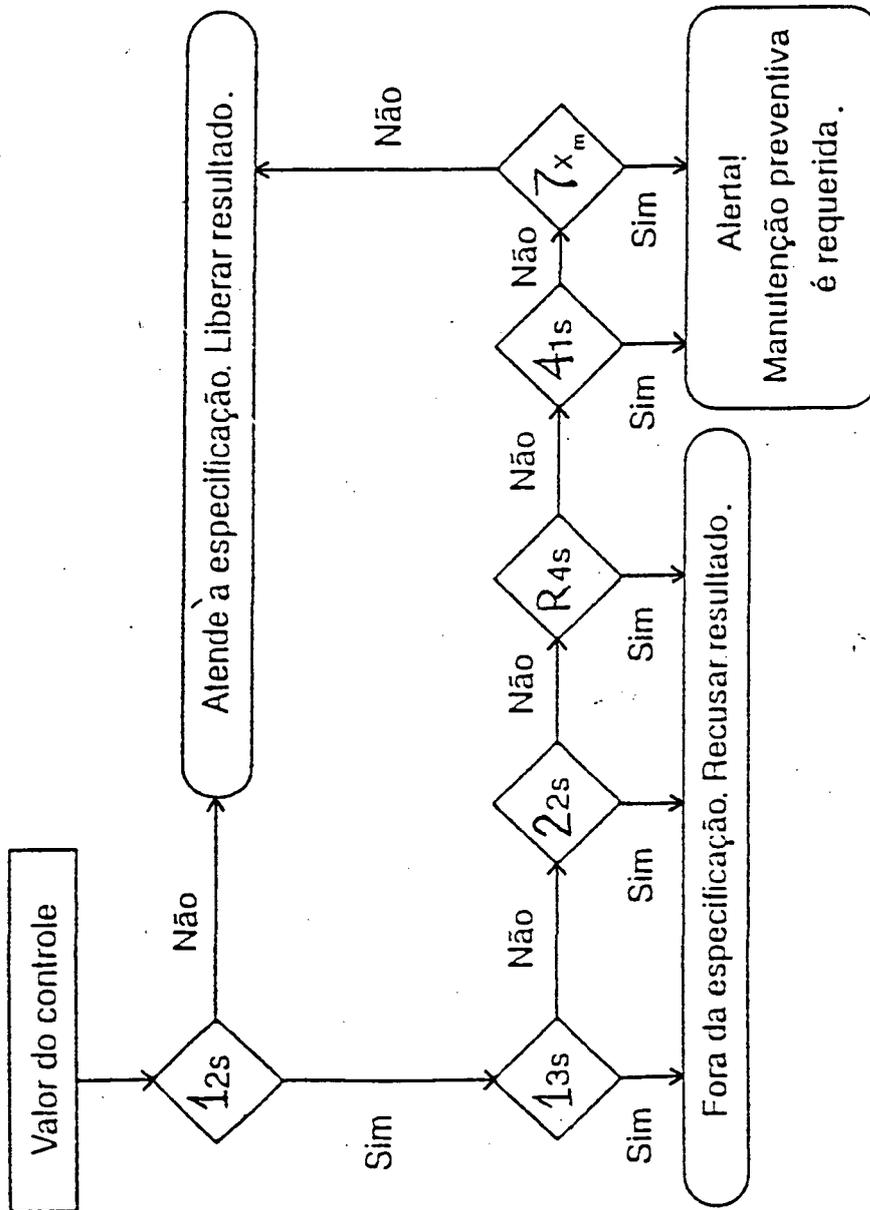


Figura 2.4 – Fluxograma para uso com as regras múltiplas [Tibúrcio, 1995].

Terminada esta fase, é hora de iniciar a fase acumulativa.

### 2. 2. 3. Fase *acumulativa*

A dinâmica desta fase chamada *acumulativa* está centralizada no uso diário da “amostra controle” antes das dosagens das amostras biológicas. Este procedimento contínuo irá criar novas baterias de dados, que podem ser mais realistas do que somente aquelas obtidas em um único mês. Essa nova matriz pode conduzir à verificação de se o procedimento analítico está operando com adequada estabilidade da precisão.

Esta fase é considerada uma das bases de sustentação do próprio programa de qualidade, por se alinhar com a busca da melhoria contínua de seus procedimentos analíticos, onde procura-se o aperfeiçoamento dos processos dentro dos laboratórios.

## CAPÍTULO III

### MÉTODO ESTATÍSTICO MULTIVARIADO DE ANÁLISE DISCRIMINANTE (FUNÇÃO DISCRIMINANTE LINEAR DE FISHER)

#### 3.1. Métodos estatísticos multivariados [Silva & Machado, 1996; Mardia *et al.*, 1979]

A análise multivariada compreende amplo conjunto de métodos e procedimentos para a descrição, exploração e análise de dados que constituem medidas de variáveis que representam mais de uma característica das unidades de amostras de uma ou mais populações sob consideração. Pode-se argumentar que quase toda pesquisa, experimental ou observacional, gera dados dessa natureza. Os métodos multivariados são particularmente apropriados quando as variáveis se relacionam, estabelecendo-se entre essas variáveis uma estrutura de dependência relevante.

Há muitos métodos multivariados que podem ser aplicados à análise de dados. O mais adequado para uma situação depende dos objetivos da análise e da forma dos dados disponíveis. Em geral, um objetivo de uma análise multivariada é a simplificação, ou seja, buscar obter o resumo de um grande conjunto de dados através de alguns poucos parâmetros

que o descrevem adequadamente. Nessas circunstâncias, muitas das técnicas multivariadas são exploratórias e visam a gerar hipóteses e não a testar hipóteses.

A análise multivariada enfoca um conjunto de variáveis coletivamente, enquanto que a análise univariada as considera individualmente. Dessa caracterização, poder-se-ia depreender que os métodos multivariados são extensões dos métodos univariados. De fato, em geral, para cada método univariado há um análogo método multivariado. Entretanto, há métodos que são especificamente multivariados. Essa observação provê uma base para classificação dos métodos multivariados em duas categorias:

- a) Extensões de métodos univariados:
  - Análise de regressão multivariada;
  - Análise de variação multivariada;
  - Análise de covariação multivariada;
  - Análise de correlação canônica.
  
- b) Métodos especificamente multivariados:
  - Análise de componentes principais;
  - Análise fatorial;
  - Análise de agrupamento;
  - Análise discriminante;
  - Análise de correspondência [Silva, 1996].

Uma outra classificação dos métodos multivariados baseia-se nos números de populações e conjuntos de dados sob consideração. Essa classificação é apresentada na tabela de dupla entrada da Figura 3. 1.

POPULAÇÕES	
UMA	DUAS OU MAIS
<p>UM</p> <p>Análise de componentes principais Análise fatorial Análise de agrupamento Análise de correspondência</p>	<p>Análise multivariada da variação Análise discriminante</p>
<p>DOIS OU MAIS</p>	<p>Análise da covariação multivariada</p>
<p>CONJUNTOS DE VARIÁVEIS</p>	

Figura 3. 1. – Classificação dos métodos multivariados, segundo o número de populações e o número de conjuntos de variáveis em consideração [Silva, 1996].

### 3. 2. Análise Discriminante

A terminologia "discriminar" foi introduzida na Estatística em 1930 por Ronald A. Fisher no primeiro tratamento moderno dos problemas de separação de conjuntos e tem sido utilizada em uma variedade de disciplinas desde sua primeira aplicação, quando era usada principalmente em ciências biológicas e comportamentais [Johnson & Wichern, 1988].

A análise discriminante permite, solucionar estatisticamente problemas que surgem com frequência em ciências sociais, como, por exemplo, a classificação de observações, com base em suas propriedades individuais [Röehrig et al., 1993].

O método discriminante tem como motivação básica a construção de uma combinação linear das propriedades diferenciadoras dos grupos. Dadas duas populações 1 e 2 de observações multivariadas  $X$ 's, de dimensão  $n$ , a idéia de Fisher foi transformar essas observações multivariadas em observações univariadas  $Y$ 's de tal modo que estejam separadas tanto quanto possível [Bennett & Mangasarian, 1992].

O método de análise está fundamentado nas variáveis discriminantes que medem características diferenciadoras dos grupos, com a ajuda de uma visualização num plano apropriado, propiciando a distinção entre dois ou mais grupos de observações estabelecidas *a priori*.

O procedimento da análise consiste em testar a hipótese de igualdade dos valores das médias dos grupos (centróides) e das matrizes de covariância dos dois grupos definidos e distintos. Se for constatada a diversidade entre os elementos padrões de cada grupo, pode-se calcular um coeficiente discriminante ( $\lambda$ ) para cada variável. A seguir, computa-se um índice crítico de separação ( $Z$ ). Com esses dois aspectos da Análise Discriminante, pode-se alocar novas observações aos grupos.

O método, então, determina um valor crítico do índice ( $Z$ ), de forma que, se uma observação se situa abaixo deste, ela pertence a um grupo, e, se acima, pertence a outro. Havendo poucas superposição das distribuições de medidas, torna-se mais fácil a classificação

das observações em grupos. Essa combinação linear das propriedades diferenciadoras dos grupos são usadas para gerar um *número índice*, por meio do qual se pode diferenciar a que grupo pertence cada observação, com alta percentagem de sucesso [Bouroche & Saporta, 1982; Silva & Machado, 1996].

Como as observações advêm de um grupo já estabelecido, é de se esperar que este dado seja classificado (alocado) através da combinação linear ótima no seu grupo de origem. Caso a observação seja classificada, erroneamente, em outro grupo ao qual ela não pertença *a priori*, pode-se concluir que houve alguma alteração no processo [Martins & Samanez, 1987].

A análise discriminante difere da análise de regressão somente quanto à natureza da variável dependente. Na função de regressão usam-se valores da variável dependente a fim de se determinar uma função linear para prever valores desta própria variável; na função discriminante, a variável dependente não assume valores quantitativos, mas sim dados categóricos, ou qualitativos, do tipo dicotômico (1 ou 2), por exemplo, homem ou mulher, empresa falida ou não, duas marcas diferentes de “amostra controle”.

Formalmente, é representada por:

$$Z_i = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \dots + \lambda_k x_k \quad (3.2.1)$$

onde:

$Z_i$  = índice diferenciador das observações dos grupos (escore da função discriminante);

$\lambda$  = coeficiente ponderado;

$X$  = valores padronizados quantitativos.

Procura-se determinar os  $\lambda$  de modo tal que o valor médio de  $Z$  para um grupo esteja a uma distância máxima normalizada da média  $Z$  do outro grupo. O valor de  $Z$  nos servirá como um indicador para diferenciar as observações dos dois grupos, e é, um plano ou linha

linha divisória, encontrado a partir da equação (3. 2. 1). O  $Z_i$  é o escore (ou índice) da função discriminante  $i$ . Os  $\lambda$  são os coeficientes ponderados, e os  $x$  são os valores padronizados das  $k$  variáveis discriminantes usadas na análise. Quanto maior o coeficiente  $\lambda_k$  da variável  $x_k$ , sem levar em consideração o sinal, mais importante é a variável para discriminar dois grupos.

Um exemplo para elucidar mais a importância de se estabelecer esse índice é o caso de um organismo financeiro que tenta alocar da melhor maneira possível a massa de crédito de que dispõe. Ele procura logicamente dar seus empréstimos a tomadores que têm a mais forte probabilidade de serem bons clientes e rejeita aqueles que têm boa chance de acabar na justiça. Cada candidato ao empréstimo deve preencher um dossiê do qual se extraem os valores padronizados quantitativos ( $x$ ).

A análise discriminante permite então elaborar os coeficientes ponderados ( $\lambda$ ). Com base nesses valores, pode-se escrever o escore da função discriminante ( $Z_i$ ). Neste caso, o método cria uma regra de decisão utilizada num segundo plano para selecionar os bons tomadores [Bouroche, 1982].

Para simplificar a discussão geométrica do problema, considerem-se duas variáveis com  $n_1$  e  $n_2$  observações, respectivamente, nos dois grupos. A equação:

$$\lambda_1 x_1 + \lambda_2 x_2 - Z = 0 \quad (3.2.2)$$

representa, então, um plano em três dimensões passando pela origem e tendo como parâmetros os  $\lambda_1$ ,  $\lambda_2$  e  $-1$ . A figura 3.2 mostra dois conjuntos, com  $n$  observações cada, que são separados por um vetor. Os valores de  $Z$  correspondentes a cada observação são indicados por pontos cheios.

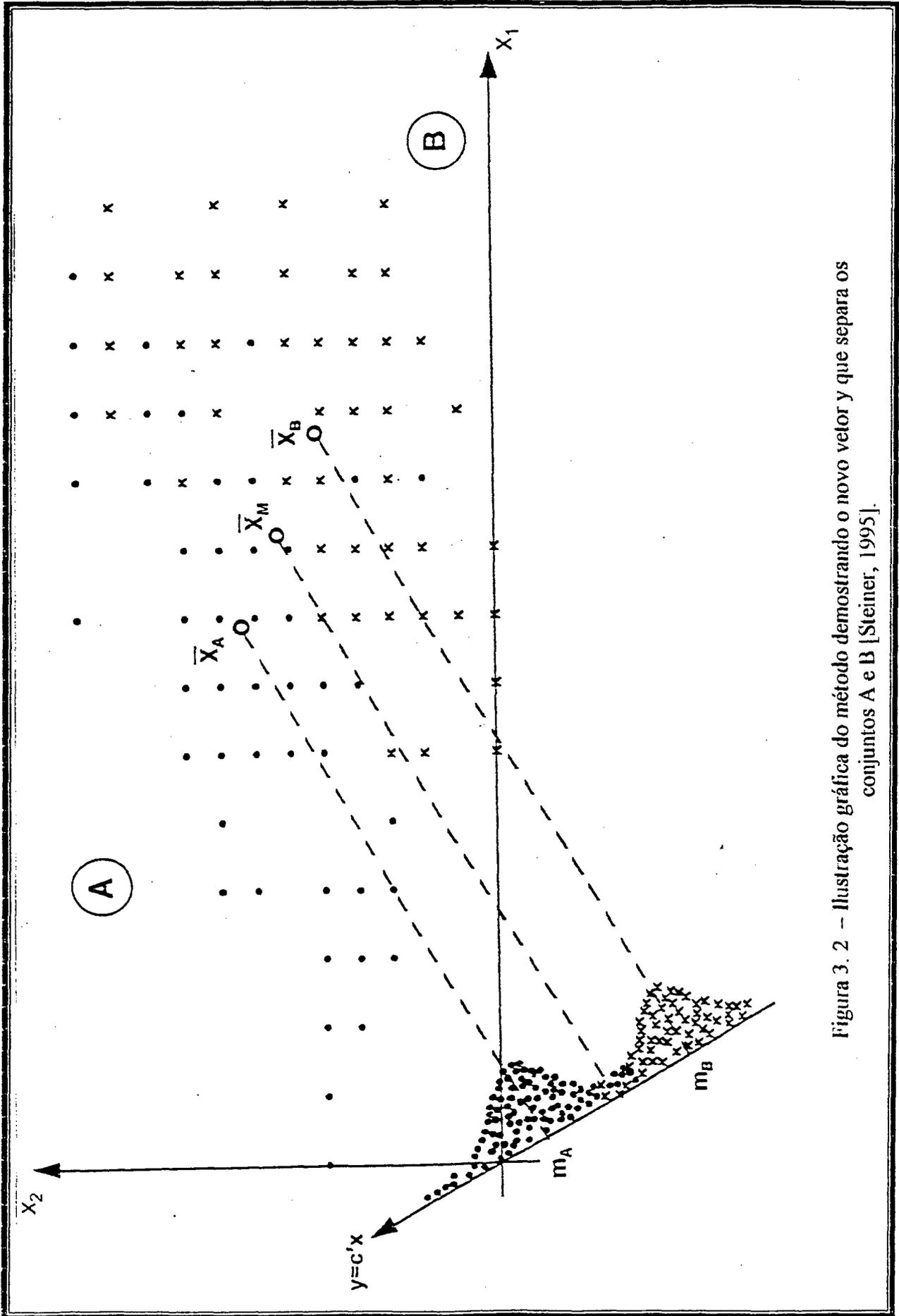


Figura 3. 2 - Ilustração gráfica do método demonstrando o novo vetor  $y$  que separa os conjuntos A e B [Steiner, 1995].

Se os dois conjuntos de pontos correspondentes a dois grupos de observações podem ser separados por meio de um plano passando pela origem (cada lado representando um grupo de observações), os valores de  $Z$  de cada grupo tornam-se crescentes, tanto positivos quando negativos. À proporção que os valores de  $Z$  crescem, as variações tornam-se, ao mesmo tempo, maiores em ambos os grupos, e há um aumento no desvio padrão; isso acontece, mesmo que a média da distribuição permita a separação entre os grupos, não necessariamente limites para a inclinação do plano, a fim de que a técnica seja eficaz.

É, portanto, desejável escolher um plano tão amplo quanto possível relativo à variação dos valores de  $Z$  dentro dos dois grupos, que separe esses valores deles. Como uma medida de separação, utiliza-se  $(\bar{Z}_1 - \bar{Z}_2)^2$ , onde  $\bar{Z}_1 - \bar{Z}_2$  são as médias respectivas de cada um dos dois grupos. Como medida de variação dos valores de  $Z$  dentro dos grupos, deve-se ter:

$$\sum_{i=1}^2 \sum_{j=1}^{n_i} (z_{ij} - \bar{Z}_i)^2 \quad (3.2.3)$$

Na Eq. 3. 2. 3.,  $Z_{ij}$  denota o valor  $Z$  da  $j$ -ésima observação no  $i$ -ésimo grupo, onde  $i=1$  ou  $2$ .

- Método

Sendo  $\bar{y}_1$  a média dos  $Y$ 's obtida dos  $X$ 's pertencentes à população 1 e  $\bar{y}_2$  a média dos  $Y$ 's obtida dos  $X$ 's pertencentes à população 2, seleciona-se a combinação linear que maximiza a distância quadrática entre  $\bar{y}_1$  e  $\bar{y}_2$  relativamente à variabilidade dos  $Y$ 's, ou seja:

$$(\bar{y}_1 - \bar{y}_2)^2 / S_y^2 \quad (3.2.4)$$

sendo :

$\bar{y}_1$  = valor esperado de uma observação multivariada de 1;

$\bar{y}_2$  = valor esperado de uma observação multivariada de 2;

e supondo que a matriz de covariância  $\mathbf{A} = E(\mathbf{X} - \bar{\mathbf{x}})(\mathbf{X} - \bar{\mathbf{x}})'$ ,  $i = 1, 2$  seja a mesma para 1 e 2, tem-se a combinação linear :

$Y = \mathbf{c}' \mathbf{X}$ , em que Y é uma variável analítica unidimensional

onde:

$\mathbf{c}'$  é matriz de ordem 1 x n;

$\mathbf{X}$  é matriz de ordem n x 1

Mas Y tem média

$$\bar{y}_1 = \mathbf{c}' \bar{\mathbf{x}}_1$$

ou

$$\bar{y}_2 = \mathbf{c}' \bar{\mathbf{x}}_2$$

(3.2.5)

dependendo da população, mas sua variância

$$S_y^2 = \text{Var}(\mathbf{c}' \mathbf{X}) = \mathbf{c}' \text{Cov}(\mathbf{X}) \mathbf{c} = \mathbf{c}' \mathbf{A} \mathbf{c} \quad (3.2.6)$$

$\mathbf{c}$  é a mesma para ambas as populações.

Substituindo (3.2.5) e (3.2.6) em (3.2.4), tem-se que :

$$(\bar{y}_1 - \bar{y}_2)^2 / S_y^2 = (\mathbf{c}' \bar{\mathbf{x}}_1 - \mathbf{c}' \bar{\mathbf{x}}_2)^2 / \mathbf{c}' \mathbf{A} \mathbf{c} = (\mathbf{c}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2))^2 / \mathbf{c}' \mathbf{A} \mathbf{c} \quad (3.2.7)$$

Os coeficientes da combinação linear  $\mathbf{c}' = [c_1, c_2, \dots, c_n]$  são aqueles que maximizam a razão (3.2.7).

Esta razão é maximizada pela escolha de  $\mathbf{c} = k \mathbf{A}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  para qualquer  $k \neq 0$ . Escolhendo  $k = 1$ , obtém-se a combinação linear :

$$\lambda = \mathbf{c}' \mathbf{X} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{A}^{-1} \mathbf{X} \quad (3.2.8)$$

que é a expressão da Função Discriminante Linear de Fisher (FDL) populacional.

Seja agora

$$y_0 = (\bar{x}_1 - \bar{x}_2)' A^{-1} x_0$$

o valor da função discriminante para uma nova observação  $x_0$  e seja

$$m = 1/2 (\bar{x}_1 - \bar{x}_2)' A^{-1} (\bar{x}_1 - \bar{x}_2)$$

o ponto médio entre as médias das duas populações univariadas. Tem-se:

$y_0 - m \geq 0$ , se  $x_0$  pertencer à população 1,  $y_0$  é maior do que  $m$  e

$y_0 - m < 0$ , se  $x_0$  pertencer à população 2,  $y_0$  é menor do que  $m$ .

Analogamente, chega-se à FDL de Fisher amostral, considerando-se as amostras A de tamanho  $m$  e B de tamanho  $k$ , respectivamente, das populações 1 e 2. Dessa forma, tem-se :

$$Y = (x_A - x_B)' A^{-1} x \quad (3.2.9)$$

onde :

$x_A$  = vetor ( $n \times 1$ ) médio amostral da população 1;

$x_B$  = vetor ( $n \times 1$ ) médio amostral da população 2;

$x$  = vetor das variáveis aleatórias correspondentes as características populacionais observadas.

$A^{-1}$  = inversa da matriz de covariância amostral conjunta :

$$A = [(m-1)S_A + (k-1)S_B] / (m+k-2);$$

$S_A$  = matriz de covariância amostral de 1;

$S_B$  = matriz de covariância amostral de 2;

Para classificação de um novo vetor  $x_0$ , a regra de decisão é a seguinte :

se  $x_0$  pertence ao grupo 1, então  $y_0 = (\bar{x}_A - \bar{x}_B)' A^{-1} x_0 \geq m = 1/2 (\bar{x}_A - \bar{x}_B)' A^{-1} (\bar{x}_A - \bar{x}_B)$

se  $x_0$  pertence ao grupo 2, então  $y_0 < m$

Com esses parâmetros e combinação linear ótima estabelecida, pode-se aplicá-la facilmente dentro do laboratório, pois o analista somente terá que comparar as novas observações utilizando o método de classificação.

### 3. 3. Métodos de análise da qualidade da discriminação

Uma vez determinada a função discriminante, pode-se proceder à análise e classificação dos dados, objetivo desta técnica, fornecendo vários subsídios para interpretação dos mesmos. Entre esses subsídios estão testes estatísticos para medir a eficácia com que as variáveis (ou características das observações) diferenciam os dados dos grupos quando combinadas em uma função discriminante [Martins & Samané, 1987; Berthoyex et al, 1988].

O procedimento da classificação, conforme já foi observado, consiste em delimitar um valor crítico de um índice  $Z$  (linha ou plano separatriz), de tal forma que, se um valor situa-se abaixo desse índice é classificado em um grupo, e se acima do mesmo, em outro. Essa observação não é de todo absoluta, pois sempre se pode encontrar uma observação fora da região correspondente ao grupo definitivo. Isso, pode induzir a erros de classificação, os quais serão menores quanto mais diferentes forem os grupos. Os erros serão do tipo I, ou seja, as observações são classificadas no grupo 2 quando deveriam constar do grupo 1, e do tipo II, caso as observações sejam classificadas no grupo 1 quando deveriam constar do grupo 2.

Devido ao aparecimento desses erros, explicitados acima, deve-se medir a significância estatística da classificação, para saber se a função discriminante dividiu com êxito a amostra em dois grupos distintos, através de três maneiras:

- a) teste  $F$ : retrata uma distribuição  $F$  e é usado para verificar se a variância de uma população é igual à variância de uma outra população, como também para comparar duas ou mais médias simultaneamente, denominada como análise da variância. O objetivo é calcular um valor crítico de  $F$  identificados por dois parâmetros: o grau de liberdade do numerador e o grau de liberdade do denominador e comparar com o valor tabelado do  $F$ . Se o resultado deste teste é elevado, há indícios de estar ocorrendo uma diferença significativa entre as médias

grupais, possibilitando rejeição da hipótese de que as observações derivam de uma mesma população.

$$F = \frac{[n_1 n_2 / (n_1 - n_2)] D^2 / k}{D / (n - k - 1)} \quad (3.3.1)$$

com  $k$  e  $(n - k - 1)$  graus de liberdade

onde:

$D = \bar{Z}_1 - \bar{Z}_2 =$  diferença dos escores médios de cada grupo;

$k =$  número de variáveis;

$n_1 =$  número de observações no Grupo I;

$n_2 =$  número de observações no Grupo II; e

$n =$  número total de observações.

- b) teste *Lambda de Wilks*: estatística geral para lidar com o teste de centróides, a qual indica que quanto mais próximo seu valor estiver de zero, significa que os grupos estão perfeitamente discriminados, e quando mais próximo a um, estão menos discriminados.
- c) tabela de classificação: mostra se os dados dos dois grupos foram identificados cada qual no seu grupo de origem, propiciando verificar uma indicação do grau de confiança da discriminação. Essa tabela é apresentada na tabela 4.3 (Martins & Samané, 1987; Pereira & Ness, 1980).

Por exemplo, se a tabela mostrar grande número de observações classificadas em outro grupo que não seja o seu de origem, pode-se inferir que os dois grupos possuem os valores de média e matrizes de covariância iguais. Portanto, o método não diferencia o grupo 1 do 2, visto que as variáveis pesquisadas são fracamente discriminantes.

Caso contrário, quando a tabela traz poucos números (3 ou menos) de observações classificadas erroneamente, verifica-se que a combinação linear é ótima e a metodologia soube diferenciar o perfil de cada grupo.

A tabela de classificação é também chamada *medida de eficiência total* por ser a mais completa e por mostrar a precisão com que as observações são classificadas. Essa medida de eficiência total é calculada por:

$$(n_1 + n_2)/n \quad (3.3.2)$$

onde:

$n_1$  = número de observações classificadas corretamente no Grupo I;

$n_2$  = número de observações classificadas corretamente no Grupo II;

$n$  = número total de observações.

O resultado final de uma perfeita discriminação através dessa medida tem que ser igual a 1: isto ocorre quando o número de observações classificadas corretamente em cada grupo é igual ao número total. Os valores classificados incorretamente num grupo não entram no cálculo.

### 3.4. Recursos Computacionais

Os cálculos envolvidos no método estatístico multivariados de análise discriminante (função discriminante linear de Fisher) são facilitados pelo uso de sistemas de programas de análise estatística que implementam facilidades para o tratamento de dados e, especialmente, para a execução de análises apropriadas. Os mais conhecidos e completos são o Statistica, Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS). Nesta dissertação, considerar-se-á a utilização do pacote computacional Statistica.

A disponibilidade desses recursos de computação diminui a necessidade de o analista ater-se a problemas referentes apenas a aspectos computacionais, particularmente a procedimentos e rotinas de cálculos [Lowry et al, 1992].

## **CAPÍTULO IV**

### **MÉTODO PROPOSTO**

O melhoramento da qualidade interna dos laboratórios tem sido um desafio constante ao corpo técnico do setor e objetivo de estudo de pesquisadores por todo mundo. Pesquisadores apontam em seus trabalhos soluções mais preocupados com o aspecto administrativo do que com o aspecto técnico. Um exemplo é o próprio Westgard (1992), que se atém ao problema pontual das práticas dos sistemas de decisões, construindo as base das regras de controle.

Neste plano, também, surge Friedrich et al (1998), que examina a aplicação de procedimentos de decisão como métodos de aceitação ou não dos resultados obtidos nas dosagens, tomando as regras de controle como indicadoras da necessidade de adotar medidas corretivas. Outros autores tratam do problema mais teoricamente do que voltados para a prática.

Neste trabalho é proposto a construção de um modelo matemático para o controle de qualidade interno, não rejeitando o método antigo e já consagrado, mas para junto com ele estabelecer uma melhoria no sistema de controle.

#### 4.1. Coleta, seleção e padronização da amostra

O critério utilizado para seleção das amostras foi o seguinte:

- foram selecionados dois lotes de “amostra controle” utilizados pelo setor de Bioquímica Clínica do Laboratório de Análises Clínicas do Hospital Universitário (HU), da Universidade Federal de Santa Catarina (UFSC), nos meses de julho/98 e agosto/98.

Essas amostras passaram a representar, respectivamente:

grupo 1 - “amostra controle” do PRO-IN (Programa Interno de Qualidade), patrocinado pela Sociedade Brasileira de Análises Clínicas e

grupo 2 - “amostra controle” comercial da Controlab (Controle Laboratorial de Qualidade), patrocinado pela Sociedade Brasileira de Patologia Clínica.

O método ora proposto preconiza que o conjunto de dados seja formado por dois grupos aleatórios de variáveis e de tamanhos idênticos. Então, foram escolhidas as variáveis que compunham as duas planilhas de registro dos valores obtidos na dosagem das “amostras controles”, arquivadas no laboratório, que não apresentavam lacunas sem resultados, tendo em vista que a técnica exige que os grupos tenham o mesmo número de variáveis.

Com base nesse pré-requisito, foram escolhidos os mesmos números de variáveis em cada grupo, sendo selecionados para cada grupo cinco compostos, todos com valores

dentro da faixa aceitável do controle interno de qualidade: ácido úrico (AU), albumina (ALB), glicose (GLI), proteínas totais (PT) e uréia (URE).

Como cada “amostra controle” possui diferentes concentrações dos compostos dosados, estudou-se uma forma de padronizar os valores coletados. A técnica aplicada foi a de considerar o valor médio exposto na bula que vem acompanhando o lote como sendo 100% do valor esperado e o valor observado na leitura como X. Assim, os valores obtidos foram uniformizados em percentuais. Os valores padronizados utilizados estão descritos na Tabela 4.1.

#### 4.3. Função discriminante (modelo matemático)

A amostra disponível composta por 210 pontos permitiu a construção do modelo de forma satisfatória:

$$Z = 0,08x_1 - 0,57x_2 - 0,70x_3 - 0,28x_4 + 0,08x_5$$

onde:

$x_1$  = valores referentes a dosagem de ALB;

$x_2$  = valores referentes a dosagem de GLI;

$x_3$  = valores referentes a dosagem de PT;

$x_4$  = valores referentes a dosagem de URE;

$x_5$  = valores referentes a dosagem de AU.

Uma vez determinado o conjunto de variáveis que proporciona discriminação satisfatória para casos com participações grupais conhecidas, pode-se derivar um conjunto de funções de classificações, a fim de se enquadrarem novas observações cujas participações são desconhecidas.

Tabela 4. 1. – Valores padronizados em porcentagem dos compostos dosados nas “amostras controles” utilizados para alimentar o *software* na construção da função discriminante, onde os dados de um a vinte e um pertencem ao conjunto de observações advindas da “amostra controle” que representa o grupo 1, do dado vinte e dois ao quarenta e dois compõe o grupo 2.

	GRUPO	ALB	GLI	PT	URE	AU
1	1	100,0	96,3	93,5	95,1	105,2
2	1	97,0	95,3	93,5	106,1	108,3
3	1	94,0	93,7	93,5	102,4	106,2
4	1	97,0	96,3	95,6	96,3	105,2
5	1	94,0	95,8	95,6	100,0	106,2
6	1	100,0	95,8	93,5	103,6	106,2
7	1	106,0	97,4	100,0	106,1	109,4
8	1	100,0	96,3	97,8	98,8	109,4
9	1	100,0	95,8	97,8	98,8	109,4
10	1	100,0	94,2	97,8	111,0	107,3
11	1	100,0	95,8	93,5	107,3	108,3
12	1	103,0	97,4	97,8	103,6	107,3
13	1	100,0	95,8	97,8	100,0	108,3
14	1	100,0	96,3	97,8	95,1	106,2
15	1	100,0	101,0	95,6	103,6	103,1
16	1	100,0	101,6	97,8	100,0	103,1
17	1	100,0	103,1	95,6	102,4	101,0
18	1	100,0	101,0	95,6	102,4	101,0
19	1	103,0	95,3	95,6	97,5	104,2
20	1	100,0	93,7	97,8	93,9	100,0
21	1	100,0	96,9	95,6	104,9	99,0
22	2	99,6	106,2	98,5	107,3	93,6
23	2	103,0	104,8	101,5	109,1	96,8
24	2	103,6	107,1	104,5	109,1	95,2
25	2	99,6	104,3	103,0	107,3	96,8
26	2	103,3	105,2	103,0	109,1	96,8
27	2	100,7	106,2	103,0	109,1	96,8
28	2	98,3	103,4	104,5	103,6	93,6
29	2	104,6	102,9	103,0	105,4	95,4
30	2	103,6	102,9	103,0	105,4	100,0
31	2	102,3	105,3	101,5	109,1	93,6
32	2	101,6	102,9	98,5	105,4	93,6
33	2	101,3	103,4	100,0	109,1	93,6
34	2	104,0	102,5	101,5	109,1	95,2
35	2	100,0	102,9	103,0	96,4	107,9
36	2	102,0	106,2	101,5	98,2	104,8
37	2	103,3	102,5	101,5	107,3	107,9
38	2	104,3	102,9	100,0	109,1	104,8
39	2	102,3	103,9	101,5	109,1	107,9
40	2	104,0	102,9	101,5	109,1	104,8
41	2	104,3	105,7	101,5	98,2	107,9
42	2	101,3	97,9	100,0	103,6	104,8

O princípio básico do procedimento é a comparação do perfil de um grupo com aquele de um grupamento alternativo. Isso feito, a observação é classificada no grupo com que se assemelha, ou, em outras palavras, naquele que apresenta o maior Z, ou no grupo de maior probabilidade de participação, conforme explicado no Capítulo III, seção 3.2. e 3.3.

O modelo matemático e outras funções classificatórias foram geradas com o auxílio do pacote estatístico STATISTICA .

A geração dos dados está sendo explicada no anexo I.

#### 4. 4. Validade da classificação

Para se saber se foi obtida uma discriminação eficaz, usam-se as três maneiras anteriormente apresentadas no capítulo III: o teste F, o teste *Lambda de Wilks* e a matriz e tabela de classificação.

- Matriz e tabela de classificação

Um uso da técnica de classificação é testar a adequação das funções discriminantes (modelo matemático proposto). Os coeficientes de classificação são fornecidos, opcionalmente, por programas contidos no pacote computacional.

O propósito, aqui, consiste na avaliação da eficiência das variáveis discriminantes. Na hipótese de ocorrer uma grande proporção de classificações incorretas, pode-se concluir que as variáveis selecionadas são discriminantes fracas. Caso a proporção seja pequena, as variáveis são ditas fortemente discriminantes.

Na tabela 4.2, apresenta-se a matriz de classificação, a qual dá a proporção e a percentagem dos grupos correta e incorretamente classificados.

Tabela 4.2. Matriz de classificação das observações das “amostras controle”, através do método escolhido, com as respectivas percentagens de acerto da discriminação, onde nas linhas está representado o número de observações que compõem os grupos *a priori* e nas colunas está o número das observações classificadas depois da aplicação do modelo matemático.

GRUPO	G_1	G_2	CLASSIFICADOS CORRETOS (%)
G_1	21	0	100,00%
G_2	1	20	95,24%
TOTAL	21	20	97,62%

As observações foram classificadas respeitando a regra geral do método, como demonstrado no capítulo III, parte 3.3: as observações do grupo 1 são aquelas que forneceram um valor  $y_0 - m \geq 0$  e as do grupo 2, aquelas que forneceram  $y_0 - m < 0$ .

Diante desses dados, verifica-se que:

- a. em todas as vinte e uma dosagens dos compostos presentes na “amostra controle” do grupo 1 e classificados corretamente, o modelo se mostrou bem eficiente, e a probabilidade de ocorrer erro do tipo I (ver 3.3) é nula;
- b. nos dados do grupo 2, vê-se que o modelo classificou corretamente vinte dosagens e teve uma única classificação errônea, como não sendo pertencente àquele grupo, pois não apresenta o mesmo perfil, resultando em uma probabilidade de 4,76% de erro do método em classificar as observações.

Esta percentagem poderá cair ainda mais, à medida que no programa sejam introduzidos novos dados e todo o processo esteja controlado.

Na tabela 4.3, estão relacionadas as dosagens corretamente classificadas e a incorretamente classificada, como resultado obtido diretamente da aplicação do *software* STATISTICA a duas “amostras controle”. Como se pode ver, a coluna da esquerda

representa os 21 dados retirados do grupo 1: quando esses são classificados como pertencendo ao grupo, são expressos como  $G_1$ . O mesmo acontece na coluna da direita, na qual foram representados os 21 dados do grupo 2, e são expressos como  $G_2$ . No caso extremo da linha 21, na coluna da direita, onde há um asterisco, percebe-se que ele pertence *a priori* ao grupo 2, mas foi classificado como tendo características do grupo 1.

Na tabela de classificação, pode-se notar que o índice gerado pela análise discriminante apresenta alto poder diferenciador sobre as variáveis. Nota-se que a maioria dos pontos possuem características com seu próprio grupo de origem e que somente uma leitura ficou mal classificada.

- Teste F e teste *lambda de Wilks*

O teste F com 5 e 36 graus de liberdade confirmou o sucesso da discriminação, apresentando resultado de 33,65149. O resultado do teste F com valor elevado dá indícios de haver diferença significativa entre os grupos, possibilitando rejeitar a hipótese de que as observações derivam de uma mesma população.

O teste *lambda de Wilks* apresentou o valor de 0,1762481. O valor do *Lambda de Wilks* está mais próximo de zero, portanto pode-se inferir que os grupos estão perfeitamente discriminados.

Observando os dados obtidos através desses testes de validação da classificação, pode-se concluir que existe uma significância estatística. Isso quer dizer que os resultados, aos níveis estatístico de confiança, não são obtidos apenas aleatoriamente, ou seja, as características das observações, na verdade, as distinguem.

Esse contexto esclarece estatisticamente que os grupos de variáveis escolhidas possuem perfil diferente entre elas, e que a combinação linear foi eficaz em diferenciá-las e será capaz de discriminar novas observações.

Tabela 4.3.- Tabela de classificação dos casos, onde se mostra se as variáveis dos dois grupos foram identificadas cada qual no seu grupo, revelando perfeita discriminação dos soros controles. Os dados mal classificados, que pertencem a um grupo mas estão classificados em outro grupo, estão representados com um \* [Origem: STATISTICA, 1999].

1	G_1	1	G_2
2	G_1	2	G_2
3	G_1	3	G_2
4	G_1	4	G_2
5	G_1	5	G_2
6	G_1	6	G_2
7	G_1	7	G_2
8	G_1	8	G_2
9	G_1	9	G_2
10	G_1	10	G_2
11	G_1	11	G_2
12	G_1	12	G_2
13	G_1	13	G_2
14	G_1	14	G_2
15	G_1	15	G_2
16	G_1	16	G_2
17	G_1	17	G_2
18	G_1	18	G_2
19	G_1	19	G_2
20	G_1	20	G_2
21	G_1	*21	G_1

#### 4. 5. Discussão

Apresenta-se, integralmente, neste capítulo, uma nova técnica para o controle “global” de qualidade interno em laboratório de análises clínicas, com a utilização de dados multivariados.

Operacionalmente, o programa computacional precisa ser informado sobre as variáveis que podem sofrer alterações durante o processo. Então, à medida que fosse melhorando o controle de qualidade “local”, ou seja, o controle sobre cada uma das variáveis do processo, mais adequado seria a obtenção do controle “global”, no processo como um todo. Objetiva-se com esta técnica obter uma nova fórmula de CEP que trabalhe com variáveis multivariadas, buscando as relações que cada composto tem entre si.

Esta técnica é nova nos Laboratórios de Análises Clínicas e é útil para se construir um novo controle de qualidade interno, onde não haja preocupação apenas com o fato de o composto dosado estar ou não dentro das especificações gerais, mas também se procura verificar se esses valores não interferem na qualidade dos outros componentes. Isso pode acontecer, pois todos os compostos advêm de uma única solução (matriz).

## CAPÍTULO V

### CONCLUSÕES DO TRABALHO E SUGESTÕES PARA TRABALHOS FUTUROS

A metodologia para Reconhecimento de Padrões, aqui desenvolvida, consta de duas grandes etapas, dentro das quais estão embutidas etapas menores :

- 1) utilização de um método para fazer o Reconhecimento de Padrões (análises discriminante) dos dados coletados "em campo";
- 2) e, conseqüentemente, construção de um modelo matemático.

Com relação ao método, ele apresenta boa *performance* para reclassificar as observações em cada grupo de origem, mas nada pode ser generalizado, pois o comportamento dos métodos de Reconhecimento de Padrões tende a sofrer alterações de acordo com :

- a1. disposição dos pontos no espaço, que será diferente para cada caso abordado;

a2. a "alimentação" do programa de treinamento. Quanto mais representativas forem as amostras dos conjuntos A e B em questão, melhor tenderá a ser o desempenho do método.

Um modelo de reconhecimento de padrões é de suma importância pois, como visto, um modelo matemático multivariado interligado a um modelo de reconhecimento de padrões fornece técnica para o controle "global" de qualidade em indústria de papel [Farnaum & Stanton, 1989].

Com os valores quantitativos das características que fazem parte da composição das "amostras controles", pode-se "treinar" o modelo que gera um índice de separação dos grupos. Dessa forma, obtém-se uma superfície separadora que, para o caso, apresentou 4,7% de erros, aproximadamente, o que pode ser considerado baixo e, portanto, satisfatório.

O presente estudo representa uma primeira colaboração em fornecer uma possível "ferramenta" para o controle de qualidade interno nos laboratórios de análises clínicas, que trabalhem com variáveis multivariadas. O método tem sua principal vantagem no potencial em analisar o arquivo inteiro, em vez de realizar o exame seqüencial de suas características individuais, removendo assim o elemento subjetivo e as possíveis ambigüidades inerentes.

Outrossim, há redução do espaço de informações, de forma que dados selecionados (com alto poder discriminante) podem resultar num incremento da eficiência alocacional de recursos e tempo destinados às tarefas rotineiras de controle interno, servindo como complemento aos processos de avaliação já existentes.

A metodologia utilizada pode parecer um tanto complexa, mas foi toda desenvolvida através do programa computacional STATISTICA, bastante difundido no meio acadêmico, sendo este um aspecto positivo importante que deve ser salientado, pois sua implementação se torna simples, bastando interpretar os resultados fornecidos.

Além disso, com a combinação linear ótima estabelecida, pode-se aplicá-la facilmente dentro do laboratório, pois o analista somente terá que comparar as novas observações utilizando o método de classificação:

$$y_0 - m \geq 0 \quad \text{e} \quad y_0 - m < 0$$

ou seja, se  $x_0$  (observação nova) pertence ao grupo 1,  $y_0$  é maior do que  $m$  e se  $x_0$  pertence ao grupo 2,  $y_0$  é menor do  $m$ .

Então, como o método exige que os grupos já estejam estabelecidos *a priori* e sejam distintos, verifica-se se a nova observação obedece às características do seu grupo de origem. Caso ela seja classificada, através da metodologia, em outro grupo diferente, conclui-se que essa observação sofreu alguma alteração no seu processo e terá que ser revista.

## 5.2 Sugestões para trabalhos futuros

Os resultados alcançados até o momento sugerem várias alternativas para futuras investigações, considerando-se que o modelo proposto é bem sucedido na tarefa de discriminar entre os dois grupos de “amostras controles”.

A seguir, estão enumeradas algumas sugestões que poderão ser objeto de estudos futuros, de modo a aperfeiçoar a metodologia aqui abordada :

- a) obter maior número de padrões para os casos reais abordados neste trabalho. Desse modo, os programas computacionais, referentes aos métodos de Reconhecimento de Padrões pesquisados, estarão mais bem "alimentados", reduzindo assim as percentagens de erros produzidos;
- b) tomar novas variáveis para o estudo de seu poder discriminante, além das levantadas neste trabalho. Por exemplo, outros compostos como colesterol, triglicerídeos, etc.;
- c) desenvolver estratégia para aplicação efetiva da metodologia apresentada nos laboratórios, explorando diversos aspectos, tais como: verificar, na prática, para o modelo construído para o controle interno de qualidade, quais são as variáveis passíveis de alterações durante o processo, como, por exemplo, eventual troca da marca de reagentes,

pois os mesmos podem sofrer mudanças nas suas características definidas pelo modelo matemático .

## REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDERSON, S. C. & COCKAYNE, S. **Clinical Chemistry**. Philadelphia: WB Saunders Company Ed., 1993, 498 p.
2. BATISTA, J. L. M. **Análise Multivariada no Planejamento de Extensão Florestal : Subsídios para uma Política de Reocupação de Áreas**. Curitiba: Universidade Federal do Paraná. (Dissertação de mestrado), 1990.
3. BEEBE, K. R.; KOWALSKI, B. R. **An introduction to multivariate calibration and analysis**. Analytical Chemistry, Sept./1987, v. 59, n.º 17, p. 1007A – 1017A.
4. BENNETT, K. P. & MANGASARIAN, O. L. **Robust Linear Programming Discrimination of Two Linearly Inseparable Sets, Optimization Methods and Software**, 1992, v.1, p. 23-34.
5. BERTHOYEX, P. M.; BOX, G. E. P. & DARJATMOKO. A. **Discriminant Upset Analysis**. Madison: Center for Quality and Productivity Improvement University of Wisconsin – Report, 1988, n.º 30, p. 1-17.

6. BERWICK, D. M.; GODFREY, A. D. & ROESSNER, J.. **Melhorando a Qualidade dos Serviços Médicos, Hospitalares e de Saúde**. São Paulo: Makron Books do Brasil Ed., 1990, 354 p.
7. BOUROCHE, J. M. & SAPORTA, G. **Análise de dados**. Rio de Janeiro: Zahar Ed., 1982.
8. BROWN, C. L. W. & BURNETT, R. W. **Practical Aspects of Implementation of the Third Generation Quality Control Procedures in Clinical Chemistry**. New Orleans: 46 American Association of Clinical Chemistry National Meeting, 1994.
9. CEMBROWSKI, G. S. & SULLIVAN, A. M. **Quality Control and Statistics**. In: BISHOP, M. L.; DUBEN-ENGELKIRK, J. L. & FODY, E. P.. *Clinical Chemistry*. Philadelphia: Lippincott-Raven Publishers Ed., 3<sup>a</sup> ed., 1996., c. 4, p. 76 – 96.
10. CROISER, R. B. **Multivariate Generalizations of Cumulative Sum Quality-Control Schemes**. *Technometrics*: August/1998, v. 30, n° 3, p. 291 – 303.
11. FARNAUM, N. R. & STANTON, L. W. **Quantitative forecasting methods**. Boston, PWS – Kent Publishing Company Ed., 1989.
12. FERGOSON, R. B. **Chemical Process Optimization utilizing Neural Network Systems, Pavilion Technologies, Inc.**, Texas: Austin Ed., 1992.
13. FRIEDRICH, M. , SERPENTINO, A. H. & SANNAZARRO, C. A. C. **Aplicação do procedimento de regras múltiplas de Shewhart no controle de qualidade em bioquímica clínica**. *Revista LAES & HAES*: 1998, n° 15, p. 124 – 132.
14. FU, K. S. **Syntatic Pattern Recognition and Applications**. New Jersey: Prentice-Hall Ed., 1982.

15. HAINLINE Jr, A. **Quality Assurance: Theoretical and Practical Aspects**. New York: 45<sup>o</sup> American Association of Clinical Chemistry National Meeting, 1993.
16. HENRY, J. B. **Diagnósticos Clínicos e Conduta Terapêutica por Exames Laboratoriais**. São Paulo: Manola Ed., 19 ed., 1995.
17. GROTSCHER, M. & WAKABAYASHI, Y. A **Cutting Plane Algorithm for a Clustering Problem**, Mathematical Programming: 1989, 45, p. 59 - 96.
18. JACSON, J. E. & MORRIS, R. H. (). **Application of Multivariate Quality Control to Photographic Processing**. Journal of the American Statistical Association: 1957. v. 52, nº 2, p. 186 – 199.
19. JOHNSON, R. A. & WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Prentice-Hall Ed., 3 ed., 1998.
20. JURAN, J. M. & GRYNA, F. M. **Controle de Qualidade**. São Paulo: Makron Books e McGraw-Hill Ed., 1993, v. 6.
21. LOWRY, C. A.; WOODWALL, W. H.; CHAMP, C. W. & RIGDON, S. E. A **Multivariate Exponentially Weighted Moving Average Control Chart**. Technometrics: February/1992, v. 34, nº 01, p. 46 – 53.
22. MANGASARIAN, O. L., SETIONO, R. & WOLBERG, W. H. **Pattern Recognition via Linear Programming : Theory and Application to Medical Diagnosis**. In: Coleman, T. F. & Li, Y. Large-Scale Numerical Optimization, Philadelphia: SIAM Ed., 1990, p.22 - 30.
23. MARDIA, K. V., KENT, J. P. & BIBBY, J. M. **Multivariate Analysis**. London: Academic Press, 1979, p. 175 - 178.

24. MARTINS, M. V. & SAMANÉZ, C. P. **Análise discriminante e os bancos comerciais privados brasileiros: uma aplicação.** São Paulo: Revista Brasileira de Mercado Capital.– dez./1987, v. 13, n° 39, p. 123 – 142.
25. NASCIMENTO, C. A. O. & YAMAMOTO, C. I. **Modelagem de Processos Químicos via Redes Neurais.** São Paulo: I Workshop em Redes Neurais, IPT, 1993.
26. PATUWO, E., HU, M. Y. & HUNG, M. S. **Two-Group Classification using Neural Networks, Decision Sciences.** 1993, v. 24, n° 4.
27. PEREIRA, R. F. & NESS JR., W. L. **Características das empresas que utilizaram o mercado primário de ações.** São Paulo: Revista Brasileira de Mercado Capital.– Mai/Ago/1980, v. 6, n° 17, p. 155 – 169.
28. RÖEHRIG C., CARNIERI, C. e STEINER, M. T. A. **Diagnóstico Diferencial da Causa da Icterícia Obstrutiva por P. L. Discriminante não Convexa.** Curitiba: XI Seminário Brasileiro de Endoscopia Digestiva, out./1993.
29. SAMAL, A. & IYENGAR, P. A. **Automatic Recognition and Analysis of Human Faces and Facial Expressions : a Survey,** Pattern Recognition, 1992, v. 25, p. 65 - 77.
30. SILVA, J. G. C. & MACHADO, A. A. **Análise multivariada.** Pelotas: Apostila. UFPel – RS, 1996.
31. STEINER, M. T. A., CARNIERI, C. **Programação Linear, Métodos Estatísticos e Redes Neurais em Reconhecimento de Padrão.** Curitiba: II Encontro Regional da Sociedade Brasileira de Matemática Aplicada e Computacional, 1995.

32. TAM, K. Y. & KIANG, M. Y. **Managerial Applications of Neural Networks: The Case of Bank Failure Predictions**, Management Sciences 1992, v. 38, n° 7, p. 926 - 947.
33. TELHADA, M. A. L. **Cartas de Controle Multivariadas**. São Paulo: Tese de Doutorado. Escola Politécnica da Universidade de São Paulo. Engenharia de Produção, 1995.
34. TIBÚRCIO, H. M. **Controle Interno de Qualidade Analítica**. São Paulo: Apostila, 1995.
35. TONKS, D. B. **Proposed Quality Specifications for the Imprecision and Inaccuracy of Analytical Systems for Clinical Chemistry**. European Journal Clinical Chemistry, 1995, n° 30, p. 311 - 317.
36. WESTGARD, J. O. **Analytical Quality Assurance Through Quality Planning and Quality Control**. Chigado: 44° American Association of Clinical Chemistry National Meeting, 1992.
37. WOLBERG, W. H. & MANGASARIAN, O. L. **Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology**. Applied Mathematics, 1990, v. 87, p. 9193 - 9196.
38. WHO - World Health Organization. **Guidelines for Program for Internal Quality Control of Quantitative Analyses in Clinical Chemistry, Norma Lab/81.3**, Geneve: 1992.

## **ANEXO**

## ANEXO I

### APLICANDO O EXEMPLO NO *SOFTWARE STATISTICA*

O método de análise discriminante se baseia nas diferentes dosagens dos compostos encontrados em dois tipos de “amostra controle”, artifício utilizada na inspeção do controle interno de qualidade nos laboratório de análises clínicas.

A proposta desta análise é saber, estatisticamente, como a metodologia pode discriminar uma variável qualitativa (tipo de amostra) baseando-se em cinco variáveis quantitativas (medidas dos compostos):

ácido úrico (AU), albumina (ALB), glicose (GLI), proteínas totais (PT) e uréia (URE).

Como cada “amostra controle” possui diferentes concentrações dos compostos dosados, estudou-se uma forma de padronizar os valores coletados. A técnica aplicada foi a de considerar o valor médio exposto na bula que vem acompanhando o lote como sendo 100% do valor esperado e o valor obtido na leitura como X.

### 1. Tabela de dados

Ao abrir o programa ele mostrará a caixa de dialogo, onde se escolhe o método de análise estatística desejado. Após aparecerá no visor a tabela para preenchimento dos dados a serem analisados (Fig. 6.1).

Nº	VAL	2	3	4	5	6
		ALB	GLI	PT	URE	AU
1	1	100,0	96,3	93,5	95,1	105,2
2	1	97,0	95,3	93,5	106,1	108,3
3	1	94,0	93,7	93,5	102,4	106,2
4	1	97,0	96,3	95,6	96,3	105,2
5	1	94,0	95,8	95,6	100,0	106,2
6	1	100,0	95,8	93,5	103,6	106,2
7	1	106,0	97,4	100,0	106,1	109,4
8	1	100,0	96,3	97,8	98,8	109,4
9	1	100,0	95,8	97,8	98,8	109,4
10	1	100,0	94,2	97,8	111,0	107,3
11	1	100,0	95,8	93,5	107,3	108,3
12	1	103,0	97,4	97,8	103,6	107,3
13	1	100,0	95,8	97,8	100,0	108,3
14	1	100,0	96,3	97,8	95,1	106,2
15	1	100,0	101,0	95,6	103,6	103,1
16	1	100,0	101,6	97,8	100,0	103,1

Fig. 6. 1 – Tabela de dados

Os dados gerados para testar a discriminação entre os grupos serviram para “alimentar” a tabela de dados contida no *software*; sendo que, para as variáveis quantitativas foram utilizadas as próprias observações retiradas da planilha de registro e a variável qualitativa, que identifica qual tipo de grupo pertence os dados, é representado com o número 1 ou 2, como estabelecido *a priori*:

grupo 1 - “amostra controle” do PRO-IN (Programa Interno de Qualidade), patrocinado pela Sociedade Brasileira de Análises Clínicas e

grupo 2 - “amostra controle” comercial da Controlab (Controle Laboratorial de Qualidade), patrocinado pela Sociedade Brasileira de Patologia Clínica.

Depois de começar o módulo análises discriminante e preencher a tabela de dados. Agora, será necessário especificar a coluna está a variável de grupo e as variáveis quantitativas, através de uma nova caixa de diálogo (Fig. 6. 2). Clica-se OK para começar a análise.

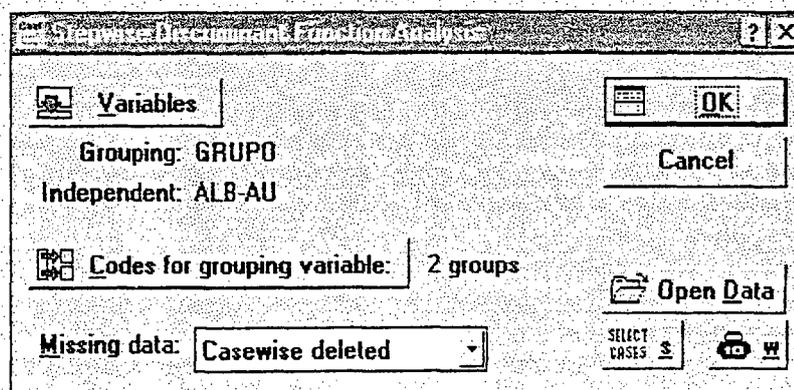


Fig. 6. 2 – Caixa de diálogo onde se especifica a coluna das variáveis de grupo e as variáveis qualitativas.

O programa abre uma nova célula (Fig. 6. 3) onde pode-se definir certas especificações, como por exemplo, o valor para tolerância de acertos, método. Geralmente, os valores determinado pelo *software* para essas especificações são estatisticamente significativos para a análise discriminantes.

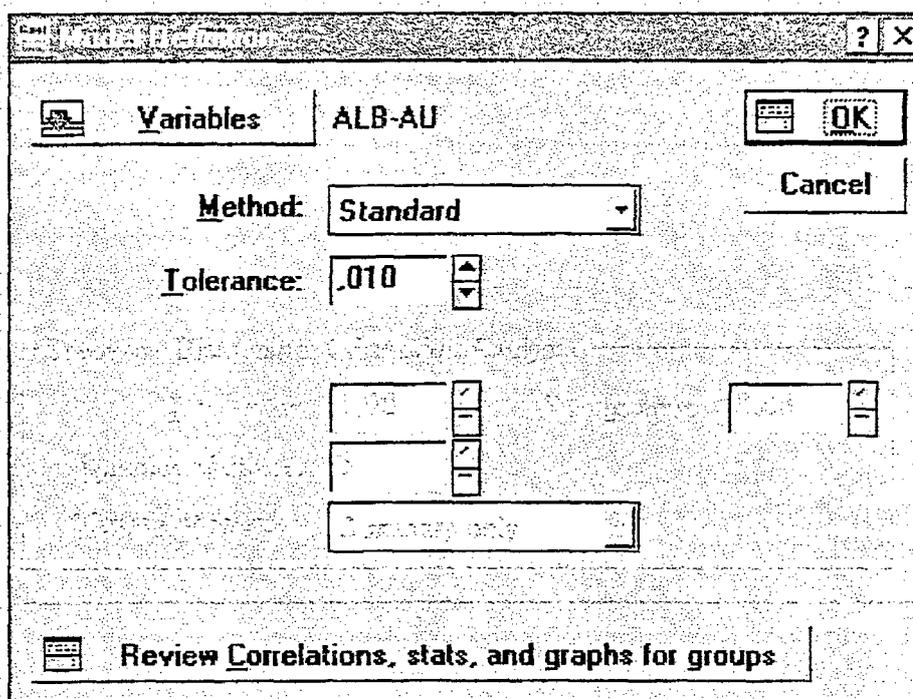


Fig. 6. 3 – Caixa para definir os parâmetros da análise discriminante.

#### Caixa de Resultados

O pacote estatístico realizará toda as operações da análise discriminante conforme o modelo. Então aparecerá uma nova caixa onde pode-se ver o número de variáveis , o valor do teste F calculado e do *Lambda de Wilks*, que são testes para verificar, estatisticamente, se a discriminação entre os dois grupos é significativa (Fig. 6. 4).

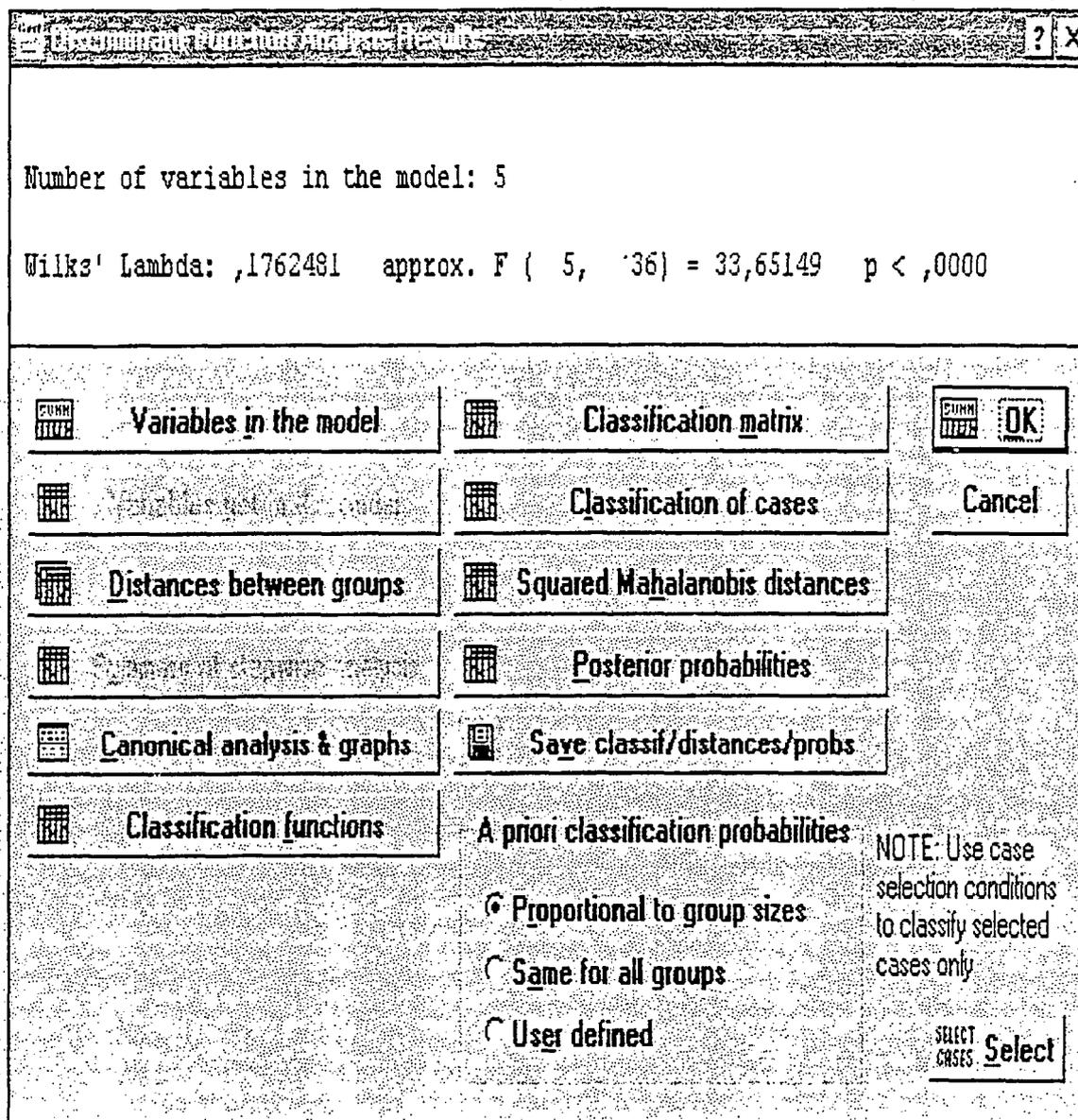


Fig. 6.4 – Caixa resultado da análise.

A partir dessa caixa pode-se obter, ainda, as funções de classificação, como a matriz e Tabela de classificação dos casos e os coeficientes padronizado que foram utilizados para a construção do modelo matemático (Fig. 6.5).

DISCRIM. ANALYSIS for Canonical Variables	
Variable	Coef.
GLI	.03688
PT	-.24604
URE	-.39041
AU	-.06618
Constant	.01697
Eigenval	64.72977
Cum Prop	4.67382
	1.00000

Fig. 6.5 - Tabela de classificação dos casos e os coeficientes padronizado.

#### Conclusão sobre o uso do *software*

O pacote estatístico *Statistica* é um programa computacional bastante difundido no meio acadêmico que fornece ao seu usuário a facilidade de desenvolver de forma rápida e satisfatória o processamento de dados através de metodologias estatística complexas, como as análises envolvendo múltiplas variáveis.