

**DETECÇÃO DE PROBLEMAS EM
INSTITUIÇÕES FINANCEIRAS
UTILIZANDO
MODELOS ESTATÍSTICOS**

Roberta Blass Staub Oliveira

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU
DE
MESTRE EM ESTATÍSTICA

Área de concentração: **Estatística**

Orientadora: **Profa. Dra. Lúcia Pereira Barroso**

- São Paulo, janeiro de 2000 -

À memória de meu pai, Augustinus Staub
e de minha avó, Cesarina Blass.

À minha mãe, Lorena e irmão, Rogério e ao meu
marido, Elmo Oliveira, pela paciência que tiveram comigo.

DETECÇÃO DE PROBLEMAS EM INSTITUIÇÕES FINANCEIRAS
UTILIZANDO MODELOS ESTATÍSTICOS

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
defendida por Roberta Blass Staub Oliveira
e aprovada pela comissão julgadora.

São Paulo, 10 de fevereiro de 2000.

Banco Examinadora:

- Profa. Dra. Lúcia Pereira Barroso (Orientadora) - IME - USP
- Profa. Dra. Clélia Maria de Castro Toloí - IME - USP
- Prof. Dr. Heron C. Esvael do Carmo - FEA - USP

Agradecimentos

Agradeço a todos que tiveram a oportunidade de conviver comigo durante a execução deste trabalho tendo paciência nos momentos de maior dedicação.

Mais especificamente, gostaria de agradecer a confiança a mim dedicada por:

- Lúcia Pereira Barroso: orientadora da dissertação e responsável pelo acompanhamento direto e pela dedicação em ler e reler minuciosamente este trabalho, enriquecendo-o constantemente;
- Clélia Maria C. Toloí, professora da USP: pela ajuda na definição do “rumo a ser seguido”, fundamental para o início efetivo do trabalho;
- Henry Corazza Sef: amigo que conheci no curso de mestrado, pelo apoio dado durante a estadia na cidade de São Paulo;
- Joviniano Melo: ex-chefe do Departamento de Cadastro e Informações do Sistema Financeiro, onde trabalho desde meu ingresso no Banco Central. Responsável pela minha liberação oficial do Banco Central para cursar o mestrado e também pela compreensão nos momentos em que precisei estudar;
- Sérgio de Almeida Lima: atual chefe do Departamento de Cadastro e Informações do Sistema Financeiro por continuar me apoiando para a finalização desta dissertação;

- Tânia Maria Lucas Dutra: pelo incentivo dado tal que eu me inscrevesse no curso de mestrado da Universidade de São Paulo - USP - e pelo incentivo profissional que com certeza foram decisivos;
- Lúcio Capelletto: orientador da dissertação no que se refere a aspectos contábeis, também agradeço pelas contribuições feitas tal que as interpretações contábeis fossem melhoradas;
- Geraldo Moreira de Melo: por esclarecer exatamente as características do banco de dados tal que as informações contábeis fossem extraídas corretamente pelo amigo Maurício;
- Maurício Soares: por ter feito um programa especialmente para a coleta dos dados utilizados na aplicação dos modelos estatísticos;
- Silvia Marques de Brito e José Reynaldo de Almeida Furlani: pela disponibilidade em esclarecer dúvidas referentes a aspectos contábeis necessários ao desenvolvimento deste trabalho.

Resumo

O Banco Central do Brasil exerce a fiscalização das instituições financeiras e tem desenvolvido metodologias de avaliação estatística com o objetivo de identificar algum tipo de deterioração na solidez ou no comportamento das mesmas. Este trabalho é desenvolvido com o objetivo de detectar tais comportamentos e apresenta uma regra de alocação das instituições financeiras em dois grupos (sem e com problemas) baseada em indicadores econômico-financeiros. São apresentados e discutidos seis modelos obtidos de Análise Discriminante, Regressão Logística e Regressão Probit.

ÍNDICE

1	Introdução	4
2	Caracterização do Problema	7
2.1	Indicadores Econômico-financeiros	7
2.2	Definições	9
2.3	Base de Dados e Período de Análise	12
2.4	Medida Resumo	13
2.5	Intervenção e Liquidação Extrajudicial	17
3	Análise Descritiva	
	e Testes de Hipóteses	19
3.1	Análise das Observações	20
3.2	Análise das Observações “Rankeadas”	24
4	Metodologia	29
4.1	Técnicas Utilizando as Variáveis Originais	29
4.1.1	Análise Discriminante	29
4.1.2	Regressão Logística	34

4.1.3	Regressão Probit	37
4.1.4	Medidas de Diagnóstico de Regressão	38
4.1.5	Seleção Automática de Variáveis (regressão)	43
4.2	Técnicas Utilizando o “Rank” das Observações	43
4.2.1	Análise Discriminante do “Rank”	43
4.2.2	Regressão Logística do “Rank”	46
4.2.3	Regressão Probit do “Rank”	48
5	Aplicação	50
5.1	Análise das Variáveis Originais	50
5.1.1	Análise Discriminante Linear	51
5.1.2	Regressão Logística	53
5.1.3	Regressão Probit	56
5.2	Análise das Variáveis “Rankeadas”	58
5.2.1	Análise Discriminante Linear	58
5.2.2	Regressão Logística	61
5.2.3	Regressão Probit	64
5.3	Comparações	66
5.4	Conclusão	72
6	Discussões Teóricas	73
6.1	Eisenbeis, 1977	74
6.2	Espahbodi, 1991	75
6.3	Pregibon, 1981	76
6.4	Perry e Cronan, 1986	76
6.5	Conover e Iman, 1980	77

6.6	Conover e Iman, 1981	78
6.7	Press e Wilson, 1978	79
6.8	So, 1998	81
	Bibliografia	82
A	“Boxplots”	87
B	Base de Dados	99
C	Probabilidades Estimadas para os Modelos 1 a 6	119

Capítulo 1

Introdução

De acordo com a Lei nº 4595 de 1964:

“Art. 10. Compete privativamente ao Banco Central do Brasil: ...

VIII - exercer a fiscalização das instituições financeiras e aplicar as penalidades previstas;...”

Com este objetivo o Banco Central (BACEN) tem desenvolvido metodologias de avaliação estatística de instituições financeiras (IF's) para identificar, previamente, algum tipo de deterioração na solidez ou no comportamento da instituição analisada.

No Brasil, entre 1994 e 1996, 29 bancos, representando 15,4% do total de depósitos, foram sujeitos à intervenção oficial, colocados sob administração especial ou receberam assistência financeira (Lindgren, Garcia e Matthew, 1996). Ocasionalmente, ocasionando significativo montante de recursos financeiros bloqueados, essas intervenções prejudicaram milhares de clientes, entre eles pessoas físicas e jurídicas, tornando-os incapazes de saldar seus compromissos. Entre os casos (famosos) temos, em 1995, as intervenções nos Bancos Econômico e Nacional.

Além disso, com a globalização, as IF's estão vulneráveis às flutuações do sistema financeiro internacional que podem gerar intensa migração de recursos devido a crises mundiais. As consequências se intensificaram no país após o Plano Real com o aumento na captação de recursos estrangeiros devido à alta taxa de juros (uma das bases de

sustentação do plano), comparativamente maior que a de outros países. Recentemente temos como exemplo, a crise asiática que obrigou a elevação da taxa de juros a fim de que o investidor estrangeiro não retirasse seu capital do país.

Portanto, tentar identificar atipicidades no comportamento de IF's que possam caracterizar problemas na sua gestão ou ocasionados por fatores externos é, mais do que nunca, fundamental, tendo em vista que não existe política econômica eficaz sem um sistema financeiro saudável.

A situação de uma IF pode ser avaliada indiretamente com base em indicadores econômico-financeiros considerados importantes pelos analistas de balanços. Este trabalho estuda modelos estatísticos que classifiquem IF's como sujeitas a falência e a problemas financeiros, ou não. Os indicadores empregados no estudo baseiam-se em três categorias: liquidez, estrutura e rentabilidade. É importante salientar que a idéia não é obter uma regra que levará diretamente à intervenção, mas sim dar um sinal de alerta para que o Banco Central, baseado também em outros estudos, possa agir da melhor maneira.

Usualmente a bibliografia encontrada a respeito utiliza análises transversais, ou seja, onde os indicadores de cada IF são observados em determinado mês/ano. Os resultados são então estendidos para outros períodos. Contudo, ajustes contábeis podem distorcer os valores obtidos para determinado indicador e informações referentes não apenas a um mês específico podem ser relevantes.

Considerando dados coletados a partir de agosto de 1994 até agosto de 1998, utilizou-se, então uma medida resumo dessas informações supondo que pudesse caracterizar a série dos indicadores e ao mesmo tempo distinguir as IF's em boa situação econômico-financeira daquelas que possuem indicativos de deterioração nos seus índices de liquidez, estrutura e rentabilidade.

Este trabalho foi desenvolvido na seguinte seqüência. O Capítulo 2 descreve os indicadores econômico-financeiros de interesse e suas características, assim como o período

de análise e a medida resumo utilizada. O Capítulo 3 apresenta uma análise descritiva dos dados com o intuito de verificar se as suposições para aplicação da análise estatística estão satisfeitas. As técnicas estatísticas são apresentadas no Capítulo 4 e aplicadas aos dados no Capítulo 5, onde comparam-se os resultados obtidos pelas diferentes técnicas utilizadas considerando o número de instituições financeiras classificadas corretamente pelos modelos. Finalmente, no Capítulo 6 são apresentadas discussões sobre técnicas estatísticas encontradas na literatura.

Capítulo 2

Caracterização do Problema

2.1 Indicadores Econômico-financeiros

Os indicadores econômico-financeiros de liquidez, estrutura e rentabilidade utilizados são descritos a seguir e definidos pelos especialistas de mercado como importantes para avaliar a situação de uma IF. Os valores podem ser calculados mensalmente a partir de dados contábeis ao nível de informação disponível para o público, do Plano Contábil das Instituições do Sistema Financeiro Nacional - COSIF, sob gestão do Banco Central do Brasil - BACEN. Os dados estão apresentados no Apêndice B.

Os indicadores de liquidez são aqueles que evidenciam a situação financeira de uma empresa frente a seus diversos compromissos passivos (Assaf Neto, 1993), ou seja, medem as suas disponibilidades para saldar suas obrigações.

Os indicadores de estrutura são utilizados para a análise e controle da estrutura de capital. São utilizados para uma avaliação da proporção de capital próprio e capital de terceiros existente na composição do passivo em relação aos ativos da empresa (Assaf Neto, 1993).

Os de rentabilidade objetivam a interpretação e análise dos resultados auferidos pela empresa. São expressos pela relação entre receitas e lucros com vários itens patrimoniais, conforme seja o objetivo da análise (Assaf Neto, 1993).

São eles:

Indicadores de Liquidez

- Liquidez Imediata (L1)
- Liquidez Corrente até 90 dias (L2)

Indicadores de Estrutura

- Imobilização do Patrimônio Líquido (L3)
- Grau de Alavancagem dos Recursos Próprios (L4)
- Participação das Exigibilidades no Ativo Operacional (L5)
- Participação dos Depósitos Interfinanceiros no Total de Depósitos (L6)
- Evolução do Ativo Operacional Médio (L7)
- Participação dos Créditos Anormais no Total das Operações de Créditos Normais (L8)
- Representatividade dos Créditos Anormais em Relação ao Patrimônio Líquido (L9)

Indicadores de Rentabilidade

- Remuneração do Ativo Operacional Médio (L10)
- Margem Líquida (L11)
- Evolução do Patrimônio Líquido (L12)
- Rentabilidade do Patrimônio Líquido Médio (L13)
- Taxa de Retorno das Aplicações (L14)
- Custo das Captações (L15)

- Participação das Rendas de Prestação de Serviços em Relação às Despesas Administrativas (L16)
- Participação das Despesas Administrativas no Ativo Total (L17)

2.2 Definições

A definição de cada um dos 17 indicadores (Capelletto, 1995) é dada a seguir.

Liquidez Imediata (L1): somam-se os valores constantes nas rubricas de disponibilidades, aplicações interfinanceiras de liquidez, títulos e valores mobiliários (TVM) dividindo-os pelos depósitos interfinanceiros adicionados de obrigações por operações compromissadas. O quociente da divisão informa quanto a instituição possui de ativos realizáveis no curto prazo relativamente aos depósitos interfinanceiros e obrigações por operações compromissadas. Quanto maior for o valor do indicador, melhor está a instituição.

Liquidez Corrente até 90 dias (L2): divide-se o valor dos ativos realizáveis até 90 dias pelo valor do passivo exigível até 90 dias. O indicador de liquidez corrente permite que se saiba como está o equilíbrio financeiro dentro de um prazo de 90 dias. Quanto maior for o valor do indicador, melhor está a instituição.

Imobilização do Patrimônio Líquido (L3): divisão do ativo permanente pelo patrimônio líquido. Permite a apuração do quanto de recursos próprios estão comprometidos em aplicações sem intenção de realização (imobilizado). O quociente informa com que volume de capital de giro próprio opera a instituição. Quanto menor for esse indicador, maior é o capital de giro próprio e melhor está a instituição.

Grau de Alavancagem dos Recursos Próprios (L4): divisão dos capitais de terceiros pelo capital próprio. Evidencia o quanto a instituição está se utilizando de recursos de terceiros para cada unidade de recursos próprios. A exposição ao risco é identificada nesse indicador, sendo oportuno ressaltar que um alto valor no grau de

alavancagem, além de evidenciar elevado risco, pode evidenciar maior potencial da instituição em gerar lucros. A princípio, para os bancos que compõem a amostra, o valor ótimo para o indicador situa-se dentro de uma faixa que varia entre sete e quinze vezes o valor do patrimônio líquido. Fora dela, quando abaixo, tem-se que o banco está com baixo nível de atividades - subutilizado - e, quando acima, está operando com recursos de terceiros mais do que o desejado, expondo-se demais ao risco.

Participação das Exigibilidades no Ativo Operacional (L5): divisão do total das exigibilidades pelo ativo operacional. Demonstra a utilização de recursos de terceiros em aplicações em operações usuais, referentes à atividade-fim da instituição. O indicador não deve ultrapassar o valor 1.

Participação dos Depósitos Interfinanceiros no Total de Depósitos (L6): divisão de depósitos interfinanceiros pelo total de depósitos. Mostra a participação dos recursos captados no mercado interbancário em relação ao total de depósitos captados. Como os recursos do interbancário são mais caros, quanto menor for esse indicador, melhor está a instituição.

Evolução do Ativo Operacional Médio (L7): divisão da média dos últimos t meses do ativo operacional pela média dos últimos $t - 1$ meses. Permite verificar o comportamento das atividades operacionais da instituição. Como o crescimento de operações é desejável para a instituição, quanto maior for o indicador, melhor.

Participação dos Créditos Anormais no Total das Operações de Créditos Normais (L8): divisão dos créditos anormais (em atraso e em liquidação) pelos créditos normais. Demonstra a capacidade de análise de crédito da IF. Indica quanto dos créditos concedidos estão vencidos e não pagos, existindo o risco de não serem honrados, relativamente aos créditos normais. Quanto menor o indicador, melhor está a instituição.

Representatividade dos Créditos Anormais em Relação ao Patrimônio Líquido (L9): divisão dos créditos anormais (em atraso e em liquidação) pelo patrimônio

líquido, adicionado pelo lucro ou prejuízo do mês. Demonstra o comprometimento do patrimônio líquido com as operações de crédito anormais, informando qual o risco destas operações para a situação líquida da instituição financeira. Quanto menor for esse indicador (valor absoluto), melhor está a instituição.

Remuneração do Ativo Operacional Médio (L10): divisão das receitas operacionais mensais pelo ativo operacional médio. Mensura a capacidade dos ativos operacionais gerarem receitas. Mostra o retorno das aplicações realizadas no ativo operacional sob a forma de receitas operacionais. O aumento da remuneração do ativo operacional é favorável para a IF.

Margem Líquida (L11): divisão do resultado líquido mensal pelo total mensal das receitas. Informa qual é a margem de lucro mensal da instituição, isto é, o quanto das receitas totais transformam-se em resultados líquidos. Quanto maior for esse indicador, melhor está a instituição.

Evolução do Patrimônio Líquido (L12): divisão entre os saldos do patrimônio líquido do mês t e o saldo do mês $t - 1$. Permite avaliar o comportamento mensal dos recursos próprios. Como o crescimento do patrimônio líquido é saudável para a instituição, quanto maior for o indicador, melhor.

Rentabilidade do Patrimônio Líquido Médio (L13): divisão do resultado líquido mensal pelo patrimônio líquido médio. Mostra o quanto a instituição obteve de retorno, sob a forma de lucro, para cada unidade de recursos próprios investida. Quanto maior o indicador, melhor está a instituição.

Taxa de Retorno das Aplicações (L14): soma das rendas mensais de operações de crédito, rendas mensais de arrendamento mercantil, rendas mensais de câmbio, rendas mensais de aplicações interfinanceiras de liquidez, rendas mensais de títulos e valores mobiliários dividida pela soma das aplicações em disponibilidades, aplicações interfinanceiras de liquidez, títulos e valores mobiliários, repasses interfinanceiros, operações de

crédito, operações de arrendamento mercantil, operações de câmbio e negociação e intermediação de valores. Mensura a taxa de retorno mensal dessas aplicações. Quanto maior for o indicador, melhor está a instituição.

Custo das Captações (L15): soma das despesas mensais de captação, despesas mensais de obrigações por empréstimos e repasses, despesas mensais de arrendamento mercantil, despesas mensais de câmbio e resultado mensal de transações com títulos e valores mobiliários dividida pelo total das exigibilidades menos as relações interfinanceiras e interdependências, menos as outras obrigações, mais os repasses interfinanceiros e as obrigações originadas por operações de câmbio e na negociação e intermediação de valores. Mensura quanto a instituição está pagando para ter esses recursos. Quanto menor esse indicador, melhor está a instituição.

Participação das Rendas de Prestação de Serviços em Relação às Despesas Administrativas (L16): divisão das rendas mensais de prestação de serviços pelas despesas administrativas mensais. Mostra o grau de cobertura dos encargos administrativos pelas rendas mensais de prestação de serviços. Revela a capacidade da instituição recuperar, via a prestação de serviços, os custos administrativos. Quanto maior for esse indicador, melhor está a instituição.

Participação das Despesas Administrativas no Ativo Total (L17): divisão das despesas administrativas mensais pelo ativo total. Demonstra o quanto a IF depende em despesas administrativas para gerir/ter aquele ativo. Quanto menor o indicador, melhor.

2.3 Base de Dados e Período de Análise

Para o presente trabalho coletaram-se os indicadores econômico-financeiros dos bancos comerciais e múltiplos tal que, ordenados em ordem decrescente de ATT médio (ativo total ajustado segundo critério do BACEN), somem 95% do ATT médio total, ou seja, são as IF's de grande e médio porte. O período de análise vai desde a implantação do

Plano Real (agosto de 1994) até agosto de 1998, tendo em vista as modificações significativas ocorridas no comportamento das IF's. Para as IF's que sofreram intervenção ou liquidação, os dados referem-se somente até essa data.

Bancos Comerciais

São instituições financeiras bancárias ou monetárias, privadas ou públicas, constituídas sob a forma de sociedade anônima, especializadas em operações de curto e médio prazos (hoje autorizadas a funcionar pelo Banco Central), com objetivo de oferecer financiamento, oportuno e adequado, ao comércio, à indústria, às empresas prestadoras de serviços, pessoas físicas, bem como ao crédito rural.

Bancos Múltiplos

Instituição financeira única, com personalidade jurídica própria, nos termos da legislação em vigor que pode agrupar as seguintes carteiras: comercial, de investimento, de desenvolvimento, de crédito imobiliário e de crédito, financiamento, investimento e arrendamento mercantil. Tal organização pode verificar-se através de processos de fusão, incorporação, cisão, transformação ou constituição direta, dependendo de prévia autorização do Banco Central do Brasil (Resolução nº 1524, de 21 setembro de 1988).

2.4 Medida Resumo

Este trabalho apresenta algumas dificuldades na modelagem estatística dos dados. As séries dos indicadores de várias IF's é bastante curta tornando inadequada a aplicação de algumas técnicas de análise de dados em séries temporais. O tamanho máximo das séries consideradas é de 48 observações (agosto de 1994 até agosto de 1998), e o mínimo de 4 observações apenas, pois as séries são obtidas a partir da implantação do Plano Real, até a data de intervenção ou liquidação extrajudicial, se ocorrer. A aplicação da análise discriminante no domínio da frequência fica conseqüentemente inviabilizada (Koyama,

1997 e Schumway, 1988). Sendo assim, optou-se por escolher uma medida resumo que representasse a série dos indicadores no sentido de discriminar as IF's nos grupos sem e com problemas, 0 e 1 respectivamente. Mais detalhes sobre utilização de medidas resumo podem ser encontrados em Domenech (1989).

Para a obtenção da medida resumo utilizada procedeu-se aos seguintes passos:

1) do arquivo contendo todas as IF's a serem analisadas no trabalho, excluir das séries das IF's com problemas as observações referentes ao período em que a IF encontrava-se sob intervenção ou liquidação extrajudicial. Como na prática não interessa caracterizar a IF após as providências do BACEN, e sim antes, esses dados foram desconsiderados;

2) a partir de 1), conforme indicações dos especialistas da área, obter as 6 últimas observações para cada indicador de cada IF. Com base na estabilidade gerada pelo Plano Real supõe-se que as condições econômicas são as mesmas no período de análise. Em alguns casos não foi possível obter as 6 últimas observações, e sim as 4 ou 5 últimas;

3) a partir de 2) extrair o valor extremo, máximo ou mínimo, das 6 últimas observações de cada série, para cada indicador, escolhendo aquela que indica a pior situação da IF.

Por exemplo, para Liquidez Imediata (L1) quanto maior for o valor do indicador melhor está a instituição. Neste caso, dos 6 últimos meses considera-se o menor valor para cada IF. Para Imobilização do Patrimônio Líquido (L3) quanto menor o indicador, melhor está a instituição. Utiliza-se, portanto, o maior valor dos 6 últimos meses. Temos, então, a pior situação possível das IF's nos últimos 6 meses.

Com base na medida resumo descrita acima ajustaram-se os modelos estatísticos descritos no Capítulo 4. Na Tabela 2.1 apresentam-se os nomes das IF's analisadas e qual o último mês utilizado para o cálculo da medida resumo.

Tabela 2.1 IF's e último mês de análise

Obs	Instituição	até	Obs	Instituição	até
1	Banco do Brasil	Ago98	24	Banco Pactual	Ago98
2	Banco de Brasília	Ago98	25	Banco Icatu	Ago98
3	Banco BNL do Brasil	Ago98	26	BBA-CREDITANSTALT	Ago98
4	Caixa Econômica Federal	Ago98	27	Banco Garantia	Ago98
5	Banco do Estado de Goiás	Jun98	28	CITIBANK	Ago98
6	Banco da Amazônia	Ago98	29	Banco ABN AMRO	Ago98
7	Banco do Estado do Pará	Nov97	30	Banco Sul América	Ago98
8	Banco Pontual	Mai98	31	Banco Rural	Ago98
9	Banco do Estado do Ceará	Mai98	32	Banco de Boston	Ago98
10	Banco BMC	Ago98	33	Banco Chase Manhattan	Ago98
11	Banco do Nordeste do Brasil	Ago98	34	Banco CCF Brasil	Ago98
12	Banco Industrial e Comercial	Ago98	35	Banco CREDIBANCO	Ago98
13	Banco do Estado de PE	Ago98	36	Banco Boavista Interatlântico	Ago98
14	Banco do Estado da Bahia	Jun98	37	Banco Bozano Simonsen	Ago98
15	Banco Marka	Ago98	38	UNIBANCO	Ago98
16	Banco Real	Ago98	39	Banco Excel Econômico	Ago98
17	Banco Mercantil do Brasil	Ago98	40	Banco Stock	Ago98
18	Banco do Estado de MG	Ago98	41	Banco Multiplic	Jul98
19	Banco General Motors	Ago98	42	Banco Interatlântico	Fev98
20	Banco Royal de Investimentos	Nov97	43	Nossa Caixa-Nosso Banco	Ago98
21	Banco Banestes	Nov97	44	Morgan Guaranty Trust Co Ny	Ago98
22	Banco ABC do Brasil	Ago98	45	Banco Norchem	Jun97
23	Banco Ômega	Ago98	46	ING Bank NV	Ago98

(continua)

Tabela 2.1 IF's e último mês de análise
(continuação)

Obs	Instituição	até	Obs	Instituição	até
47	Continental Banco	Ago98	69	Lloyds Bank PLC.	Ago98
48	Credinvest CFI	Dez97	70	Banco Tokyo	Mar96
49	Banco Safra	Ago98	71	Banco Sogeral	Ago98
50	Banco Fibra	Ago98	72	Banco Itamarati	Jan97
51	Banco Inter American Express	Ago98	73	Banco Patente	Ago98
52	Banco Votorantim	Ago98	74	Deustche Bank	Ago98
53	AGF Braseg Leasing	Jan97	75	Banco do Estado do Paraná	Nov97
54	Dresdner Bank Lateinamerika	Ago98	76	Banco do Estado de SC	Ago98
55	Banco Tokyo-Mitsubishy Brasil	Ago98	77	Banco Meridional do Brasil	Ago98
56	Banco Sumitomo Brasileiro	Ago98	78	Banco do Estado do RS	Fev98
57	Banco Digibanco	Mar97	79	Banco Itabanco	Ago98
58	Banco Santander Noroeste	Ago98	80	Caixa Econ. Estadual do RS	Ago98
59	Banco Itaú	Ago98	81	Banco Banorte	Nov96
60	Banco Bradesco	Ago98	82	Banco do Estado de Alagoas	Dez94
61	Banco Francês Brasileiro	Ago98	83	Banco Econômico	Jul95
62	Banco de Crédito Nacional	Ago98	84	Banco Nacional	Out95
63	Banco Sudameris	Ago98	85	Banco Progresso	Jan97
64	Banco Mercantil de São Paulo	Ago98	86	Banerj	Nov94
65	Banco Bandeirantes	Ago98	87	Banco Rosa	Fev95
66	Banco Europeu p/ Am. Latina	Ago98	88	Banco Open	Dez94
67	Banco Dibens	Ago98	89	Banespa	Nov94
68	Banco América do Sul	Ago98	90	Banco Bamerindus	Fev97

Exemplificando: para o Banco do Brasil, índice de Liquidez Imediata (L1) obtiveram-se as 6 últimas observações correspondentes aos meses de março até agosto de 1998 e dentre esses dados selecionou-se o menor valor.

2.5 Intervenção e Liquidação Extrajudicial

De acordo com a lei nº 6024, de 13 de março de 1974, “as instituições financeiras privadas e as públicas não federais, assim como as cooperativas de crédito, estão sujeitas, nos termos desta lei, à intervenção ou à liquidação extrajudicial, em ambos os casos efetuada e decretada pelo Banco Central do Brasil...”.

Além disso, pela citada lei: “Far-se-á a intervenção quando se verificarem as seguintes anormalidades nos negócios sociais da instituição:

I - a entidade sofrer prejuízo, decorrente de má administração, que sujeite a riscos os seus credores;

II - forem verificadas reiteradas infrações a dispositivos da legislação bancária não regularizadas após as determinações do Banco Central do Brasil, no uso das suas atribuições de fiscalização;

III - na hipótese de ocorrer qualquer dos fatos mencionados nos arts. 1 e 2 do Decreto-lei nº 7661, de junho de 1945 (Lei de Falências), houver possibilidade de evitar-se a liquidação extrajudicial.”

A liquidação extrajudicial fica decretada quando (Lei nº 6024):

“I - ex officio:

a) em razão de ocorrências que comprometam sua situação econômica ou financeira especialmente quando deixar de satisfazer, com pontualidade, seus compromissos ou quando se caracterizar qualquer dos motivos que autorizem a declaração de falência;

b) quando a administração violar gravemente as normas legais e estatutárias que dis-

ciplinam a atividade da instituição, bem como as determinações do Conselho Monetário Nacional ou do Banco Central do Brasil, no uso de suas atribuições legais;

c) quando a instituição sofrer prejuízo que sujeite a risco anormal seus credores quirografários;

d) quando, cassada a autorização para funcionar, a instituição não iniciar, nos 90 (noventa) dias seguintes, sua liquidação ordinária, ou quando, iniciada esta, verificar o Banco Central do Brasil que a morosidade de sua administração pode acarretar prejuízo para os credores;

II - a requerimento dos administradores da instituição - se o respectivo estatuto social lhes conferir esta competência - ou por proposta do interventor, expostos circunstanciadamente os motivos justificadores da medida.”

As seguintes IF's estiveram sujeitas à intervenção ou liquidação extra-judicial durante o período de análise:

Tabela 2.2 Instituições Financeiras que sofreram intervenção

Instituição Financeira	Data da intervenção ou liquidação
Banco Banorte	Mai de 1996
Banco do Est. de Alagoas	Janeiro de 1995
Banco Econômico	Agosto de 1995
Banco Nacional	Novembro de 1995
Banco Progresso	Fevereiro de 1997
Banerj	Dezembro de 1994
Banco Rosa	Março de 1995
Banco Open	Janeiro de 1995
Banespa	Novembro de 1995
Bamerindus	Março de 1997

Capítulo 3

Análise Descritiva e Testes de Hipóteses

A primeira idéia para atingir os objetivos deste trabalho é aplicar uma Análise Discriminante (Johnson e Wichern, 1998) para tentar separar os dois grupos de IF's (sem e com problemas). Essa técnica requer que os dados sigam uma distribuição normal multivariada, podendo requerer também igualdade das matrizes de covariâncias dos dois grupos, dependendo se a função discriminante é linear ou não. Mais detalhes sobre a metodologia podem ser encontrados no Capítulo 4.

Com o objetivo de conhecer o comportamento dos indicadores econômico-financeiros e avaliar se as suposições requeridas para a aplicação das técnicas estatísticas estão satisfeitas, apresentam-se neste capítulo, a análise descritiva, testes para normalidade e testes de diferenças de médias univariados.

Conforme citado anteriormente, foi calculada uma medida resumo dos indicadores para cada IF, sobre as quais baseiam-se todas as análises mostradas a seguir.

3.1 Análise das Observações

O primeiro passo foi construir “boxplots” (Bussab e Morettin, 1987) para cada um dos indicadores. Esses gráficos, de números A1 a A17 estão no Apêndice A. Para os indicadores L1 e L2 havia uma observação com comportamento fortemente diferente das demais (Banco Multiplic e Banco Interatlântico, respectivamente). Assim, estas observações foram retiradas da análise descritiva e dos testes de hipóteses. Os gráficos apresentados já excluem essas observações.

Pelos gráficos pode-se verificar que existe uma quantidade grande de dados discrepantes; o único indicador a não apresentá-los é o L6, cujos valores variam de 0 a 1 pois constituem-se da Participação dos Depósitos Interfinanceiros no Total dos Depósitos. Outro ponto a ser considerado é que, para todos os indicadores, as caixas dos Grupos 0 e 1 estão próximas, indicando que cada indicador separadamente não deve discriminar bem os grupos. Em termos gerais, as caixas são maiores para o Grupo 1.

As médias, os desvios-padrão e os tamanhos das amostras para cada grupo encontram-se na Tabela 3.1, assim como o sentido do indicador, denotando quanto maior (menor) o valor do indicador, melhor a situação da IF.

É de se esperar que o Grupo 0 (IF's sem problemas) tenha desempenho médio melhor do que o Grupo 1 (IF's com problemas). Em geral, as médias amostrais seguem esse sentido, com exceção apenas do indicador L14, que apresenta média amostral da Taxa de Retorno das Aplicações maior no Grupo 1 do que no Grupo 0.

Tabela 3.1 Médias, desvios-padrão e tamanhos das amostras dos indicadores por grupo

Indicador	Grupo 0			Grupo 1			Sentido (melhor)
	n	Média	Desvio-padrão	n	Média	Desvio-padrão	
L1	74	3,84	9,57	10	0,86	0,71	>
L2	79	1,44	2,54	10	0,96	0,10	>
L3	80	0,72	1,06	10	2,25	2,46	<
L4	80	16,83	32,62	10	22,72	15,43	<
L5	80	0,96	0,25	10	1,08	0,17	<
L6	76	0,29	0,31	10	0,41	0,30	<
L7	80	0,95	0,13	10	0,60	0,44	>
L8	79	0,70	2,29	10	1,03	2,60	<
L9	80	0,78	2,09	10	7,58	16,65	<
L10	80	0,04	0,07	10	0,02	0,02	>
L11	80	-0,12	0,40	10	-0,24	0,26	>
L12	80	-0,11	0,37	10	-0,61	0,97	>
L13	78	-0,15	0,56	10	-1,32	2,15	>
L14	79	0,02	0,01	10	0,07	0,06	>
L15	77	0,14	0,12	10	0,53	0,76	<
L16	80	0,21	0,20	10	0,18	0,16	>
L17	80	0,04	0,09	10	0,06	0,05	<

O teste de Anderson-Darling para normalidade (D’Augustino e Stevens, 1986) foi aplicado a cada indicador. Os resultados dos testes (estatísticas e níveis descritivos) são apresentados na Tabela 3.2. Somente para os indicadores L1, L2, L6, L11, L16 e L17, no Grupo 1, a normalidade não foi rejeitada ao nível de 5%. No Grupo 0, a normalidade foi rejeitada para todos os indicadores a níveis menores do que 0,1%. É importante salientar que o Grupo 1 tem poucas IF’s e por isso é possível não estarmos detectando

a não normalidade.

Tabela 3.2 Teste de Anderson-Darling para normalidade dos indicadores

Indicador	Grupo 0		Grupo 1	
	A^2	P	A^2	P
L1	27,82	0,000	0,52	0,137
L2	29,96	0,000	0,37	0,352
L3	6,80	0,000	1,24	0,002
L4	12,03	0,000	0,70	0,047
L5	11,20	0,000	0,95	0,010
L6	4,36	0,000	0,30	0,509
L7	10,14	0,000	1,05	0,005
L8	22,00	0,000	2,66	0,000
L9	13,40	0,000	2,21	0,000
L10	19,94	0,000	0,82	0,021
L11	13,84	0,000	0,55	0,115
L12	18,83	0,000	1,46	0,000
L13	19,56	0,000	1,31	0,001
L14	6,33	0,000	0,79	0,026
L15	8,95	0,000	2,02	0,000
L16	2,44	0,000	0,62	0,076
L17	13,17	0,000	0,45	0,216

Na Tabela 3.3 temos o valor da estatística t-Student, o número de graus de liberdade e o respectivo nível descritivo (P) dos testes univariados de igualdade de médias. Para os indicadores L5, L6, L8, L11, L16 e L17 foi aplicado o teste de comparação de médias para amostras independentes e variâncias iguais, equivalente à análise de variância com um fator. Para os indicadores L1, L2, L3, L4, L7, L9, L10, L12, L13, L14 e L15 foi

usado o teste t aproximado, que permite que as variâncias sejam diferentes (Bussab e Morettin, 1987). Essa decisão foi guiada pelos valores dos desvios-padrão apresentados na Tabela 3.1.

Tabela 3.3 Testes de igualdade de médias dos indicadores

Variável	t	gl	P
L1	2,62	78	0,011
L2	1,67	79	0,099
L3	-1,94	9	0,084
L4	-0,97	21	0,340
L5	-1,47	88	0,140
L6	-1,21	84	0,230
L7	2,51	9	0,033
L8	-0,43	87	0,670
L9	-1,29	9	0,230
L10	1,96	52	0,055
L11	0,96	88	0,340
L12	1,62	9	0,140
L13	1,72	9	0,120
L14	-2,66	9	0,026
L15	-1,62	9	0,140
L16	0,45	88	0,660
L17	-0,77	88	0,440

Pela Tabela 3.3, apresentaram diferença de médias entre os Grupos 0 e 1 ao nível de significância de 5%, as variáveis L1, L7 e L14. É importante ressaltar que os testes podem ser afetados pela não normalidade, em especial o Grupo 1, que tem poucas observações. Salienta-se ainda, que não necessariamente as variáveis que apresentaram diferença de

médias entre os grupos serão as que melhor contribuem para a correta classificação das IF's. Como exemplo, em Espahbodi (1991), as técnicas de seleção automática de variáveis escolheram para os modelos finais variáveis consideradas não significantes na análise univariada e Cochran (1964) mostrou que variáveis não significantes no contexto univariado podem ser importantes quando combinadas com outras.

3.2 Análise das Observações “Rankeadas”

Considerando a não normalidade dos dados verificou-se a necessidade de transformá-los de maneira que se adequassem melhor à técnica a ser aplicada. Algumas transformações tradicionais não foram utilizadas porque os dados contêm valores negativos. Quando isso ocorre, uma alternativa é somar uma constante a todas as observações para que estas não sejam negativas e então aplicar a transformação. Outra opção seria usar transformações diferentes para indicadores diversos, mas isto levaria a difíceis interpretações. Optou-se então por utilizar a análise discriminante aplicada ao “rank” dos indicadores, denominados de RL1 a RL17.

Em primeiro lugar fizemos os testes de Anderson-Darling para normalidade. As correspondentes estatísticas dos testes e os níveis descritivos encontram-se na Tabela 3.4. Pode-se notar que houve uma sensível melhora nos resultados após a transformação. Considerando o nível de 1%, rejeita-se a normalidade somente para o indicador RL16 no Grupo 0 e para os indicadores RL1, RL12 e RL13 no Grupo 1.

É importante lembrar que mesmo que a normalidade fosse não rejeitada individualmente para cada indicador, ainda assim isso não indicaria normalidade multivariada. O que parece é que a fuga de normalidade é menor para os dados “rankeados”.

Tabela 3.4 Teste de Anderson-Darling para Normalidade do “rank” dos indicadores

Indicador	Grupo 0		Grupo 1	
	A^2	P	A^2	P
RL1	0,709	0,062	1,023	0,006
RL2	0,801	0,038	0,404	0,287
RL3	0,880	0,023	0,531	0,129
RL4	0,990	0,012	0,460	0,202
RL5	0,825	0,032	0,363	0,365
RL6	0,938	0,017	0,224	0,757
RL7	0,926	0,018	0,659	0,059
RL8	0,979	0,013	0,166	0,912
RL9	0,871	0,024	0,478	0,182
RL10	0,837	0,030	0,823	0,021
RL11	0,738	0,052	0,719	0,041
RL12	0,665	0,080	1,233	0,002
RL13	0,628	0,098	1,001	0,007
RL14	0,790	0,039	0,942	0,010
RL15	0,738	0,052	0,379	0,332
RL16	1,094	0,007	0,172	0,901
RL17	0,794	0,038	0,918	0,012

Os “boxplots” dos dados transformados estão no Apêndice A (Gráficos A18 a A34). As médias, medianas, desvios-padrão e tamanhos das amostras para cada grupo em estudo (Tabela 3.5) estão a seguir.

Tabela 3.5. Médias, medianas, desvios-padrão e tamanhos das amostras do “rank” dos indicadores

Indicador (“rank”)	Grupo 0				Grupo 1				Sentido (melhor)
	n	Média	Mediana	Desvio- padrão	n	Média	Mediana	Desvio- padrão	
RL1	75	45,75	46,0	23,73	10	22,40	12,0	22,79	>
RL2	80	47,71	48,5	25,55	10	27,80	21,5	25,01	>
RL3	80	42,65	41,5	25,31	10	68,30	77,0	21,86	<
RL4	80	43,91	44,5	26,00	10	58,20	54,0	24,81	<
RL5	80	43,60	43,0	25,70	10	60,70	66,5	25,71	<
RL6	76	41,96	40,5	25,25	10	55,20	54,5	19,99	<
RL7	80	49,08	50,5	25,04	10	16,90	14,5	14,89	>
RL8	79	44,23	44,0	25,98	10	51,10	52,5	24,98	<
RL9	80	43,03	42,5	25,33	10	65,30	73,5	24,93	<
RL10	80	47,08	47,5	24,67	10	32,90	19,5	34,75	>
RL11	80	48,01	48,5	25,84	10	25,40	18,5	19,63	>
RL12	80	48,15	48,5	24,41	10	24,30	10,5	30,96	>
RL13	78	47,26	47,5	24,17	10	23,00	12,5	27,07	>
RL14	79	41,44	41,0	24,44	10	73,10	79,5	19,00	>
RL15	77	40,73	40,0	24,53	10	69,20	71,0	14,73	<
RL16	80	45,69	46,5	26,67	10	44,00	42,5	22,37	>
RL17	80	43,90	42,5	25,12	10	58,30	68,5	31,70	<

Pela Tabela 3.5 e Gráficos A18 a A34 (Apêndice A) nota-se que os problemas com as observações discrepantes da análise anterior e a grande variabilidade apresentada por alguns indicadores são contornados. Somente o indicador RL15 apresenta um dado discrepante no Grupo 1. Assim, como no caso das variáveis originais, para o “rank” da

variável L14 (Taxa de Retorno das Aplicações) as médias e medianas amostrais dos grupos não condizem com o que se espera do indicador, ou seja, era esperado Taxa de Retorno das Aplicações maior no Grupo 0.

Os gráficos mostram que as medianas dos indicadores RL4, RL5, RL6, RL8, RL10, RL16 e RL17 parecem ser iguais para os dois grupos, as medianas dos indicadores RL1, RL2, RL7, RL11, RL12 e RL13 parecem maiores no Grupo 0, ocorrendo o oposto para os indicadores RL3, RL9, RL14 e RL15.

Na Tabela 3.6 apresentam-se os resultados dos testes de igualdade de médias e medianas dos dois grupos. Para a média, foi utilizado o teste t com a suposição de normalidade e igualdade de variâncias pois, embora a normalidade tenha sido rejeitada para poucos indicadores, o teste não é altamente afetado nessa situação (Cochran, 1947). Para a mediana foi aplicado o teste de Mann Whitney (Conover, 1971)

Apresentaram diferença de médias e de medianas entre os grupos ao nível de significância de 5% as variáveis RL1, RL2, RL3, RL7, RL9, RL11, RL12, RL13, RL14 e RL15, confirmando a análise feita sobre os gráficos. Para as variáveis originais (Tabela 3.3), somente três diferenças haviam sido detectadas.

Tabela 3.6 Testes de igualdades de médias e medianas do “rank” dos indicadores

Variável	Teste para a média			Teste para a mediana	
	t	gl	P	M	P
RL1	2,93	83	0,004	3431	0,005
RL2	2,33	88	0,022	3817	0,023
RL3	-3,06	88	0,003	3412	0,004
RL4	-1,65	88	0,100	3513	0,104
RL5	-1,98	88	0,050	3488	0,052
RL6	-1,59	84	0,120	3189	0,116
RL7	3,96	88	0,000	3926	0,000
RL8	-0,79	87	0,430	3494	0,432
RL9	-2,63	88	0,010	3442	0,011
RL10	1,63	88	0,110	3766	0,107
RL11	2,67	88	0,009	3841	0,010
RL12	2,83	88	0,006	3852	0,007
RL13	2,95	86	0,004	3686	0,005
RL14	-3,94	87	0,000	3274	0,000
RL15	-3,58	85	0,001	3136	0,001
RL16	0,19	88	0,850	3655	0,852
RL17	-1,66	88	0,100	3512	0,102

Capítulo 4

Metodologia

Neste capítulo apresentam-se as diferentes técnicas estatísticas aplicadas ao problema, que se trata de discriminação de duas populações. As técnicas abaixo descritas podem ser estendidas para mais do que duas populações, mas este não é o caso de interesse. As referências aqui citadas também trazem essa extensão.

4.1 Técnicas Utilizando as Variáveis Originais

4.1.1 Análise Discriminante

A análise discriminante consiste em técnica multivariada que visa a obtenção de função que melhor discrimine distintos conjuntos de observações e também permita a alocação de novas observações aos grupos previamente definidos.

Apresenta-se a metodologia da análise discriminante apenas para o caso em que temos um vetor de variáveis aleatórias $\mathbf{x}_{(p \times 1)}$ de uma de duas populações (π_0 e π_1), com funções densidade de probabilidade f_0 e f_1 e respectivos vetores de médias μ_0 e μ_1 . Quando as matrizes de variância-covariância dos grupos (Σ_0 e Σ_1) são idênticas temos a análise discriminante linear; com diferentes matrizes de dispersão, temos a análise discriminante quadrática. Esta metodologia pode ser encontrada em livros de Análise Multivariada como Mardia et al. (1979), Johnson e Wichern (1998) e em SAS/STAT User's Guide Volume 1 (1989).

A análise discriminante baseia-se em obter a função dos dados que minimize o Custo

Esperado de Classificação Errada (*CEE*), que pode levar em consideração o custo de se classificar um objeto na população errada e também a probabilidade a priori de um objeto qualquer pertencer a cada uma das populações. O *CEE* é dado por

$$CEE = c(1|0)p(1|0)p_0 + c(0|1)p(0|1)p_1,$$

onde

$c(i|j)$: custo de classificar a observação na população π_i quando ela realmente pertence à população π_j ;

$p(i|j)$: probabilidade de classificar a observação na população π_i quando ela realmente pertence à população π_j ;

p_i : probabilidade a priori de classificar a observação na população π_i .

Considerando $p_0 = p_1$ e $c(1|0) = c(0|1)$, minimiza-se o *CEE* ao classificar uma nova observação \mathbf{x}_c nas populações

$$\pi_0 \text{ se } f_0(\mathbf{x}_c) \geq f_1(\mathbf{x}_c), \quad (4.1)$$

$$\pi_1 \text{ se } f_1(\mathbf{x}_c) > f_0(\mathbf{x}_c). \quad (4.2)$$

Se assumirmos que as duas populações seguem distribuições normais multivariadas com matrizes de variância-covariância que não coincidem, a função densidade de probabilidade para a i -ésima população é dada por

$$f_i(\mathbf{x}) = [(2\pi)^{p/2} |\Sigma_i|^{1/2}]^{-1} \exp[-1/2(\mathbf{x} - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)], \quad (4.3)$$

onde

$i = 0, 1$;

$\mathbf{x} = [x_1, x_2, \dots, x_p]'$: vetor de variáveis;

$\boldsymbol{\mu}_i$ = vetor de médias da i -ésima população;

Σ_i = matriz de variância-covariância da i -ésima população.

Para o caso de duas populações normais, com mesmas matrizes de variância-covariância, a função densidade para a i -ésima população é dada por

$$f_i(\mathbf{x}) = [(2\pi)^{p/2} |\Sigma|^{1/2}]^{-1} \exp[-1/2(\mathbf{x} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)], \quad (4.4)$$

onde

$i = 0, 1$;

$\mathbf{x} = [x_1, x_2, \dots, x_p]'$: vetor de variáveis;

$\boldsymbol{\mu}_i$ = vetor de médias da i -ésima população;

Σ = matriz de variância-covariância comum.

No caso quadrático, substituindo (4.3) em (4.1) e (4.2), classifica-se a observação \mathbf{x}_c em

$$\pi_0 \text{ se } -(\mathbf{x}_c - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_0) + (\mathbf{x}_c - \boldsymbol{\mu}_1)' \Sigma_1^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_1) \geq \log \frac{|\Sigma_0|}{|\Sigma_1|}, \quad (4.5)$$

$$\pi_1 \text{ se } -(\mathbf{x}_c - \boldsymbol{\mu}_0)' \Sigma_0^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_0) + (\mathbf{x}_c - \boldsymbol{\mu}_1)' \Sigma_1^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_1) < \log \frac{|\Sigma_0|}{|\Sigma_1|}. \quad (4.6)$$

No caso linear, substituindo (4.4) em (4.1) e (4.2), temos a observação \mathbf{x}_c classificada em

$$\pi_0 \text{ se } -(\mathbf{x}_c - \boldsymbol{\mu}_0)' \Sigma^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_0) \geq -(\mathbf{x}_c - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_1), \quad (4.7)$$

$$\pi_1 \text{ se } -(\mathbf{x}_c - \boldsymbol{\mu}_0)' \Sigma^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_0) < -(\mathbf{x}_c - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x}_c - \boldsymbol{\mu}_1). \quad (4.8)$$

Segundo a literatura, a falta de normalidade é mais crítica e afeta mais os resultados no caso quadrático. Na prática, obtendo-se amostras aleatórias de tamanhos n_0 e n_1 de π_0 e π_1 , respectivamente, a regra de classificação amostral é obtida substituindo os parâmetros pelos seus estimadores, dados na Tabela 4.1.

Tabela 4.1. Estimadores dos parâmetros

Parâmetro	Estimador
μ_0	$\bar{x}_0 = \frac{1}{n_0} \sum_1^{n_0} x_{0j}$
μ_1	$\bar{x}_1 = \frac{1}{n_1} \sum_1^{n_1} x_{1j}$
Σ_0	$S_0 = \frac{1}{n_0-1} \sum_1^{n_0} (x_{0j} - \bar{x}_0)(x_{0j} - \bar{x}_0)'$
Σ_1	$S_1 = \frac{1}{n_1-1} \sum_1^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)'$
Σ	$S = \left[\frac{n_0-1}{(n_0-1)+(n_1-1)} \right] S_0 + \left[\frac{n_1-1}{(n_0-1)+(n_1-1)} \right] S_1$

Com relação à interpretação da função discriminante, três são os principais critérios encontrados na literatura. O primeiro é a interpretação através dos coeficientes das variáveis independentes: variáveis com altos coeficientes contribuem mais para o poder de discriminação da função. O sinal indica se a contribuição é positiva ou negativa. O segundo é a interpretação pelas cargas discriminantes, que são as correlações lineares entre as variáveis independentes e a função discriminante: as cargas indicam a contribuição relativa de cada variável independente. Por fim, quando o procedimento “stepwise” é utilizado, uma interpretação do poder discriminante relativo de cada variável independente pode ser feita através do valor das estatísticas F parciais: grandes valores de F indicam maior poder de discriminação.

Em qualquer um dos casos, principalmente se o procedimento “stepwise” foi usado, a interpretação deve ser feita com cuidado. Um coeficiente pequeno pode acontecer não porque a variável independente correspondente não seja importante, mas sim por causa de multicolinearidade fazendo com que uma variável correlacionada com esta já esteja no modelo. O valor dos F 's parciais seguem a mesma ordem do valor dos coeficientes da função discriminante. Uma discussão mais detalhada sobre este aspecto pode ser encontrada em Hair et al (1995).

Com a finalidade de estimar o percentual de classificação errada, calcula-se a Taxa Aparente de Erro - TAE com base no número de observações da amostra alocadas a cada população, esquematizada abaixo. A TAE estima a Taxa Real de Erro.

	Predito		
Real	π_0	π_1	Total
π_0	n_{00}	n_{01}	n_0
π_1	n_{10}	n_{11}	n_1
Total	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

$$TAE = \frac{n_{01} + n_{10}}{n},$$

onde

n_{ij} = número de observações da população π_i alocadas em π_j ;

$i, j = 0, 1$;

$\frac{n_{01}}{n_0}$ e $\frac{n_{10}}{n_1}$ são, respectivamente, as taxas de erro de observações de π_0 classificadas em π_1 e de π_1 classificadas em π_0 .

O uso simultâneo de todas as observações para avaliar o índice de classificação da análise discriminante apresenta resultados viesados, pois as mesmas observações são usadas tanto para construir a função de classificação quanto para fazer a avaliação que, dessa forma, subestima a Taxa Real de Erro. Uma alternativa, sugerida por Lachenbruch e Mickey (1968), é um tipo de “crossvalidation”, onde cada uma das observações é retirada da amostra (uma de cada vez), a função de classificação é construída e a observação é alocada em uma das populações com base nessa função. Esse procedimento é repetido n vezes, um para cada observação. A TAE é calculada como acima.

Seleção Automática de Variáveis

Na análise discriminante a seleção automática de variáveis é feita considerando o nível de significância de um teste F para análise de covariância (Wilk’s lambda, critério da razão de verossimilhança, Mardia et al (1979)). Na primeira iteração, considerando o caso em que nenhuma variável foi incluída no modelo, dá-se início à seleção considerando cada variável quantitativa separadamente. Essas atuam como variáveis dependentes

em modelos de análise de covariância e a variável classificatória atua como variável independente. Nas iterações seguintes, as variáveis escolhidas atuam como covariáveis, juntamente com a variável classificatória. As demais variáveis a serem incluídas no modelo atuam como variáveis dependentes em modelos separados, um modelo para cada uma. (SAS/STAT, volume 2, 1989).

Existem três métodos para a obtenção de um conjunto parcimonioso de variáveis capazes de discriminar entre os dois grupos:

- “forward”: em cada iteração, a variável que melhor satisfizer o critério de seleção é adicionada, até que não reste nenhuma outra;
- “backward”: começa com todas as variáveis incluídas e elimina as piores covariáveis, uma por uma, até que todas as restantes no modelo sejam necessárias;
- “stepwise”: combina os 2 procedimentos anteriores; as variáveis vão sendo introduzidas no modelo, mas podem ser retiradas quando outras variáveis são adicionadas.

4.1.2 Regressão Logística

Nesta seção discorreremos sobre a regressão logística que tem se constituído num dos principais métodos de modelagem estatística de dados (Paula, 1997). Maiores informações sobre o assunto podem ser encontradas em Hosmer e Lemeshow (1989), McCullagh e Nelder (1989), Paula (1997) e SAS/STAT Volume 2 (1989). Aqui, estamos definindo a regressão logística somente para o caso em que a variável dependente é dicotômica (Grupo 0 e Grupo 1), que é o caso de interesse neste trabalho.

Tendo em vista que utilizando a regressão logística é possível atribuir a cada observação probabilidades de que ela pertença aos grupos de interesse, pode-se então utilizá-la como técnica classificatória, assim como a análise discriminante. A observação é classificada no grupo que apresenta maior probabilidade.

A variável dependente Y é dicotômica e indica se a observação pertence ao Grupo 0, ou seja,

$$Y_j = \begin{cases} 1 & \text{se a observação } j \text{ pertence ao Grupo 0,} \\ 0 & \text{caso contrário.} \end{cases}$$

O modelo logístico é um caso particular dos modelos lineares generalizados, cuja função de ligação é definida por

$$\log \left\{ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right\} = \boldsymbol{\beta}'\mathbf{x},$$

onde

$\mathbf{x} = [1, x_1, x_2, \dots, x_p]'$: vetor de variáveis independentes;

$\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]$: vetor de parâmetros desconhecidos;

$\pi(\mathbf{x}) = E(Y = 1|\mathbf{x}) = P(Y = 1|\mathbf{x})$: probabilidade da observação pertencer ao Grupo 0 dado o vetor de variáveis \mathbf{x} ;

$-\infty < \boldsymbol{\beta}'\mathbf{x} < +\infty$.

Conseqüentemente,

$$\pi(\mathbf{x}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}\}}.$$

Pode-se escrever o modelo como

$$Y = \pi(\mathbf{x}) + \epsilon,$$

onde $\pi(\mathbf{x}) = E(Y|\mathbf{x})$.

Como Y pode assumir somente dois valores, ϵ também pode assumir somente dois valores, ou seja

- se $Y = 1$, então $\epsilon = 1 - \pi(\mathbf{x})$, com probabilidade $\pi(\mathbf{x})$,
- se $Y = 0$, então $\epsilon = -\pi(\mathbf{x})$, com probabilidade $1 - \pi(\mathbf{x})$,

ou seja, ϵ segue uma distribuição com média 0 e variância $\pi(\mathbf{x})(1 - \pi(\mathbf{x}))$. A distribuição condicional de Y é binomial com probabilidade de sucesso igual a $\pi(\mathbf{x})$.

É possível estimar $\pi(\mathbf{x})$ através da estimativa de $\beta' \mathbf{x}$. A regra de classificação consiste em estimar $\pi(\mathbf{x})$ para uma nova observação \mathbf{x}_c : se $\hat{\pi}(\mathbf{x}_c) \geq 0,50$, classificar a observação no Grupo 0, caso contrário, no Grupo 1.

Os estimadores de máxima verossimilhança são obtidos pela solução das $p+1$ equações correspondentes à diferenciação do logaritmo da função de verossimilhança em relação aos $p+1$ elementos do vetor β . Assim, as soluções são obtidas por métodos iterativos, em especial o método de mínimos quadrados ponderados, tal que o passo $(m+1)$ do algoritmo leva a

$$\beta^{(m+1)} = (\mathbf{X}' \mathbf{V}^m \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^m \mathbf{z}^m, \quad (4.9)$$

onde

$m = 0, 1, \dots;$

$\pi_j = P(Y_j = 1/x_j)$: probabilidade de que a j -ésima observação seja classificada no Grupo 0;

$\mathbf{X}_{(n \times p)}$: matriz das variáveis independentes;

$\mathbf{V} = \text{diag}[\pi_1(1 - \pi_1), \pi_2(1 - \pi_2), \dots, \pi_n(1 - \pi_n)]$;

$\mathbf{z} = (z_1, z_2, \dots, z_n)'$: variável dependente modificada;

$z_j = \eta_j + \frac{(y_j - \pi_j)}{[\pi_j(1 - \pi_j)]}$, $j = 1, 2, \dots, n$;

$\eta_j = \log\left\{\frac{\pi_j}{1 - \pi_j}\right\}$.

Para estimar as variâncias e covariâncias dos estimadores dos parâmetros, o caminho mais comum é utilizar a teoria de verossimilhança, ou seja, obter estimadores pela inversão da matriz de informação dada pelas segundas derivadas parciais do logaritmo da função de verossimilhança na última iteração. A variância assintótica de $\hat{\beta}$ é $(\mathbf{X}' \mathbf{V} \mathbf{X})^{-1}$.

O ajuste do modelo é feito da maneira mais usual, ou seja, através do teste da razão de verossimilhança. A estatística do teste é

$$D = -2 \sum_{j=1}^n \left[y_j \log \left(\frac{\hat{\pi}_j}{y_j} \right) + (1 - y_j) \log \left(1 - \frac{\hat{\pi}_j}{y_j} \right) \right],$$

onde $\hat{\pi}_j = \hat{\pi}(\mathbf{x}_j)$ que, sob a hipótese nula, segue uma distribuição de qui-quadrado.

O teste de Wald (Rao, 1973) é uma alternativa para o teste univariado sobre cada parâmetro ($H : \beta_k = 0, k = 0, 1, 2, \dots, p$). A estatística do teste é dada por

$$w_k = \frac{\hat{\beta}_k}{\hat{EP}(\hat{\beta}_k)},$$

onde $\hat{EP}(\hat{\beta}_k)$ é o estimador do erro padrão de $\hat{\beta}_k$. Sob a hipótese nula, essa estatística tem distribuição assintótica normal padrão.

Para avaliar a regra de classificação, pode-se proceder como na análise discriminante, calculando a *TAE* (Taxa Aparente de Erro) para reclassificação ou validação cruzada. Nesta última, ajustam-se n modelos, cada um deles retirando uma das observações. Cada observação será alocada com base no modelo de cujo ajuste não fez parte.

A vantagem da regressão logística é que é flexível e fácil de usar, permitindo ainda interpretação dos parâmetros, que é feita através da razão de chances. Quando as variáveis independentes são contínuas, o coeficiente é a variação na função logit para o incremento de uma unidade da variável a que corresponde.

4.1.3 Regressão Probit

O modelo de regressão probit é similar ao modelo de regressão logística, somente com uma mudança na função de ligação que é dada por

$$\Phi^{-1}(\pi(\mathbf{x})) = \boldsymbol{\beta}'\mathbf{x},$$

onde $\Phi(\mathbf{x})$: função de distribuição normal acumulada;

$\mathbf{x} = [1, x_1, x_2, \dots, x_p]'$: vetor de variáveis independentes;

$\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]$: vetor de parâmetros desconhecidos;

$\pi(\mathbf{x}) = E(Y = 1/\mathbf{x}) = P(Y = 1/\mathbf{x})$: probabilidade da observação pertencer ao Grupo 0 dado o vetor de variáveis \mathbf{x} ;

$-\infty < \boldsymbol{\beta}'\mathbf{x} < +\infty$.

Conseqüentemente,

$$\pi(\mathbf{x}) = \Phi(\boldsymbol{\beta}'\mathbf{x}).$$

O modelo de regressão probit pode ser ajustado pela PROC LOGISTIC do SAS, com função de ligação probit (normit). Neste caso os estimadores de mínimos quadrados ponderados também são obtidos por métodos iterativos, com o passo $m + 1$ levando a

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}'\mathbf{V}^m\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^m\mathbf{z}^m, \quad (4.10)$$

onde

$$m = 0, 1, \dots;$$

$$\pi_j = \Phi(\boldsymbol{\beta}'\mathbf{x}_j);$$

$$\eta_j = \Phi^{-1}(\pi_j);$$

$\frac{\partial \pi_j}{\partial \eta_j}$: derivada de π_j relativamente a η_j ;

$$\mathbf{V} = \text{diag}\left[\left\{\frac{\partial \pi_j}{\partial \eta_j}\right\}^2 / \pi_j(1 - \pi_j)\right];$$

$\mathbf{z} = (z_1, z_2, \dots, z_n)'$: variável dependente modificada;

$$z_j = \eta_j + (y_j - \pi_j) \frac{\partial \eta_j}{\partial \pi_j}, \quad j = 1, 2, \dots, n.$$

O processo iterativo neste caso é o de Newton-Raphson e os outros testes são similares aos aplicados na seção anterior.

4.1.4 Medidas de Diagnóstico de Regressão

Medidas resumo são habitualmente utilizadas com o intuito de verificação do ajuste geral do modelo. Nos modelos de regressão aqui expostos há várias possíveis medidas das diferenças entre os valores observados e os valores ajustados pelo modelo (resíduos).

Apresentam-se a seguir as medidas de diagnóstico de regressão desenvolvidas por Pregibon (1981) e utilizadas para verificar a adequabilidade do modelo. Paula (1997), Hosmer e Lemeshow (1989), SAS/STAT Volume 2 (1989), SAS/System for Regression (1991) e Souza (1998) também podem ser consultados a respeito dessas medidas.

Considere a notação:

$\mathbf{x}' = (x_1, x_2, \dots, x_p)$: vetor de p variáveis independentes;

L : número de distintos valores observados de \mathbf{x} ;

\mathbf{b} : vetor estimador de $\boldsymbol{\beta}$;

$m_l, l = 1, 2, \dots, L$: número de observações com $\mathbf{x} = \mathbf{x}_l$;

y_l : número de vezes que Y assume o valor 1 dentre as observações que apresentam $\mathbf{x} = \mathbf{x}_l$;

$\hat{\pi}_l = \hat{\pi}(\mathbf{x}_l)$: estimativa de $\pi(\mathbf{x})$ quando $\mathbf{x} = \mathbf{x}_l$.

Pode-se verificar que $\sum_{l=1}^L y_l = n_1$, ou seja, é o número de elementos pertencentes ao Grupo 0. Além disso, tem-se que $L \leq n$ e quando as variáveis independentes são contínuas, L tende a ser próximo ou até mesmo igual a n .

Resíduo de Pearson e Desvio Residual

Duas medidas de diferença entre os valores observados e os valores ajustados são os resíduos de Pearson e o desvio residual, que ajudam a identificar observações que não são bem explicadas pelo modelo. Neste caso, tem-se observações atípicas para Y , que é a variável que classifica a observação em um dos grupos. A atipicidade é identificada caso Y não possa ser predito pelo modelo.

O resíduo de Pearson para o l -ésimo valor de \mathbf{x} é

$$\chi_l = \frac{(y_l - m_l \hat{\pi}_l)}{\sqrt{m_l \hat{\pi}_l (1 - \hat{\pi}_l)}}.$$

A estatística resumo é

$$\chi^2 = \sum_{l=1}^L \chi_l^2.$$

O desvio residual para o l -ésimo valor de \mathbf{x} é definido por

$$d_l = \pm \left\{ 2 \left[y_l \log \left(\frac{y_l}{m_l \hat{\pi}_l} \right) + (m_l - y_l) \log \left(\frac{(m_l - y_l)}{m_l (1 - \hat{\pi}_l)} \right) \right] \right\}^{1/2}$$

onde o sinal mais/menos é o mesmo sinal de $(y_l - m_l \hat{\pi}_l)$.

A estatística resumo é

$$D = \sum_{l=1}^L d_l^2.$$

Sob a suposição de que o modelo ajustado é adequado, as distribuições das estatísticas χ^2 e D são supostas qui-quadrado com $L - (p + 1)$ graus de liberdade. Valores altos da estatística indicam que o modelo não é adequado. Quando L é próximo de n , McCullagh e Nelder (1983) argumentam que os valores esperados de χ^2 e D são menores do que o estabelecido na distribuição de qui-quadrado com $L - p - 1$ graus de liberdade e sugerem um fator de correção para a estatística. Entretanto, Hosmer e Lemeshow (1989) argumentam que a correção traz uma redução muito grande e que na prática o uso de $L - p - 1$ graus de liberdade fornece uma estimativa razoável do valor esperado quando o modelo é adequado.

Teste de Hosmer e Lemeshow

Hosmer e Lemeshow propuseram um teste de ajuste baseado nas probabilidades estimadas, segundo dois métodos:

1. baseado nos percentis;
2. baseado em valores fixados.

As probabilidades estimadas são ordenadas da menor para a maior e g agrupamentos são formados. No primeiro método, cada agrupamento tem n/g observações; no segundo, os agrupamentos são formados de acordo com os percentis de ordem g . Se, por exemplo, $g = 10$, o primeiro agrupamento tem as observações tais que as probabilidades estimadas sejam menores ou iguais a 0,1; no segundo, as observações com probabilidades estimadas entre 0,1 e 0,2 e assim por diante até o décimo agrupamento que contém as observações estimadas maiores do que 0,9.

A estatística do teste é obtida pelo cálculo da estatística de qui-quadrado de Pearson

aplicado a uma tabela $2 \times g$, ou seja, é dada por

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)},$$

onde

n'_k : número de observações no k -ésimo agrupamento;

o_k : número de observações no k -ésimo agrupamento com $Y = 1$, ou seja, pertencentes ao Grupo 0;

$\bar{\pi}_k$: média das probabilidades estimadas das observações pertencentes ao k -ésimo agrupamento.

Com base em estudos de simulação, Hosmer e Lemeshow (1989) mostraram que, quando $L = n$ e o modelo ajustado é correto, a distribuição de \hat{C} é bem aproximada pela distribuição de qui-quadrado com $g - 2$ graus de liberdade. Valores altos da estatística indicam falta de ajuste.

Matriz “Chapéu” Diagonal

Os elementos da matriz “chapéu” diagonal são úteis para identificar medidas atípicas (pontos extremos) para as variáveis exógenas (\mathbf{x}).

Os elementos da matriz são dados por

$$h_j = \hat{\pi}_j(1 - \hat{\pi}_j)(1, \mathbf{x}'_j)(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}(1, \mathbf{x}'_j)',$$

onde $j = 1, 2, \dots, n$ e $\hat{\pi}_j$ e \mathbf{x}'_j são os mesmos como definidos no início desta seção, $\mathbf{X}_{(n \times p)}$ é a matriz das variáveis independentes e $\mathbf{V}_{(n \times n)}$ é uma matriz diagonal com os elementos dados por $\hat{\pi}_j(1 - \hat{\pi}_j)$.

Quando $h_j > \frac{2(m+1)}{n}$ (onde n é o número de observações e $m + 1$ é o número de parâmetros do modelo) tem-se que a j -ésima observação é atípica para os valores da matriz \mathbf{X} .

Destaca-se, porém, que uma observação atípica pode ou não ser influente. Uma

observação é influente se alguma estatística resultante do modelo for sensível (estiver sujeita à mudanças) à sua presença.

DFBetas

Esta estatística pode ser usada para identificar observações influentes no modelo. Os *DFBetas* “medem” a influência da j -ésima observação individualmente no k -ésimo parâmetro do modelo. Caso essa influência exista, tem-se que a sua retirada implicará em significativa mudança no parâmetro em questão. A estatística é definida por

$$DFBeta(k_j) = \frac{\Delta_k b_j^1}{\hat{\sigma}(b_k)},$$

onde

$$k = 0, 1, \dots, p;$$

$\hat{\sigma}(b_k)$ é o erro padrão estimado do k -ésimo componente de \mathbf{b} ;

$\Delta_k b_j^1$ é o k -ésimo componente da primeira diferença Δb_j^1 ;

$$\Delta b_j^1 = \frac{(y_j - \hat{\pi}_j)}{1 - h_j} (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1} \mathbf{x}_j.$$

Valores altos da estatística indicam influência da j -ésima observação no k -ésimo coeficiente de β .

C e Cbar

Medidas baseadas na distância de Cook proveniente da teoria da regressão linear. A distância de Cook mede mudanças nas estimativas ocorridas devido à retirada da j -ésima observação. C e $Cbar$ indicam a influência de observações individualmente no vetor de parâmetros \mathbf{b} .

$$C_j = \frac{\chi_j^2 h_j}{(1 - h_j)^2} \quad \text{e} \quad Cbar_j = \frac{\chi_j^2 h_j}{(1 - h_j)},$$

onde χ_j^2 é o j -ésimo resíduo de Pearson e h_j é o j -ésimo elemento da matriz “chapéu” diagonal. Essas estatísticas medem o decréscimo no valor da estatística qui-quadrado de Pearson quando a j -ésima observação é excluída. Valores altos indicam influência da observação correspondente.

4.1.5 Seleção Automática de Variáveis (regressão)

O teste escore utilizado para selecionar variáveis nos modelos de regressão aqui apresentados é dado por

$$T = U'(\beta_0)I^{-1}(\beta_0)U(\beta_0),$$

onde

$U(\beta) = \frac{\partial L(\beta;y)}{\partial \beta}$: vetor de derivadas parciais do logaritmo da função de verossimilhança em relação ao vetor de parâmetros β ;

$I(\beta) = - \frac{\partial^2 L(\beta;y)}{\partial \beta \partial \beta'}$: negativo do vetor de segundas derivadas parciais do logaritmo da função de verossimilhança em relação ao vetor de parâmetros β .

Sob a hipótese nula $H_0 : \beta = \beta_0$, o teste escore (T) tem distribuição assintótica χ^2 com $p + 1$ graus de liberdade.

Ao utilizarem-se as opções “forward”, “backward” ou “stepwise” é calculada a estatística χ^2 residual, a qual consiste em comparar os modelos sob as hipóteses H_0 (nula) e H_1 (alternativa). Essa estatística tem distribuição assintótica χ^2 com $s - t$ graus de liberdade, sendo s o número de parâmetros do modelo sob H_0 e t o número de parâmetros do modelo sob H_1 .

Para maiores informações ver Paula (1997) e SAS/STAT Volume 2 (1989).

4.2 Técnicas Utilizando o “Rank” das Observações

4.2.1 Análise Discriminante do “Rank”

Já vimos na Subseção 4.1.1 que o método tradicional de análise discriminante requer que as observações sigam uma distribuição normal multivariada. Se as matrizes de covariâncias são iguais, usa-se em geral a regra linear de Fisher; se são diferentes, a regra quadrática.

Se a suposição de normalidade não pode ser assumida, essas regras de classificação

podem não ter bom desempenho. Há na literatura procedimentos que não exigem a normalidade dos dados, entretanto, esses procedimentos são em geral de difícil aplicação, e exigem grandes amostras para que as densidades possam ser estimadas.

Outra alternativa é usar transformações de modo que as variáveis transformadas tenham distribuição aproximadamente normal. Conover e Iman (1981) apresentam a transformação “rank” como uma solução para o caso.

Os procedimentos que utilizam a transformação “rank” baseiam-se em aplicar técnicas paramétricas usuais aos dados “rankeados”. Nesse caso, as análises discriminantes são aplicadas ao “rank” das observações como se esses fossem normais.

Segundo Conover e Iman (1981), esse procedimento é simples e relativamente robusto. Eles apresentam suas conclusões com base em estudos de simulação, comparando seus resultados com outros apresentados na literatura em estudos similares, desenvolvidos por Gessaman e Gessaman (1972), Lachenbruch, Sneeringer e Revo (1973), Johnson e Ramberg (1977) e Koffler e Penfield (1979). Esses estudos consideram distribuições normais com matrizes de covariâncias iguais e diferentes, distribuições não normais e métodos não paramétricos.

Os métodos foram comparados através da Taxa Aparente de Erro (TAE), para duas populações. Denotemos por:

- FDL: função discriminante linear tradicional;
- FDQ: função discriminante quadrática tradicional;
- RFDL: função discriminante linear aplicada ao “rank” das observações;
- RFDQ: função discriminante quadrática aplicada ao “rank” das observações.

As principais conclusões foram:

- quando as populações são normais, os resultados obtidos por FDL ou RFDL são equivalentes e muito pouco se perde ao usarmos RFDQ no lugar de FDQ;
- quando as populações não são normais e os tamanhos das amostras são iguais, em geral, RFDL é melhor do que FDL ou FDQ e não há grande diferença entre usar RFDL ou RFDQ;
- quando as populações não são normais e as amostras têm tamanhos diferentes, o RFDQ tem desempenho pior do que o RFDL, sendo este pior do que o FDQ;
- em termos gerais, o RFDL é preferível a quatro métodos não paramétricos apresentados no estudo.

O argumento usado por Conover e Iman (1981) para a aplicação do “rank” é que, quando as populações não são normais com covariâncias iguais, o procedimento ótimo é aplicar uma transformação para que as densidades se adequem à normal e aplicar a análise discriminante linear tradicional. Muitas vezes essas transformações são monotônicas e tomar o “rank” das variáveis originais ou das variáveis transformadas é o mesmo. Como já foi dito, FDL e RFDL são equivalentes nessa situação. Assim, pode-se esperar que aplicar a função linear discriminante ao “rank” das observações é equivalente ao procedimento ótimo.

Outras referências sobre o uso do “rank” na análise discriminante são Lachenbruch (1975) e Moore e Smith (1975).

O “rank” dos dados $r(\mathbf{x})$ é obtido pela ordenação das observações. Para cada variável é atribuído o valor 1 para o primeiro menor, 2 para o segundo menor, ..., e o valor n para o maior. Em caso de empate, a média dos “ranks” é atribuída.

No caso linear, substituindo o valor observado em (4.4) pelo seu respectivo “rank”,

tem-se a função densidade de probabilidade, dada por

$$f_i(\mathbf{r}(\mathbf{x})) = [(2\pi)^{p/2} |\Sigma_R|^{1/2}]^{-1} \exp[-1/2(\mathbf{r}(\mathbf{x}) - \boldsymbol{\mu}_{iR})' \Sigma_R^{-1} (\mathbf{r}(\mathbf{x}) - \boldsymbol{\mu}_{iR})], \quad (4.11)$$

onde

$$i = 0, 1;$$

$$\mathbf{r}(\mathbf{x}) = [r(x_1), r(x_2), \dots, r(x_p)]';$$

$\boldsymbol{\mu}_{iR}$ é o vetor de médias dos “ranks” da i -ésima população;

Σ_R é a matriz de variância-covariância do “rank” das observações.

Para classificar uma nova observação \mathbf{x}_c em uma das populações, o seu “rank” é calculado por interpolação linear entre os dois “ranks” mais próximos. A regra de classificação, no caso linear, continua sendo a definida em (4.1) e (4.2), ou seja, classificar \mathbf{x}_c em

$$\pi_0 \text{ se } -(\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{0R})' \Sigma_R^{-1} (\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{0R}) \geq -(\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{1R})' \Sigma_R^{-1} (\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{1R}),$$

$$\pi_1 \text{ se } -(\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{0R})' \Sigma_R^{-1} (\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{0R}) < -(\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{1R})' \Sigma_R^{-1} (\mathbf{r}(\mathbf{x}_c) - \boldsymbol{\mu}_{1R}).$$

Neste caso, também obtêm-se amostras aleatórias de tamanhos n_0 e n_1 das duas populações e substitui-se os respectivos parâmetros da função densidade pelas suas estimativas amostrais.

4.2.2 Regressão Logística do “Rank”

Neste caso o vetor das variáveis $\mathbf{x} = [1, x_1, x_2, \dots, x_p]'$ é substituído pelo vetor dos seus respectivos “ranks” $(\mathbf{r}(\mathbf{x}))$.

O modelo logístico fica definido por

$$\log\left\{\frac{\pi(\mathbf{r}(\mathbf{x}))}{1 - \pi(\mathbf{r}(\mathbf{x}))}\right\} = \boldsymbol{\beta}' \mathbf{r}(\mathbf{x}),$$

onde

$$\mathbf{r}(\mathbf{x}) = [1, r(x_1), r(x_2), \dots, r(x_p)]';$$

$$\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \dots, \beta_p];$$

$\pi(\mathbf{r}(\mathbf{x})) = E(Y = 1/\mathbf{r}(\mathbf{x})) = P(Y = 1/\mathbf{r}(\mathbf{x}))$: probabilidade de classificação da observação no Grupo 0 dado o vetor de variáveis “rankeadas” $\mathbf{r}(\mathbf{x})$.

Logo,

$$\pi(\mathbf{r}(\mathbf{x})) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{r}(\mathbf{x})\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{r}(\mathbf{x})\}}.$$

Neste caso, os parâmetros da regressão logística também podem ser obtidos por um processo iterativo de mínimos quadrados ponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{R}(\mathbf{X})'\mathbf{V}_r^m\mathbf{R}(\mathbf{X}))^{-1}\mathbf{R}(\mathbf{X})'\mathbf{V}_r^m\mathbf{z}_r^m,$$

onde

$$m = 0, 1, \dots;$$

$$j = 1, 2, \dots, n;$$

$\pi_{jr} = P(Y_j = 1|\mathbf{r}_j(\mathbf{x}))$: probabilidade de que a j -ésima observação “rankeada” seja classificada no Grupo 0;

$$\mathbf{V}_r = \text{diag}[\pi_{1r}(1 - \pi_{1r}), \pi_{2r}(1 - \pi_{2r}), \dots, \pi_{nr}(1 - \pi_{nr})];$$

$$\mathbf{z}_r = (z_{1r}, z_{2r}, \dots, z_{nr});$$

$$z_{jr} = \eta_{jr} + \frac{(y_j - \pi_{jr})}{[\pi_{jr}(1 - \pi_{jr})]};$$

$$\eta_{jr} = \log\left\{\frac{\pi_{jr}}{1 - \pi_{jr}}\right\}.$$

O mesmo método de validação cruzada deve ser aplicado neste caso para o posterior cálculo da TAE . A metodologia é a mesma para as observações originais só que neste caso é aplicada ao “rank” das observações. Maiores detalhes são apresentados na Seção 4.1.2. Para aprofundamento no assunto consultar Conover e Iman (1981), Hosmer e Lemeshow (1989), McCullagh e Nelder (1989), Paula (1997) e SAS/STAT Volume 2 (1989).

4.2.3 Regressão Probit do “Rank”

O modelo probit é dado por:

$$\Phi^{-1}(\mathbf{r}(\mathbf{x})) = \boldsymbol{\beta}'\mathbf{r}(\mathbf{x}),$$

onde

$\Phi(\mathbf{r}(\mathbf{x}))$: função de distribuição normal acumulada;

$\mathbf{r}(\mathbf{x}) = [1, r(x_1), r(x_2), \dots, r(x_p)]'$;

$\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \dots, \beta_p]$;

$\pi(\mathbf{r}(\mathbf{x})) = E(Y = 1/\mathbf{r}(\mathbf{x})) = P(Y = 1/\mathbf{r}(\mathbf{x}))$: probabilidade de classificação da observação no Grupo 0 dado o vetor $\mathbf{r}(\mathbf{x})$;

$-\infty < \boldsymbol{\beta}'\mathbf{r}(\mathbf{x}) < +\infty$.

Conseqüentemente,

$$\pi(\mathbf{r}(\mathbf{x})) = \Phi(\boldsymbol{\beta}'\mathbf{r}(\mathbf{x}))$$

e as estimativas dos parâmetros são dadas por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}'\mathbf{V}^m\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^m\mathbf{z}^m,$$

onde

$m = 0, 1, \dots$;

$j = 1, 2, \dots, n$;

$\pi_{jr} = P(Y_j = 1 | \mathbf{r}_j(\mathbf{x}))$: probabilidade de que a j -ésima observação seja classificada no Grupo 0;

$\mathbf{V} = \text{diag}\left\{\left\{\frac{\partial \pi_{jr}}{\partial \eta_{jr}}\right\}^2 / \pi_{jr}(1 - \pi_{jr})\right\}$;

$\mathbf{z}_r = (z_{1r}, z_{2r}, \dots, z_{nr})'$;

$z_{jr} = \eta_{jr} + (y_j - \pi_{jr})\frac{\partial \eta_{jr}}{\partial \pi_{jr}}$;

$\eta_{jr} = \Phi^{-1}(\pi_{jr})$.

Como no caso anterior, a metodologia foi descrita na Seção 4.1.3 e maiores informações podem ser encontradas em Conover e Iman (1981), Hosmer e Lemeshow (1989), McCullagh e Nelder (1989), Paula (1997) e SAS/STAT Volume 2 (1989).

Capítulo 5

Aplicação

Um problema semelhante ao tratado neste trabalho foi considerado por Espahbodi (1991), que utilizou a análise discriminante sob a suposição de normalidade e a regressão logística. Aqui, além destas técnicas, optou-se por considerar também a regressão probit com base nos dados originais, como também estas três técnicas aplicadas ao “rank” das observações. Um resumo sobre cada uma dessas técnicas é apresentado no Capítulo 4. Como o objetivo principal é a correta classificação de IF's nos grupos, o critério utilizado para definição do melhor modelo é a menor taxa de erro nas classificações, juntamente com a significância dos parâmetros nos modelos considerados.

Apresenta-se primeiramente a análise das observações originais e em seguida a análise das observações “rankeadas”. O “rank” das observações é obtido ordenando cada variável do menor para o maior valor, sendo o valor 1 atribuído para o primeiro menor valor, 2 para o segundo menor, ..., e n para o maior, onde n é o número de observações válidas. Em caso de empate, a média dos “ranks” é atribuída.

As análises foram feitas no software SAS com a aplicação das PROC's DISCRIM, STEPDISC e LOGISTIC esta última com funções de ligação logit e probit (normit).

5.1 Análise das Variáveis Originais

Esta seção apresenta os resultados da aplicação das técnicas de análise discriminante linear, regressão logística e regressão probit, baseados nos dados originais.

Aplicou-se a análise discriminante quadrática nas variáveis L3, L4, L6, L7, L9 e L14 (seleção automática de variáveis). Apesar de haver indicação de que as matrizes de covariâncias são diferentes, a regra quadrática apresentou pior resultado do que a regra linear. Isso deve ter ocorrido por causa da falta de normalidade.

5.1.1 Análise Discriminante Linear

Pela seleção automática de variáveis para análise discriminante (métodos “forward”, “backward” e “stepwise”) foram selecionadas as variáveis L3, L4, L6, L7, L9 e L14. Os coeficientes da função discriminante ajustada aos dados (Modelo 1) são mostrados na Tabela 5.1.

Tabela 5.1 Coeficientes da função discriminante (Modelo 1)

	Grupo 0	Grupo 1
Constante	-31,37	-26,23
L3	-1,62	0,73
L4	0,07	0,02
L6	6,21	9,49
L7	57,07	40,64
L9	-0,17	0,00
L14	235,39	306,77

A classificação de uma nova unidade é feita segundo a regra dada por (4.1) e (4.2), ou seja, as duas funções discriminantes são aplicadas à nova observação que é classificada na população cujo valor foi o maior.

Hair et al (1995) indicam que a amostra deve ter por volta de 20 unidades por preditor incluído na função discriminante, mas 15 unidades por preditor, como no caso em análise, não seria problemático. Os autores indicam também que o menor grupo deve ter mais do que o número de preditores e que cada grupo tenha pelo menos 20 unidades. Esta última indicação não ocorre no conjunto em análise e além disso, os grupos têm

número de unidades muito diferentes, o que pode ter impacto na função discriminante e na classificação.

A Tabela 5.2 apresenta as taxas de reclassificação e validação cruzada, de acordo com o Modelo 1.

Tabela 5.2 Reclassificação e validação cruzada (Modelo 1)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	76	0	76	0	75	1	76
	100,00	0,00	100,00		98,68	1,32	100,00
1	4	6	10	1	4	6	10
	40,00	60,00	100,00		40,00	60,00	100,00
Total	80	6	86	Total	79	7	86
	93,02	6,98	100,00		91,86	8,14	100,00
	TAE = 4,65%				TAE = 5,81%		

Segundo essa tabela, por exemplo, no caso da validação cruzada, cinco bancos foram classificados de maneira errada. A Tabela 5.3 mostra quais foram os bancos mal classificados.

Tabela 5.3 Classificações erradas (Modelo 1)

Obs	Banco	De - Para
68	América do Sul	0 - 1
81	Banorte	1 - 0
83	Econômico	1 - 0
84	Nacional	1 - 0
90	Bamerindus	1 - 0

Pode-se verificar pela Tabela 5.3 que quatro das cinco IF's mal classificadas são bancos que sofreram intervenção (Grupo 1) mas foram classificados como bancos sem

problemas (Grupo 0). Nesse grupo a taxa de classificação errada foi de 40% e esse é justamente o erro considerado de maior gravidade neste problema.

5.1.2 Regressão Logística

Utilizando-se a seleção automática de variáveis para regressão logística, nenhum modelo satisfatório foi obtido, pois as taxas de erro de classificação do Grupo 1 para o 0 foram sempre superiores a 40%. Contudo, a partir das variáveis escolhidas no modelo anterior, L3, L4, L6, L7, L9 e L14, os resultados melhoraram.

As variáveis escolhidas foram respectivamente L3, L6, L7 e L14 (“forward”), L3, L6 e L14 (“backward”) e L3, L6 e L14 (“stepwise”). A melhor classificação obtida foi para o ajuste com os indicadores L3, L6, L7 e L14, cuja significância dos parâmetros sugere a retirada do intercepto. O coeficiente obtido para o intercepto neste modelo foi 4,91, e o valor obtido para o teste de Wald para testar a sua significância foi 1,15 com “p-value” correspondente a 0,28, o qual nos leva a considerá-lo não significativo.

Considerando o diagnóstico do modelo (item 4.1.4, Metodologia) e a *TAE*, optou-se pela retirada das observações 18, 54, 68 e 90 (Modelo 2). O diagnóstico baseou-se nas medidas abaixo descritas as quais destacaram as seguintes observações:

- Resíduo de Pearson e Desvio Residual: destacam-se as observações 83, 84 e 90 para o primeiro e 83 e 90 para o segundo;
- Matriz “chapéu”: 18, 47, 54, 68, 72, 81 e 89;
- DFBetas: referente a L3: 18 e 72; a L6: 54, 72, 84 e 90; a L7: 18, 54, 72 e 90; a L14: 54 e 90;
- C e Cbar: 18, 54 e 90.

As estimativas do modelo ajustado encontram-se na Tabela 5.4, com apenas o coeficiente do indicador L6 não significativo ao nível de 10%. A sua retirada implica no aumento da *TAE* e por isso foi mantido. O teste de Hosmer e Lemeshow resultou em nível descritivo maior do que 99%, mostrando que o modelo está bem ajustado.

Tabela 5.4 Coeficientes, erros-padrão e testes de Wald (Modelo 2)

Variável	GL	Coeficiente	Erro-padrão	Teste de Wald	p-value
L3	1	- 2,73	1,34	4,13	0,04
L6	1	- 5,49	3,74	2,15	0,14
L7	1	14,23	6,61	4,63	0,03
L14	1	-152,6	89,64	2,89	0,09

Interpretação dos Sinais dos Coeficientes

As variáveis L3 (Imobilização do Patrimônio Líquido) e L6 (Participação dos Depósitos Interfinanceiros no Total de Depósitos) possuem coeficientes negativos no modelo indicando que quanto menor os valores assumidos pelos indicadores, menor a chance da IF ter problemas, coerente com o esperado, ou seja, é melhor ter baixo índice de recursos próprios comprometidos em aplicações sem intenção de realização e de recursos captados no mercado interbancário, os quais são mais caros.

A baixa Imobilização do Patrimônio Líquido implica maior disponibilidade para a IF e conseqüentemente maiores recursos aplicados em ativos operacionais. Considera-se que predominantemente, os recursos não aplicados no permanente são aplicados no ativo operacional, tendo em vista as características do Plano Contábil das Instituições do Sistema Financeiro Nacional - COSIF.

Conforme verifica-se no Manual de Normas e Instruções do Banco Central - MNI de número 02 - 01 - 02 "... o total de recursos aplicados no Ativo Permanente não pode ultrapassar 80% (oitenta por cento) do valor do PLA das instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil, ..." (PLA - Patrimônio Líquido Ajustado), salvo algumas considerações. O limite é estabelecido com a finalidade de impedir a alta imobilização que impossibilite as IF's de cumprirem com as suas obrigações, devido à falta de liquidez.

O coeficiente de L7 (Evolução do Ativo Operacional Médio) significa que o aumento médio das atividades operacionais da instituição diminui a chance da IF ter problemas. Este resultado está coerente com o resultado obtido com a Imobilização do Patrimônio Líquido, tendo em vista que a sua diminuição implica no aumento do ativo operacional.

O aumento de L14 (Taxa de Retorno das Aplicações), condiz com maior probabilidade da IF sofrer liquidação ou intervenção extrajudicial. Na prática, entende-se que, não necessariamente a maior Taxa de Retorno das Aplicações implica na maior chance da IF sofrer liquidação ou intervenção. Isso porque o maior volume de operações significa o crescimento das atividades operacionais da IF e a possibilidade de obtenção de resultados melhores.

Além disso, L14 está correlacionada com L15 (Custo das Captações, o qual não aparece nesse modelo). A correlação amostral entre essas variáveis é de 0,54. Em busca do maior retorno de suas aplicações a IF está sujeita à seguinte situação: de um lado os tomadores pagam altas taxas de juros e os aplicadores exigem alta taxa de remuneração pelos seus depósitos (alto custo de captação). Este fato provoca a elevação do risco da IF, pois os tomadores que aceitam pagar taxas maiores são justamente os de maior risco. Ou seja, os analistas de balanço entendem que o maior risco, maior custo de captação e maior retorno de aplicações “andam juntos”. E esse risco devido ao maior retorno pode ou não se consumir.

Outro aspecto que pode resultar em uma elevada taxa de retorno, acima da média dos demais bancos, é a existência de operações não realizáveis (congeladas). Essas operações anormais (atraso ou liquidação) permanecem apropriando receitas quando não reclassificadas, provocando um aumento indevido das receitas auferidas que são apenas contábeis, não refletindo a realidade do banco.

As *TAE* e correspondentes IF's classificadas erroneamente estão nas Tabelas 5.5 e 5.6.

Tabela 5.5 Reclassificação e validação cruzada (Modelo 2)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	73	0	73	0	70	3	73
	100,00	0,00	100,00		95,90	4,10	100,00
1	2	7	9	1	2	7	9
	22,22	77,77	100,00		22,22	77,77	100,00
Total	75	7	82	Total	72	10	82
	91,46	8,54	100,00		87,80	12,20	100,00
TAE = 2,44%			TAE = 6,09%				

Tabela 5.6 Classificações erradas (Modelo 2)

Obs	Banco	De - Para
15	Marka	0 - 1
30	Sul América	0 - 1
72	Itamarati	0 - 1
83	Econômico	1 - 0
84	Nacional	1 - 0

5.1.3 Regressão Probit

A partir das variáveis selecionadas para a análise discriminante (L3, L4, L6, L7, L9 e L14), aplicou-se a regressão probit selecionando-se automaticamente as variáveis L3, L6, L7 e L14 (“forward”), L3, L6 e L14 (“backward”) e L3, L6 e L14 (“stepwise”), sendo que nos dois últimos casos a validade do modelo é questionável. Assim como no caso em que foi utilizada a função de ligação logit, a melhor classificação obtida foi para o ajuste com os indicadores L3, L6, L7 e L14, cuja significância dos parâmetros também sugere a retirada do intercepto (intercepto = 2,43, teste de Wald = 1,09 e “p-value” = 0,30).

O diagnóstico do modelo destaca as unidades abaixo descritas, optando-se então pela retirada das observações 18 e 54, consideradas influentes (Modelo 3). Nota-se que

as observações mal classificadas (83, 84 e 90) no modelo resultante (Modelo 3), são justamente aquelas indicadas pela análise de resíduos.

Medidas de diagnóstico e observações que se destacaram:

- Resíduo de Pearson e Desvio Residual: 83, 84 e 90 para ambas as medidas;
- Matriz “chapéu”: 18, 54 e 87;
- DFBetas: referente a L3: 18; a L6: 72; a L7: 18; a L14: 54;
- C e Cbar: 18 e 54.

As estimativas do modelo ajustado encontram-se na Tabela 5.7, com apenas o coeficiente do indicador L6 não significativo ao nível de 10%. Como no caso anterior e pelos mesmos motivos, o indicador foi mantido no modelo. O nível descritivo do teste de Hosmer e Lemeshow foi maior do que 99%.

Tabela 5.7 Coeficientes, erros-padrão e testes de Wald (Modelo 3)

Variável	GL	Coeficiente	Erro-padrão	Teste de Wald	p-value
L3	1	-1,08	0,44	6,09	0,01
L6	1	-2,37	1,50	2,50	0,11
L7	1	5,91	2,17	7,45	0,01
L14	1	-59,23	32,52	3,32	0,07

Interpretação dos Sinais dos Coeficientes

O modelo obtido pela aplicação da regressão probit (Modelo 3) é similar ao modelo sugerido através da regressão logística (Modelo 2). Os dois dependem das mesmas variáveis, cujos coeficientes tem os mesmos sinais. Portanto, a interpretação do comportamento das variáveis (aumento ou decréscimo) no risco da IF sofrer liquidação ou intervenção e conseqüentes interpretações não se modificam.

As Tabelas 5.8 e 5.9 mostram, respectivamente, as taxas de reclassificação e as IF's mal classificadas segundo a classificação cruzada, no Modelo 3.

Tabela 5.8 Reclassificação e validação cruzada (Modelo 3)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	73	1	74	0	72	2	74
	98,65	1,35	100,00		97,30	2,70	100,00
1	3	7	10	1	3	7	10
	30,00	70,00	100,00		30,00	70,00	100,00
Total	76	8	84	Total	75	9	84
	90,48	9,52	100,00		89,29	10,71	100,00
			TAE = 4,76 %				TAE = 5,95 %

Tabela 5.9 Classificações erradas (Modelo 3)

Obs	Banco	De - Para
68	América do Sul	0 - 1
72	Itamarati	0 - 1
83	Econômico	1 - 0
84	Nacional	1 - 0
90	Bamerindus	1 - 0

5.2 Análise das Variáveis “Rankeadas”

Esta seção apresenta os resultados da aplicação das técnicas de análise discriminante linear, regressão logística e regressão probit baseadas nos dados transformados. A análise discriminante quadrática apresentou piores resultados do que a linear, com índice de erro nas classificações do Grupo 1 para o Grupo 0 equivalente a 40%, por isso não será apresentada.

5.2.1 Análise Discriminante Linear

Os métodos de seleção automática de variáveis para análise discriminante selecionaram respectivamente os “ranks” das variáveis L6, L7, L10, L14, L15 e L17 (“for-

ward”); L6, L9, L10, L14, L15 e L17 (“backward”) e L3, L6, L10, L14, L15, L17 (“step-wise”). Nota-se que apesar dos três métodos não apresentarem as mesmas variáveis selecionadas como na análise discriminante linear com as variáveis originais, predominam as variáveis L6, L10, L14, L15 e L17.

A menor TAE foi proporcionada pela seleção “backward” (Modelo 4) com $TAE = 4,65\%$ (reclassificação) e $6,97\%$ (validação cruzada). A Tabela 5.10 mostra as funções de classificação desse modelo.

Tabela 5.10 Coeficientes da função discriminante (Modelo 4)

	Grupo 0	Grupo 1
Constante	-7,98	-22,28
RL6	0,10	0,14
RL9	0,07	0,11
RL10	-0,01	-0,16
RL14	0,05	0,16
RL15	0,08	0,18
RL17	0,07	0,15

As Tabelas 5.11 e 5.12 mostram respectivamente as TAE (reclassificação e validação cruzada) e as observações mal classificadas sob o Modelo 4.

Tabela 5.11 Reclassificação e validação cruzada (Modelo 4)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	74	2	76	0	73	3	76
	97,37	2,63	100,00		96,05	3,95	100,00
1	2	8	10	1	3	7	10
	20,00	80,00	100,00		30,00	70,00	100,00
Total	76	10	86	Total	76	10	86
	88,37	11,63	100,00		88,37	11,63	100,00
	TAE = 4,65 %				TAE = 6,97 %		

Tabela 5.12 Classificações erradas (Modelo 4)

Obs	Banco	De - Para
68	América do Sul	0 - 1
69	Lloyds Bank	0 - 1
80	Caixa Econômica Estadual do RS	0 - 1
83	Econômico	1 - 0
84	Nacional	1 - 0
90	Bamerindus	1 - 0

Os Bancos Econômico, Nacional e Bamerindus possuem as seguintes características as quais podem ter sido a causa da má classificação dessas IF's:

- Bancos Econômico (RL9 = 34 e RL10 = 79) e Nacional (RL9 = 59 e RL10 = 83): possuem baixo "rank" de Representatividade de Créditos Anormais em Relação ao Patrimônio Líquido (L9), comparativamente ao Grupo 1 e altos "ranks" de Remuneração do Ativo Operacional Médio (L10). Posteriormente descobriu-se que essa remuneração era apenas contábil, não representando a realidade;

- Banco Bamerindus (RL14 = 34 e RL15 = 36): a Taxa de Retorno das Aplicações (L14) e o Custo das Captações (L15) são os mais baixos entre todas as instituições que sofreram intervenção ou liquidação extrajudicial. De fato, caso a diferença entre as taxas de retorno e de captação (spread) seja baixa, o resultado do banco pode não ser suficiente para cobrir as despesas operacionais.

5.2.2 Regressão Logística

O método de seleção automática de variáveis (“stepwise”) escolheu os “ranks” de L1, L7 e L14. Para as seleções “forward” e “backward”, as estimativas de máxima verossimilhança não existem devido à completa separação entre os grupos (So, 1998).

No diagnóstico do modelo com todas as observações, com o intercepto e as variáveis RL1, RL7 e RL14 salientam-se as observações abaixo, por medidas de diagnóstico:

- Resíduo de Pearson e Desvio Residual: 90;
- Matriz “chapéu”: 8, 15, 19, 30, 40, 72, 73, 77, 81, 82, 83, 84, 85, 86, 87 e 89;
- DFBetas: relativo ao intercepto: 73, 87 e 90; a RL1: 8, 86, 87, 89; a RL7: 73, 84, 87; a RL14: 8, 15, 19, 84 e a 90;
- C e Cbar: 8, 15, 19, 84, 86, 89 e 90.

Retirando-se as observações 8, 15 e 19, e considerando a melhor taxa de classificação obtida, além da significância dos parâmetros (todos são significativos, exceto o intercepto), obteve-se o Modelo 5 da Tabela 5.13. O nível descritivo do correspondente teste de Hosmer e Lemeshow foi maior do que 99%.

Tabela 5.13 Coeficientes, erros-padrão e testes de Wald (Modelo 5)

Variável	GL	Coeficiente	Erro-padrão	Teste de Wald	p-value
Intercepto	1	1,85	2,97	0,39	0,53
RL1	1	0,16	0,07	5,30	0,02
RL7	1	0,15	0,06	5,51	0,09
RL14	1	-0,15	0,06	6,55	0,01

Interpretação dos Sinais dos Coeficientes

Há evidências de que quanto maior o valor das variáveis L1 (Liquidez Imediata) e L7 (Evolução do Ativo Operacional Médio) maior a chance da IF não entrar em intervenção ou liquidação extrajudicial e quanto maior a Taxa de Retorno das Aplicações (L14) maior a chance da IF ter problemas, mantidas as demais condições fixas. No caso do índice L14 o oposto era esperado e isto já havia sido observado na análise descritiva.

Quanto maior o indicador L1, maior a possibilidade da IF ser capaz de saldar as suas obrigações no curto prazo, devido a sua disponibilidade financeira imediata. Maiores valores para L7 indicam o aumento do direcionamento de capital para as atividades operacionais da IF e conseqüente diminuição da Imobilização do Patrimônio Líquido.

As considerações anteriormente descritas para o indicador L14 também são válidas para esse caso.

As taxas de reclassificação e validação cruzada e a relação de IF's mal classificadas segundo o Modelo 5 são apresentadas nas Tabelas 5.14 e 5.15, respectivamente.

Tabela 5.14 Reclassificação e validação cruzada (Modelo 5)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	71	1	72	0	70	2	72
	98,61	1,39	100,00		97,22	2,78	100,00
1	1	9	10	1	2	8	10
	10,00	90,00	100,00		20,00	80,00	100,00
Total	72	10	82	Total	72	10	82
	87,80	12,19	100,00	Total	87,80	12,19	100,00
	TAE = 2,43 %				TAE = 4,88 %		

Tabela 5.15 Classificações erradas (Modelo 5)

Obs	Banco	De - Para
30	Sul América	0 - 1
73	Patente	0 - 1
89	Banespa	1 - 0
90	Bamerindus	1 - 0

As características do Bancos Sul América, Patente, Banespa e Bamerindus que podem ter causado a má classificação dessas IF's são:

- Banco do Estado de São Paulo (RL1 = 59): alto “rank” de Liquidez Imediata (L1) comparativamente às demais IF's com problemas;

- Banco Bamerindus (RL14 = 34): baixo “rank” da Taxa de Retorno das Aplicações (L14) comparativamente às demais IF's com problemas;

- Banco Sul América (RL7 = 9 e RL14 = 78): baixo “rank” da Evolução do Ativo Operacional (L7) e alto “rank” da Taxa de Retorno das Aplicações (L14), comparativamente às demais IF's sem problemas;

- Banco Patente (RL7 = 11 e RL1 = 18): baixo “rank” da Evolução do Ativo Operacional (L7) e baixo “rank” de Liquidez Imediata (L1), comparativamente às demais IF's sem problemas.

Retirando-se o intercepto do modelo, os resultados são praticamente os mesmos, mas com uma observação a mais classificada erroneamente do Grupo 1 para o Grupo 0 (Banco do Estado do Rio de Janeiro) e conseqüente $TAE = 6,10$ (validação cruzada). Por isso, embora o intercepto não seja significantemente diferente de zero (Tabela 5.13), ele foi mantido no modelo.

5.2.3 Regressão Probit

A partir do “rank” dos indicadores, selecionou-se as seguintes variáveis:

- método “backward”: RL2, RL10, RL14 e RL15, contudo o modelo obtido não é confiável;

- método “stepwise”: RL10 e RL14. Neste caso a TAE de 6,74% não foi satisfatória, sendo 60% o erro na classificação de IF's originalmente no Grupo 1 para o Grupo 0 e 0% na classificação de IF's do Grupo 0 para o Grupo 1;

- método “forward”: RL1, RL3, RL4, RL7, RL10 e RL14, $TAE = 8,23$ (30% de 1 para 0 e 5,33% de 0 para 1). Neste caso apenas o coeficiente relativo ao “rank” de L14 é significativo.

Considerando as variáveis selecionadas pelo método “forward”, a significância das variáveis e diagnóstico dos modelos obtidos, conforme apresentado nos modelos anteriores e com base no Capítulo 4, item 4.1.4, optou-se pelo ajuste do modelo retirando-se o “rank” das variáveis RL3 e RL4, o intercepto e as observações 80 e 90 (Modelo 6). O nível descritivo apresentado no teste de Hosmer e Lemeshow foi maior do que 99%. A Tabela 5.16 fornece as estimativas desse modelo.

Tabela 5.16 Coeficientes, erros-padrão e testes de Wald (Modelo 6)

Variável	GL	Coeficiente	Erro-padrão	Teste de Wald	p-value
RL1	1	0,27	0,15	3,39	0,07
RL7	1	0,05	0,04	1,54	0,21
RL10	1	0,17	0,10	2,91	0,09
RL14	1	-0,21	0,11	3,72	0,05

Interpretação dos Sinais dos Coeficientes

O coeficiente do “rank” de L14, mesma conclusão do modelo logístico anterior, indica que quanto maior a Taxa de Retorno das Aplicações maior a chance da IF ter problemas. Os coeficientes do “rank” das demais variáveis (L1: Liquidez Imediata; L7: Evolução do Ativo Operacional Médio; L10: Remuneração do Ativo Operacional Médio) são positivos, significando que quanto maior os seus valores maior a chance das IF’s serem saudáveis. As *TAE* (Tabela 5.17) obtidas também foram melhores do que para a análise discriminante linear.

As Tabelas 5.17 e 5.18 mostram as taxas de reclassificação e validação cruzada, bem como as unidades mal classificadas segundo o Modelo 6.

Tabela 5.17 Reclassificação e validação cruzada (Modelo 6)

De - Para	Grupo 0	Grupo 1	Total	De - Para	Grupo 0	Grupo 1	Total
0	73	1	74	0	72	2	74
	98,65	1,35	100,00		97,30	2,70	100,00
1	1	8	9	1	2	7	9
	11,11	88,89	100,00		22,22	77,77	100,00
Total	74	9	83	Total	74	9	83
	89,16	10,84	100,00		89,16	10,84	100,00
	TAE = 2,40 %				TAE = 4,82 %		

Tabela 5.18 Classificações erradas (Modelo 6)

Obs	Banco	De - Para
8	Pontual	0 - 1
19	General Motors	0 - 1
81	Banorte	1 - 0
83	Econômico	1 - 0

De acordo com os valores dos ranks obtidos para os Grupos 0 e 1 temos as seguintes indicações do porquê das IF's mal classificadas (Tabela 5.18):

- Bancos Banorte (RL7 = 31 e RL10 = 73) e Econômico (RL7 = 33 e RL10 = 79) possuem "ranks" maiores do que as demais IF's do Grupo 1 para os indicadores L7 e L10. A maior Evolução do Ativo Operacional Médio, assim como sua respectiva remuneração, é característica de bancos pertencentes ao Grupo 0, ou seja, em boa situação econômico-financeira;

- Bancos Pontual (RL1 = 12 e RL14 = 76) e General Motors (RL1 = 2 e RL14 = 74) possuem "ranks" baixos para o indicador L1 (Liquidez Imediata), ou seja, baixos ativos realizáveis a curto prazo. O "rank" da Taxa de Retorno das Aplicações (L14) é alto; pelos analistas de mercado este comportamento seria adequado para o Grupo 0.

5.3 Comparações

Na seqüência confrontam-se as *TAE*, os bancos mal classificados e os indicadores selecionados nos modelos apresentados. A Tabela 5.19 mostra as *TAE* dos seis modelos selecionados e também as taxas de classificação erradas do Grupo 1 para o Grupo 0. A Tabela 5.20 apresenta os bancos do Grupo 1 classificados erroneamente no Grupo 0 e as respectivas probabilidades de classificação nos grupos. A Tabela 5.21 apresenta os bancos mal classificados sob cada um dos modelos e a 5.22 os indicadores que os compõem.

As *TAE* obtidas foram menores para os modelos que consideram o "rank" das observações nas regressões logística e probit. Isto verifica-se para as *TAE* gerais, que consideram os erros do Grupo 0 para o Grupo 1 e vice-versa, assim como para as *TAE* do Grupo 1 para o 0.

Entre as regressões (logística e probit) do "rank" das observações, os resultados são muito semelhantes. Pelas Tabelas 5.14 e 5.17 nota-se que o número de observações classificadas erroneamente do Grupo 1 para o 0 e do Grupo 0 para o 1 são as mesmas

(contudo, as observações mal classificadas pela validação cruzada são diferentes nos dois modelos (5 e 6)). As taxas de erro são diferentes porque em um modelo temos o ajuste baseado em 82 IF's (Modelo 5) e no outro em 83 IF's. Portanto, em termos de resultados não há motivos para preferir um a outro. Mesmo assim, a “popularidade” da regressão logística facilita o entendimento do modelo aplicado para pessoas que não sejam da área estatística.

Tabela 5.19 Taxas aparentes de erro

Modelo	TAE			
	Reclassificação		Validação cruzada	
	Grupo 1 para o 0	Geral	Grupo 1 para o 0	Geral
Discriminante linear	40,00	4,65	40,00	5,81
Regressão logística	22,22	2,44	22,22	6,09
Regressão probit	30,00	4,76	30,00	5,95
“Rank”				
Discriminante linear	20,00	4,65	30,00	6,97
Regressão logística	10,00	2,43	20,00	4,88
Regressão probit	11,11	2,40	22,22	4,82

A análise discriminante apresentou menor *TAE* geral para as observações originais. Apesar disso, a sua taxa de erro do Grupo 1 para o 0 foi das mais altas (40%) entre todos os modelos, tanto para a reclassificação quanto para a validação cruzada.

Para a análise das variáveis transformadas a análise discriminante apresentou os piores resultados. Possivelmente esses resultados não foram os melhores devido ao problema de não normalidade dos dados.

Todos os modelos apresentam altas taxas de classificação do Grupo 1 para o Grupo 0 (maior do que 20% na validação cruzada) e esse é justamente o tipo de erro mais grave. No entanto, os bancos mal classificados nesse caso (bancos com numeração acima de 80) são quase sempre os mesmos como se pode ver na Tabela 5.21. São eles os Bancos

Econômico, Nacional e Bamerindus e, como já foi comentado anteriormente, os dois primeiros apresentaram indicadores contábeis que não traduziam a realidade, tornando muito difícil a detecção de problemas com base nesses indicadores. O Bamerindus fazia parte de um conglomerado, no qual um de seus integrantes teve problemas, que refletiram no banco, mas não nos seus indicadores econômico-financeiros. Conseqüentemente, isso ocasionou a classificação do banco no Grupo 0 (IF's sem problemas).

Na Tabela C1 do Apêndice C temos as probabilidades de classificar as IF's no Grupo 0 (sem problemas) para os modelos ajustados (Modelos 1 até 6).

Mesmo existindo um modelo o qual proporciona a melhor *TAE*, não se deve descartar a informação de outras IF's que aparecem classificadas erroneamente em mais de um modelo.

Tabela 5.20 Probabilidades de classificação no Grupo 1

Banco	Modelo					
	1	2	3	4	5	6
Banorte	0,569	-	-	-	-	0,501
Econômico	0,983	0,855	0,819	0,751	-	0,711
Nacional	0,984	0,806	0,772	0,719	-	-
Banespa	-	-	-	-	0,5152	-
Bamerindus	0,951	-	0,858	0,778	0,916	-

Na Tabela 5.20 apresentam-se as probabilidades de classificação no Grupo 0 dos bancos do Grupo 1 que foram classificados erroneamente nos modelos. As probabilidades dos Modelos 4, Banco Nacional; Modelo 5, Banespa e Modelo 6, Banorte, referem-se ao modelo ajustado sem essas observações, ou seja, temos as probabilidades pelo método de validação cruzada porque no modelo normal essas observações não foram mal classificadas.

Tabela 5.21 Bancos classificados erroneamente

Modelo	8	15	19	30	68	69	72	80	81	83	84	89	90
Discriminante linear					x				x	x	x		x
Regressão logística		x		x			x			x	x		
Regressão probit					x		x			x	x		x
“Rank”													
Discriminante linear					x	x		x		x	x		x
Regressão logística				x				x				x	x
Regressão probit	x		x						x	x			

8 = Banco Pontual, 15 = Banco Marka, 19 = Banco General Motors, 30 = Banco Sul América, 68 = Banco América do Sul, 69 = Lloyds Bank PLC., 72 = Banco Itamarati, 80 = Caixa Econômica Estadual do RS, 81 = Banco Banorte, 83 = Banco Econômico, 84 = Banco Nacional, 89 = Banespa e 90 = Bamerindus

É essencial salientar que as IF’s são alocadas nos grupos de acordo com probabilidades de classificação, e determinado ponto de corte, neste caso 50%. Ou seja, probabilidades acima de 50% implicam em destinar a observação ao Grupo 0 e probabilidades menores do que 50% ao Grupo 1. Pode acontecer que a IF foi erroneamente classificada, contudo sua probabilidade de classificação está bem próxima do ponto de corte, portanto, merecendo atenção. Sendo assim, é importante para os usuários de modelos de classificação para uso prático em fiscalização, ordenar as IF’s pelas respectivas probabilidades a elas atribuídas, tendo em vista priorizar atitudes de controle nos casos, por ordem, mais críticos.

Pela Tabela 5.20, percebe-se que apenas os Banco Banorte e o Banespa possuem as suas probabilidades de classificação no Grupo 0 próximas do limite de corte (0,50) para classificação nos Grupos 0 ou 1 e indicam a necessidade de maior atenção do Banco Central. Os demais são justamente aqueles bancos que apresentavam indicadores contábeis que não traduziam a realidade, conforme mencionado anteriormente.

Os indicadores utilizados na tese possuem certa estabilidade, ou seja, se desde agosto de 1998 até agora o banco não apresentou problemas, não há motivos para supor que a IF pudesse estar em má situação, considerando L1 até L17. Contudo, o Banco Marka e o Banco Pontual mereceriam atenção nesse sentido pois, como é sabido, o Banco Marka “quebrou” e o Banco Pontual encontra-se sob intervenção do Banco Central, sugerindo talvez que já estivessem com problemas no período de análise no qual este trabalho baseia-se.

As causas da “quebra” do Banco Marka têm sido amplamente divulgadas pela imprensa. Desconsiderando a crença do mercado de que a moeda brasileira sofreria desvalorização cambial, os responsáveis pelo banco apostaram na Bolsa de Mercadorias & Futuros - BM&F - diferentemente dos demais investidores. O banco tinha 12.650 contratos que representavam US\$ 1,2 bilhões (Revista Isto é, 21 de abril de 1999). Com a desvalorização da moeda não foram capazes de saldar seus contratos da BM&F.

O caso acima descrito caracteriza-se por mudanças rápidas no mercado. Quando ocorre a valorização ou desvalorização da moeda, lucros ou prejuízos podem concretizar-se rapidamente, situação esta que não é identificada pelos indicadores em estudo.

E de que adiantaria a análise de dados de demonstrações contábeis de instituições financeiras se pode acontecer dela tornar-se insolvente e o fato não influenciar os indicadores a tempo de identificar o problema? O fato dos indicadores utilizados não refletirem posições tal como os comprometimentos do banco na BM&F não exclui a hipótese de que podem identificar outras causas, pois o oposto também poderia ocorrer, ou seja, dados de balancete indicando problemas na IF os quais talvez não fossem caracterizados de outra forma. Além disso, este tipo de análise não é a única base de informação disponível para analistas de mercado.

Ao contrário do Banco Marka, os problemas do Banco Pontual já mostravam indícios na presença do “rank” do indicador L1 (Liquidez Imediata), cujo valor é 12. Tais problemas foram realmente detectados posteriormente.

No caso do Banco Nacional os dados disponibilizados ao BACEN não refletem de maneira alguma a sua real situação econômico-financeira, pois foram fraudados. Obviamente, as alterações aproximam os indicadores dos valores que qualificam IF's no Grupo 0 (IF's sem problemas).

Na Tabela 5.22 apresentam-se os indicadores que fazem parte dos modelos aplicados.

Tabela 5.22 Indicadores que compõem os Modelos 1 a 6

Modelo	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Discriminante linear			x	x		x	x		x								x
Regressão logística			x			x	x										x
Regressão probit			x			x	x										x
“Rank”																	
Discriminante linear						x			x	x				x	x		x
Regressão logística	x						x							x			
Regressão probit	x						x		x					x			

Os indicadores, ou seus “ranks”, abaixo listados não foram considerados importantes na presença das demais variáveis em caso algum:

- L2: Liquidez Corrente até 90 dias;
- L5: Participação das Exigibilidades no Ativo Operacional;
- L8: Participação dos Créditos Anormais no Total das Operações de Créditos Normais;
- L11: Margem Líquida;
- L12: Evolução do Patrimônio Líquido;
- L13: Rentabilidade do Patrimônio Líquido Médio;

- L16: Participação das Rendas de Prestação de Serviços em Relação às Despesas Administrativas.

Vale ressaltar que os testes de normalidade efetuados no Capítulo 3, para o “rank” dos indicadores, rejeitou a normalidade para RL16 no Grupo 0 e RL1, RL12 e RL13 no Grupo 1 e destas apenas RL1 apareceu nos modelos.

Tendo em vista que a regressão logística e a regressão probit (Modelos 2 e 3) foram aplicadas nas variáveis selecionadas para a análise discriminante (Modelo 1), obviamente baseiam-se em subgrupos dessas variáveis. Utilizando o “rank” na regressão logística e regressão probit (Modelos 5 e 6) predominam as mesmas variáveis escolhidas, diferentemente da análise discriminante (Modelo 4).

5.4 Conclusão

Em suma, pode-se concluir:

- considerando a *TAE*, os resultados indicam superioridade das regressões logística e probit aplicada ao “rank” das observações para ajustar os dados em questão;
- predominantemente as mesmas observações do Grupo 1 são mal classificadas nos diferentes modelos (Econômico, Nacional e Bamerindus) ou aparecem como influentes nas regressões logísticas (Bamerindus);
- a interpretação dos coeficientes das regressões estão de acordo com o significado esperado pelos analistas de balanço, exceto para a variável L14 (Taxa de Retorno das Aplicações), cujo coeficiente sinaliza o contrário;
- a medida resumo utilizada é eficiente no sentido que proporciona *TAE* satisfatórias, superiores às obtidas em vários outros estudos, e viabiliza a comparação das diversas IF's, pois algumas possuem apenas poucas observações em suas séries e outras até 48 observações no tempo.

Capítulo 6

Discussões Teóricas

As análises discriminante linear e quadrática têm sido amplamente utilizadas com o objetivo de obter um modelo estatístico baseado em observações prévias de distintos grupos e com base neste modelo classificar novas observações nesses grupos, objetivo deste trabalho. Contudo algumas suposições da técnica não têm sido observadas, prejudicando os resultados obtidos.

Para uma análise discriminante paramétrica, por exemplo, temos os seguintes pré-requisitos (suposições):

- discriminante linear: normalidade dos dados e igualdade das matrizes de variância-covariância;
- discriminante quadrática: normalidade dos dados;
- discriminante de Fisher: igualdade das matrizes de variância-covariância.

Embora a técnica de Fisher tenha sido desenvolvida sem a suposição de normalidade, o seu critério de alocação coincide com o critério da análise discriminante linear.

Estudos anteriores com colocações a respeito da melhor técnica a ser utilizada para análise dos dados deste trabalho, suas vantagens e desvantagens são descritos a seguir. Apresentam-se as conclusões identificando-as pelo seu autor.

6.1 Eisenbeis, 1977

Na prática, considerando dados econômicos, a suposição de normalidade é raramente obtida. Eisenbeis (1977) descreve alguns problemas da aplicação da técnica de análise discriminante a dados econômicos e financeiros.

A violação da suposição de normalidade, segundo Eisenbeis (1977), pode viesar os testes de significância e taxas de erro estimadas e que a maioria dos testes disponíveis são para normalidade univariada e não multivariada. Salienta também a dificuldade em derivar a função densidade de probabilidade conjunta dos dados em caso de não normalidade.

Lembra que as distribuições marginais de uma normal multivariada são normais, contudo a normalidade marginal não necessariamente garante a normalidade conjunta.

Para dados contínuos, sugere uma transformação dos dados, se possível, para aproximar à normalidade, e então aplicar testes para determinar se a técnica linear ou quadrática deve ser usada. No caso de uma distribuição assimétrica à direita o efeito da transformação log natural é tornar mais simétrica a distribuição dos dados.

De Eisenbeis, a desigualdade da matriz de dispersão pode afetar os testes de igualdade da média dos grupos, assim como a regra de classificação nos grupos, quando em vez de usar a forma quadrática a linear é usada.

Sendo assim, sugere antes de efetuar o teste de igualdade da média dos grupos e estimar a taxa de erro na classificação, verificar a igualdade das matrizes de dispersão.

Além disso, cita estudos cujos resultados da análise linear e quadrática diferem a medida que diferem as matrizes de dispersão dos grupos e o número de variáveis aumenta.

Caso uma variável “dummy” ou dicotômica seja incluída na análise, a hipótese de mesma dispersão irá muito provavelmente ser rejeitada, indicando o uso da discriminante quadrática. Motivo: considere uma variável Bernoulli X que assume os valores 0 e 1 ($P(X$

$= 1) = p$ e $P(X = 0) = 1 - p = q$). A média de X no grupo 1 é $\mu_{x,1} = p_1$ e a variância $\sigma_{x,1}^2 = p_1(1 - p_1) = p_1q_1$. Para a análise discriminante linear ser adequada é necessário que $\sigma_{x,1}^2 = \sigma_{x,2}^2$, o que ocorre apenas se $p_1 = p_2$, ou se $p_1 = q_2$. No primeiro caso não interessa aplicar a análise discriminante, X não é um bom discriminador, e o segundo caso é muito específico. Então, se p_1 é diferente de p_2 , ou se p_1 é diferente de q_2 implica que $\Sigma_1 \neq \Sigma_2$.

Reclassificar a amostra original e a partir daí calcular a *TAE* conduz a uma taxa viesada (previsão otimista) do desempenho das regras de classificação na população. Calcular a *TAE* pelo método de Lachenbruch (validação cruzada) é uma alternativa.

6.2 Espahbodi, 1991

Espahbodi, também preocupado com a falta de normalidade dos dados e a desigualdade da matriz de dispersão entre os grupos na aplicação da análise discriminante para distinguir bancos falidos e não-falidos, comparou em seu trabalho as técnicas discriminante e regressão logística.

Percebendo esses problemas, recentes estudos, além de Espahbodi, aplicaram técnicas baseadas em estimadores de máxima verossimilhança, tais como modelos de regressão logit e probit. Os estimadores de máxima verossimilhança são vantajosos devido a suas propriedades estatísticas: consistentes, assintoticamente eficientes e possuem distribuição assintótica conhecida. Além disso, a significância dos coeficientes do modelo pode ser testada e estatísticas para verificar a adequabilidade do ajuste também estão disponíveis.

O autor salienta a importância de não se analisar apenas a probabilidade de falência dos bancos (definindo dois distintos grupos), mas a ordem em que as IF's estariam sujeitas à falência, pois, o viés causado pela amostra que serviu para a construção do modelo não altera a ordem dos bancos em termos das suas probabilidades de falência (Palepu, 1986).

A amostra final do estudo de Espahbodi foi baseada em IF's que faliram em 1983

e outras que não faliram. Coletou os dados em 1981 e 1982 com a finalidade de prever a falência com um e dois anos de antecedência aplicando-se as técnicas de análise discriminante e regressão logística.

A regressão logística apresentou taxa geral de acerto de 87,67% e 75,71% para os modelos com 1 e 2 anos de antecedência e a análise discriminante 86,30% e 84,28%, sugerindo que o modelo logístico classifica melhor com 1 ano de antecedência e o modelo discriminante com 2 anos.

Tendo em vista que a utilização de modelos para reclassificar a própria amostra com que foram desenvolvidos acarreta taxa de classificação viesada, os modelos necessitam ser validados. O trabalho validou os modelos classificando amostras não utilizadas na sua construção. Na validação dos modelos para um ano antes da falência foi obtida taxa de acerto de 83% para o modelo logístico e 79% para a análise discriminante, favorecendo o desempenho do primeiro para previsões. Na análise discriminante linear a piora nas classificações, de acordo com Espahbodi, poderia ser admitida devido a violação da suposição de normalidade dos dados a qual teria subestimado a taxa de erro.

Pela superioridade técnica da regressão logística, neste caso melhor adequabilidade da técnica aos dados, e desempenho tão bom quanto a análise discriminante, o autor indica que essa análise poderia ser considerada mais apropriada.

6.3 Pregibon, 1981

Uma das vantagens do uso da regressão logística para classificar observações nos grupos é a possibilidade de diagnóstico se o ajuste é adequado ou não aos dados. As técnicas de diagnóstico de regressão são descritas pelo autor.

6.4 Perry e Cronan, 1986

Perry e Cronan (1986) utilizam a alternativa proposta por Conover e Iman (1980) para alcançar bons resultados considerando a não normalidade dos dados, obtendo o

“rank” das observações para cada variável e então aplicando a análise discriminante.

Apesar do ordenamento das observações ignorar a magnitude das variáveis e a magnitude da diferença entre elas, as observações discrepantes não precisam ser descartadas neste tipo de análise.

Comparando os resultados obtidos por Peavy e Edgar (1983), com as variáveis originais, e os obtidos por Perry e Cronan, variáveis “rankeadas”, observou-se melhor desempenho (taxa de acerto) no segundo caso: 96,4% e 94% (reclassificação e validação cruzada) contra 91,6% e 88% no primeiro caso.

6.5 Conover e Iman, 1980

Os autores apresentam a análise discriminante baseada no “rank” das observações como opção para o caso de não normalidade dos dados e apresentam simulações que comprovam a robustez da técnica para alguns casos.

Alternativas para estimar a função de densidade necessária à análise discriminante são alguns dos procedimentos não paramétricos. Contudo esses métodos requerem amostras grandes de cada população, porque não só alguns parâmetros serão estimados, mas também a função de densidade completa.

Outra opção é aplicar a transformação do “rank” das observações a cada variável e então aplicar a análise discriminante como se as populações fossem normais. Resultados de simulações indicam que se os dados são normais o desempenho é tão bom quanto a análise discriminante tradicional, caso contrário, o método é tão bom quanto os demais métodos não paramétricos apresentados no seu trabalho.

Nos estudos envolvendo distribuições normais e mesmo tamanho de amostras a discriminante linear do “rank” é quase tão efetiva quanto à discriminante linear. A média de erros da segunda é aproximadamente 95% da média de erros da primeira.

Quando as populações normais possuem diferentes matrizes de dispersão, a disci-

minante quadrática é o melhor método estudado, possuindo em torno de 90% dos erros da discriminante quadrática dos “ranks”. Quando os grupos não possuem distribuição normal as análises baseadas nos “ranks” são de fato melhores do que as que se baseiam nas observações originais.

6.6 Conover e Iman, 1981

No artigo “Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics” os autores apresentam alguns testes paramétricos com as respectivas alternativas não paramétricas baseadas nos “ranks” das observações. Definem os procedimentos baseados na transformação “rank” como aqueles nos quais as técnicas paramétricas usuais são aplicadas aos “ranks”.

Destacam como solução para o caso de não normalidade dos dados, quando esta é necessária, a sua transformação ou a utilização de procedimentos livres desta suposição. Este segundo incluiu os métodos baseados no “rank” dos dados.

Como exemplo de comparação entre os testes paramétricos e não paramétricos (outros estão disponíveis) apresenta-se o exemplo para testar a hipótese $E(X) = E(Y)$ no caso de duas amostras independentes X_1, \dots, X_n e Y_1, \dots, Y_m . O teste paramétrico adequado é comparar a estatística t abaixo com os quantis da distribuição t com $N - 2$ graus de liberdade

$$t = \frac{\bar{X} - \bar{Y}}{[(\Sigma(X_i - \bar{X})^2 + \Sigma(Y_i - \bar{Y})^2)(N/(nm(N-2)))]^{1/2}},$$

onde $N = n + m$, \bar{X} e \bar{Y} são as médias amostrais.

O teste não paramétrico análogo (Wilcoxon-Mann-Whitney) é baseado em

$$T = \frac{S - n(N+1)/2}{[\frac{nm}{N(N-1)} \Sigma R_i^2 - \frac{nm(N+1)^2}{4(N-1)}]},$$

onde R_i é o “rank” da i -ésima observação na amostra correspondente à variável X e S é

a soma desses valores na amostra. A estatística deve ser comparada com a distribuição normal padrão e, em caso de empates e tamanho da amostra menor do que 20, tabelas exatas.

Com o teste t baseado no “rank” (RTP) obtém-se a seguinte estatística

$$t_R = \left[\frac{S}{n} - \frac{1}{m} \left(\frac{N(N+1)}{2} - S \right) \right] / \left[\left(\sum R_i^2 - \frac{1}{n} S^2 - \frac{1}{m} \left(\frac{N(N+1)}{2} - S \right)^2 \right) \frac{N}{nm(N-2)} \right]^{1/2},$$

a qual deve ser comparada com a distribuição t com $N - 2$ graus de liberdade.

A relação entre T e t_R é dada por

$$t_R = \frac{T}{\left[\frac{N-1}{N-2} - \frac{T^2}{n-2} \right]^{1/2}}.$$

Como t_R é uma função de T monotonicamente crescente, os dois testes são equivalentes quando os valores críticos exatos são usados.

As vantagens da RTP são:

- menor rigor nas suposições;
- facilidade de sua aplicação;
- comparações com técnicas paramétricas difundidas e disponíveis em softwares estatísticos.

6.7 Press e Wilson, 1978

No artigo, os autores concluem que no caso de normalidade dos dados a análise discriminante é mais adequada do que a regressão logística para discriminar observações nos grupos. No caso de não normalidade, o contrário seria válido.

Uma vantagem da modelagem logística é que a mesma também serve de alternativa para análise de tabelas de contingência.

Outra questão entre as duas técnicas é a estimação de seus parâmetros. A eficiência

dos estimadores é geralmente importante. Salienta que Efron (1975) mostrou que os estimadores da regressão logística são $1/2$ e $2/3$ tão eficientes quanto os estimadores da função discriminante quando os dados são normais multivariados com iguais matrizes de dispersão, prevalecendo neste caso o uso da análise discriminante.

Em caso de não normalidade os estimadores da análise discriminante não são consistentes, ou seja, mesmo em grandes amostras não há garantias de que bons ajustes ou previsões serão obtidos por esse método. Uma solução prática seria usar estimadores consistentes como os de máxima verossimilhança.

Quando a suposição de normalidade é violada variáveis que não são importantes para o modelo podem ser incluídas tendo em vista a significância dos coeficientes estimados. Por exemplo, um coeficiente que é zero pode ser considerado significativo na análise discriminante.

Além disso, Halperin, Blackwelder e Verter (1971) concluíram que sob não normalidade os ajustes proporcionados pelos estimadores de máxima verossimilhança são melhores considerando a classificação das observações, justificada inclusive pela teoria.

A regressão logística também é aconselhada por proporcionar “sinais” em alguns problemas no ajuste. No caso da completa separação entre os grupos, por exemplo, os estimadores de máxima verossimilhança não vão existir porque não existe uma única solução. Este fato não aparece na discriminante onde os estimadores são calculados normalmente.

Os estimadores de máxima verossimilhança são funções de estatísticas suficientes, enquanto os estimadores da função discriminante não são. E pelo Teorema de Rao-Blackwell o menor erro médio quadrático é sempre obtido por esses estimadores.

Nas aplicações práticas apresentadas pelo autor a regressão logística apresenta melhores resultados com base no número de observações classificadas corretamente.

6.8 So, 1998

O autor explora um dos problemas encontrados ao ajustar a regressão logística: parâmetros infinitos. Este caso refere-se à situação em que a equação de verossimilhança não tem solução finita, ou seja, os estimadores de máxima verossimilhança não existem, os quais dependem dos dados amostrais.

As estimativas não existirão em 2 casos:

- separação completa;
- separação quase completa.

Completa separação dos dados corresponde ao caso em que existe um vetor \mathbf{b} que aloca corretamente todas as observações ao seu respectivo grupo resposta, tal que se

$$\mathbf{b}' \mathbf{x}_i > 0, y_i = 0;$$

$$\mathbf{b}' \mathbf{x}_i < 0, y_i = 1.$$

Neste caso as estimativas de máxima verossimilhança não existem e o logaritmo da verossimilhança tende a 0 com o aumento do número de iterações.

Separção quase completa ocorre quando existe um vetor \mathbf{b} , tal que

$$\mathbf{b}' \mathbf{x}_i \geq 0, y_i = 0;$$

$$\mathbf{b}' \mathbf{x}_i \leq 0, y_i = 1,$$

para pelo menos uma das observações. A estimativa de máxima verossimilhança também não existe e a matriz de dispersão torna-se ilimitada.

Quando nem a completa separação dos pontos, nem a quase completa ocorre, temos a terceira situação: sobreposição. Neste caso a estimativa de máxima verossimilhança existe e é única.

Bibliografia

- Assaf Neto, A. (1993). *Estrutura e Análise de Balanços: um enfoque econômico-financeiro*. 3a ed. São Paulo, Atlas.
- Banco Central do Brasil. *Resolução 1524/88*. 21 de setembro de 1988.
- Brasil, Congresso Nacional. *Decreto-Lei 7661/45*. Diário Oficial, 21 de junho de 1945.
- Brasil, Congresso Nacional. *Lei 4595/64*. Diário Oficial, 31 de dezembro de 1964.
- Brasil, Congresso Nacional. *Lei 6024/74*. Diário Oficial, 13 de março de 1974.
- Bussab, W. O. e Morettin, P. A. (1987). *Estatística Básica*, 4. ed. Atual Editora
- Capelletto, L. R. (1995). *Contribuição à Avaliação do Desempenho Operacional de Bancos, Antes e Depois da Implantação do "Plano Real"*. Dissertação de mestrado. Departamento de Administração, Universidade de Brasília.
- Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, **3**, 22-38.
- Cochran, W. G. (1964). On the performance of the linear discriminant function. *Technometrics*, **6**, 179-190.
- Conover, W. J. (1971). *Practical Nonparametric Statistics*, John Wiley and Sons: New York.
- Conover, W. J. and Iman, R. J. (1980). The rank transformation as a method of discrimination with some examples. *Communications in Statistics - Theory and*

Methods, **9**, 465-487.

Conover, W. J. and Iman, R. J. (1981). Rank transformation as a bridge between parametric and nonparametric statistics. *The American Statistician*, **35**, 465-487.

D'Augustino, R. B. and Stevens, M. A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

Domenech, C. H. (1989). *Métodos Exploratórios e Uso de Medidas Resumo para Análise de Dados Longitudinais*. Dissertação de mestrado. Instituto de Matemática e Estatística, Universidade de São Paulo.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, **70**, 892-898.

Eisenbeis, R. A. (1977). Pitfalls in the application of discriminant analysis in business, finance and economics. *Journal of Finance*, **32**, 875-900.

Espahbodi, P. (1991). Identification of problem banks and binary choice models. *Journal of Banking and Finance*, **15**, 53-71.

Gessaman, M. P. and Gessaman, P. H. (1972). A comparison of some multivariate discrimination procedures. *Journal of the American Statistical Association*, **67**, 468-472.

Hair Jr., J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. (1995). *Multivariate Data Analysis*, 4 Ed. Prentice Hall: New Jersey.

Halperin, W., Blackwelder, C. and Verter, J. (1971). Estimation of the multivariate logistic risk function: a comparison of the discriminant function and maximum likelihood approaches. *Journal of Chronic Disease*, **24**, 125-158.

Hosmer, Jr. D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*, John Wiley

& Sons: New York.

- Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, Prentice Hall: New Jersey.
- Johnson, M. E. and Ramberg, J. S. (1977). Transformations of the Multivariate Normal Distribution with Applications to Simulation. *Technical Report LA-UR-77-2595*, Los Alamos Scientific Laboratory, New Mexico.
- Koffler, S. L. and Penfield, D. A. (1979). Nonparametric discrimination procedures for nonnormal distributions. *Journal of Statistical Computation and Simulation*, **8**, 281-299.
- Koyama, S. M. (1997). *Análise Discriminante para Séries Temporais*. Dissertação de mestrado. Instituto de Matemática e Estatística, Universidade de São Paulo.
- Lachenbruch, P. A. (1975). A problem in discrimination using ranks. *Paper presented at the 8th Annual Symposium on Interface of Computer Science and Statistics, at U. C. L. A.*
- Lachenbruch, P. A. and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, **10**, 1-11.
- Lachenbruch, P. A., Sneeringer, C. and Revo, L. T. (1973). Robustness of the linear and quadratic discriminant function to certain types of normality. *Communications in Statistics - Theory and Methods*, **1**, 39-56.
- Lindgren, C., Garcia, G. and Saal, M. I. (1996). *Bank Soundness and Macroeconomic Policy*, International Monetary Fund.
- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press: London.
- McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*, Chapman and Hall:

New York.

- Moore, K. K. and Smith, W. B. (1975). A rank order approach to discriminant analysis. *Proceedings of the Business and Economics Statistical Section of the American Statistical Association*, 451-455
- Palepu, K. G., (1986). Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics*, 288-301.
- Paula, G. A. (1997). *Modelos de Regressão*, Instituto de Matemática e Estatística - Universidade de São Paulo.
- Peavy, J. W. and Edgar, M. D. (1983). A multiple discriminante analysis of BHC commercial paper ratings. *Journal of Banking and Finance*, 7, 160-173.
- Perry, L. G. and Cronan, T. P. (1986). A note on rank transformation discriminant analysis. *Journal of Banking and Finance*, 10, 605-610.
- Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9, 705-724.
- Press, J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73, 699-705.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2 Ed. John Wiley and Sons: New York.
- SAS Institute Inc., SAS/Language: Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc., 1990.
- SAS Institute Inc., SAS/Guide to Macro Processing, Version 6, Second Edition, Cary, NC: SAS Institute Inc. 1990.
- SAS Institute Inc., SAS/Procedures Guide, Release V.03 Edition. Cary, NC: SAS Institute Inc., 1988.

- SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc., 1989.
- SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989.
- SAS Institute Inc., SAS/System for Regression, Second Edition, Cary, NC: SAS Institute Inc., 1991.
- Shumway, R. H. (1988). *Applied Statistical Time Series Analysis*, Prentice hall: New Jersey.
- So, Y. (1998). TS-450: A tutorial on logistic regression, *SUGII8 Proceedings*.
- Souza, G. S. (1998). *Introdução aos Modelos de Regressão Linear e Não Linear*, Embrapa - Ministério da Agricultura.

Apêndice A

“Boxplots”

Gráfico A1 “Boxplot” de L1 (exclui a obs. 41)

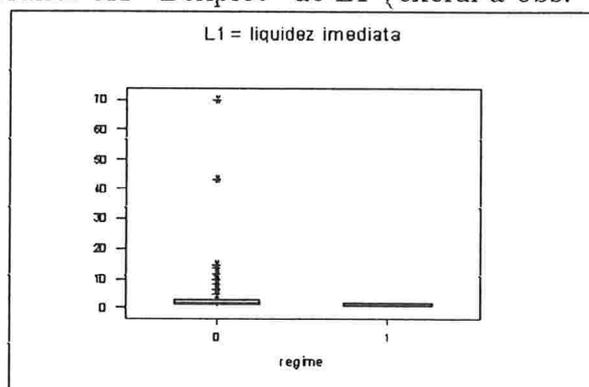


Gráfico A2 “Boxplot” de L2 (exclui a obs. 42)

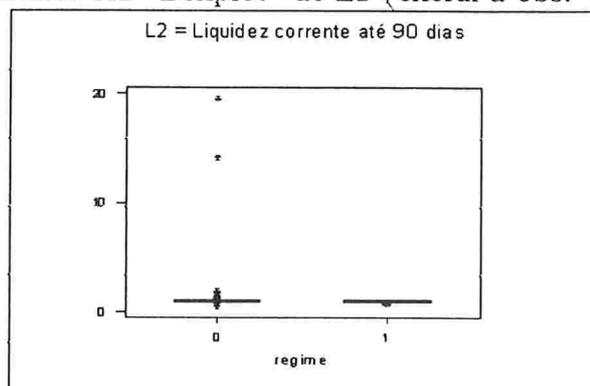


Gráfico A3 "Boxplot" de L3

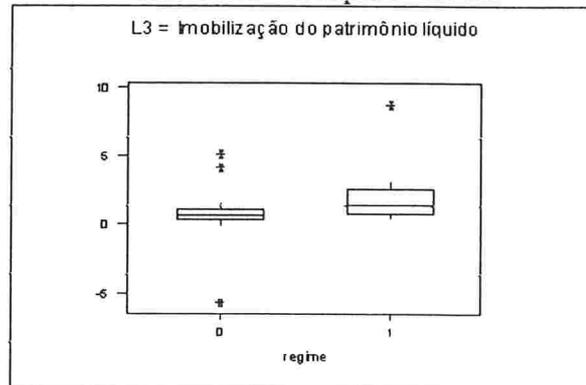


Gráfico A4 "Boxplot" de L4

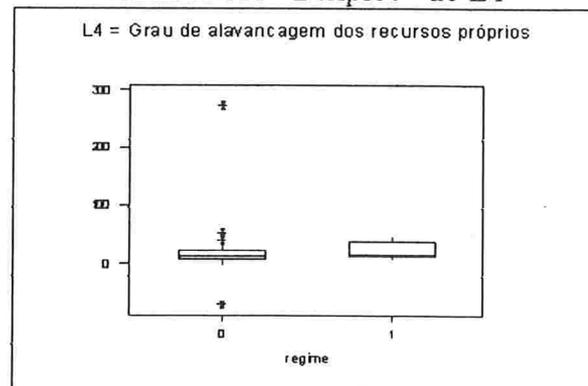


Gráfico A5 "Boxplot" de L5

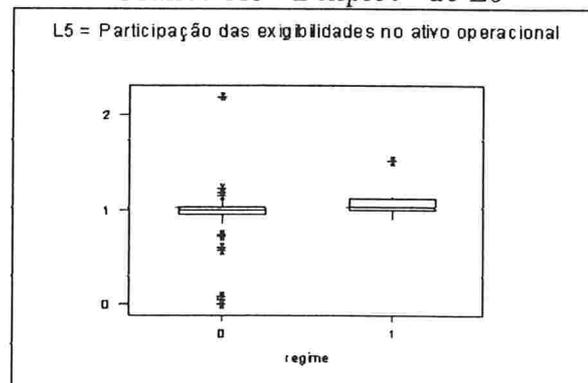


Gráfico A6 "Boxplot" de L6

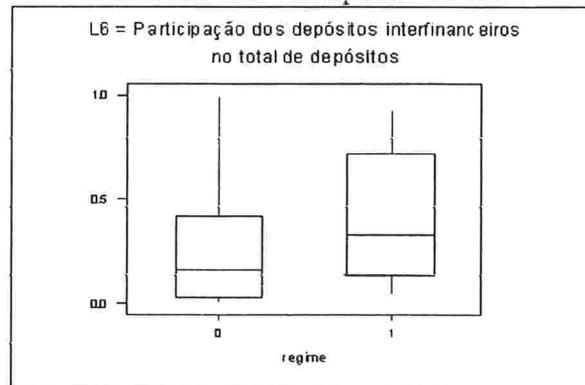


Gráfico A7 "Boxplot" de L7

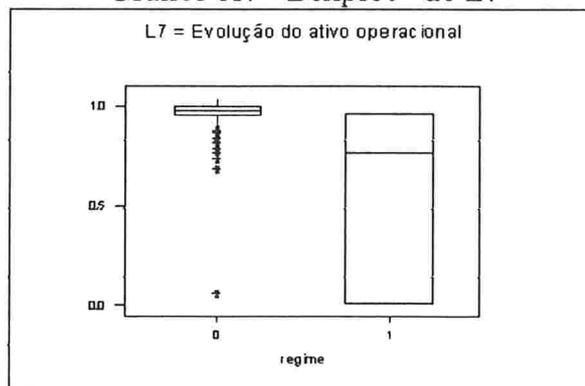


Gráfico A8 "Boxplot" de L8

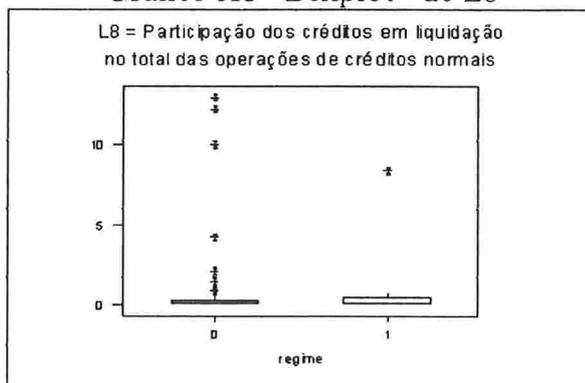


Gráfico A9 "Boxplot" de L9

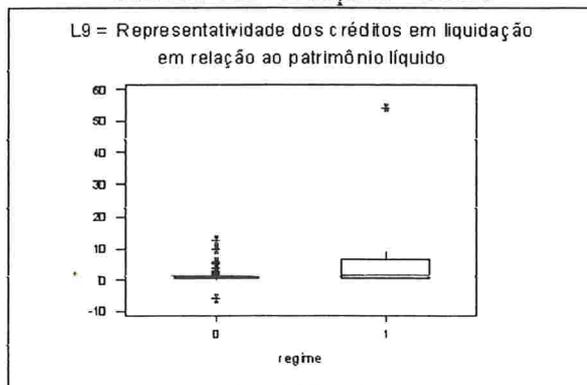


Gráfico A10 "Boxplot" de L10

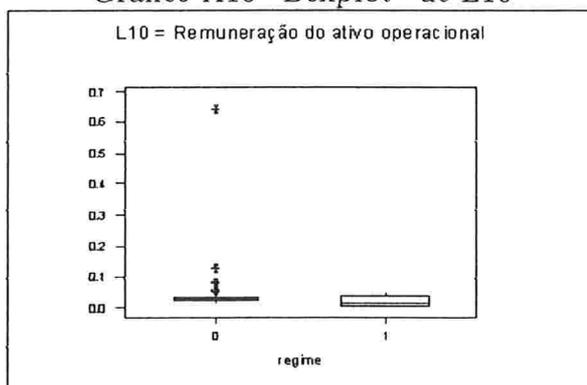


Gráfico A11 "Boxplot" de L11

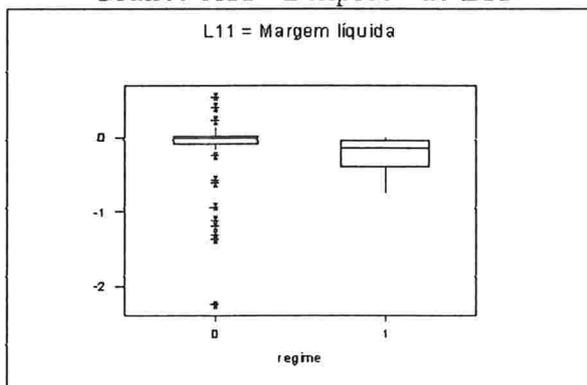


Gráfico A12 "Boxplot" de L12

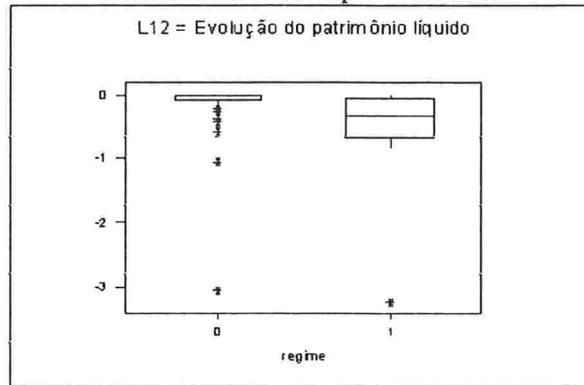


Gráfico A13 "Boxplot" de L13

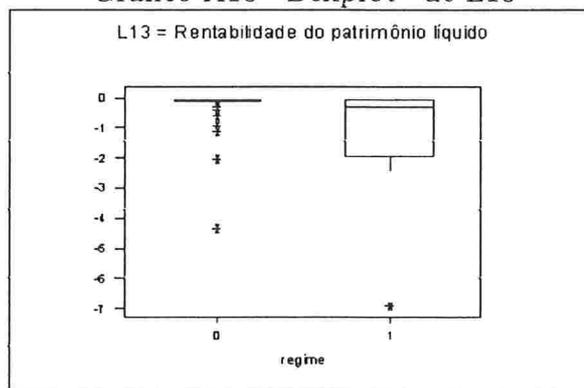


Gráfico A14 "Boxplot" de L14

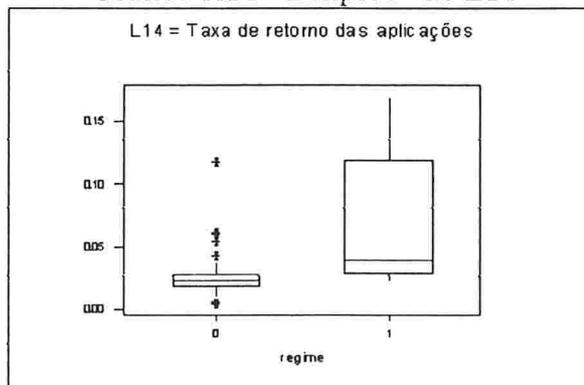


Gráfico A15 "Boxplot" de L15

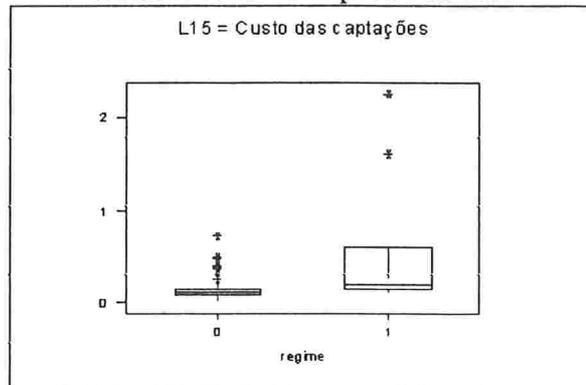


Gráfico A16 "Boxplot" de L16

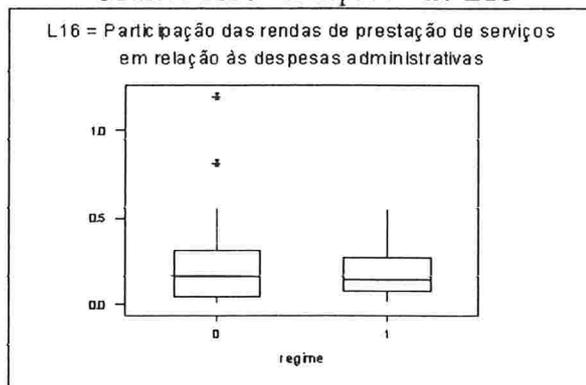


Gráfico A17 "Boxplot" de L17

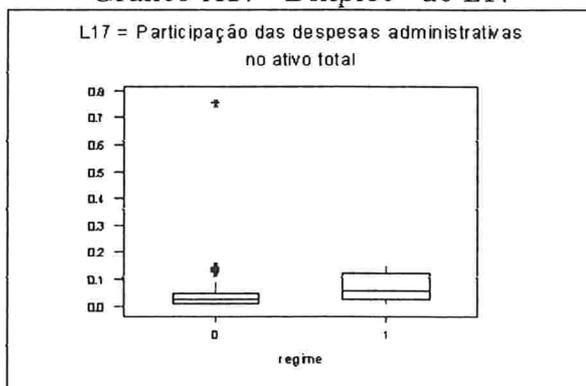


Gráfico A18 "Boxplot" de RL1

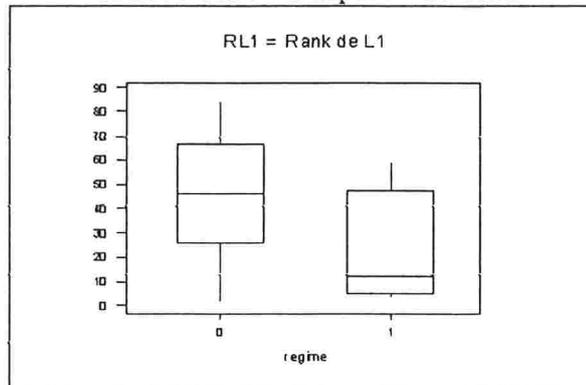


Gráfico A19 "Boxplot" de RL2

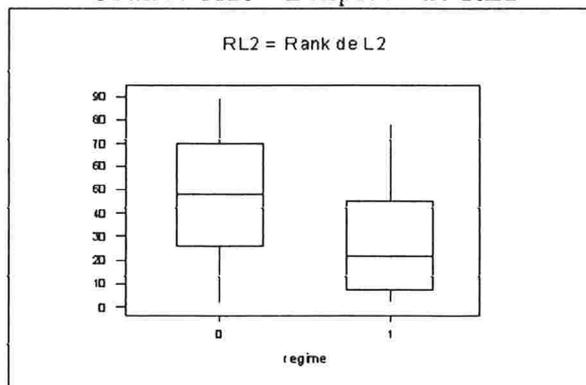


Gráfico A20 "Boxplot" de RL3

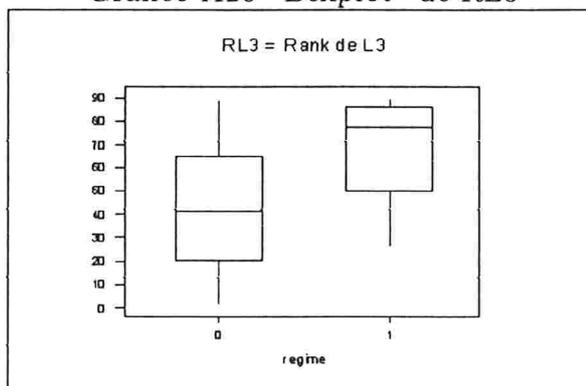


Gráfico A21 "Boxplot" de RL4

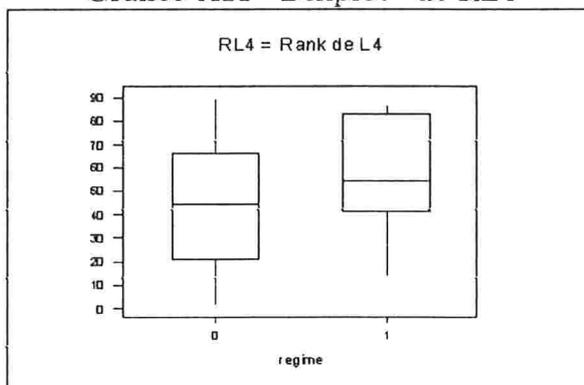


Gráfico A22 "Boxplot" de RL5

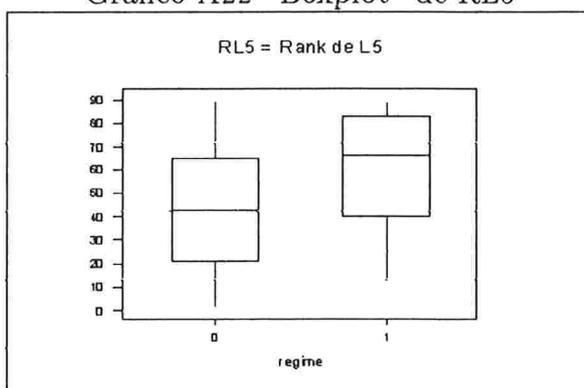


Gráfico A23 "Boxplot" de RL6

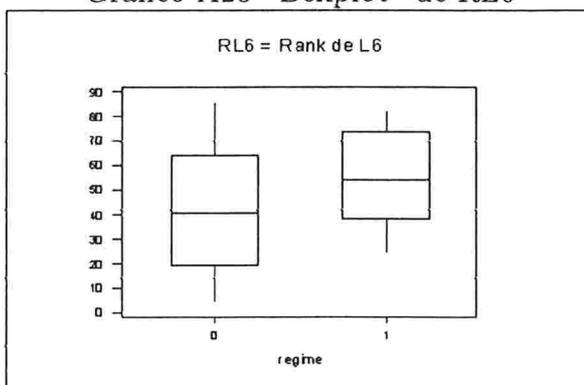


Gráfico A24 "Boxplot" de RL7

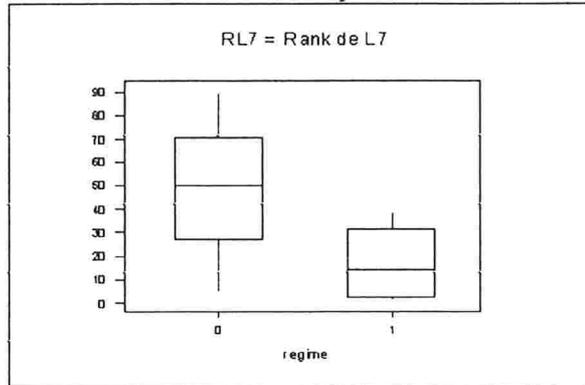


Gráfico A25 "Boxplot" de RL8

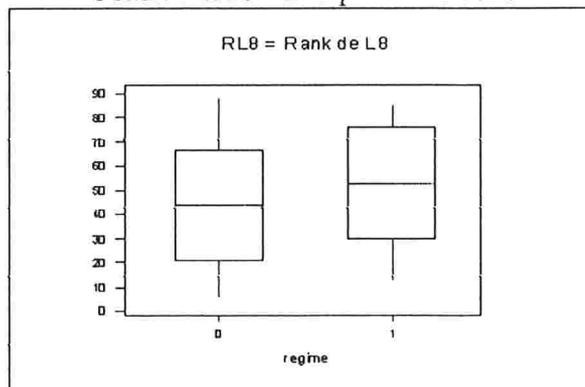


Gráfico A26 "Boxplot" de RL9

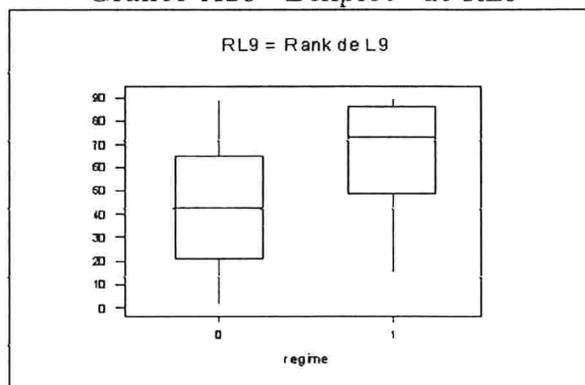


Gráfico A27 "Boxplot" de RL10

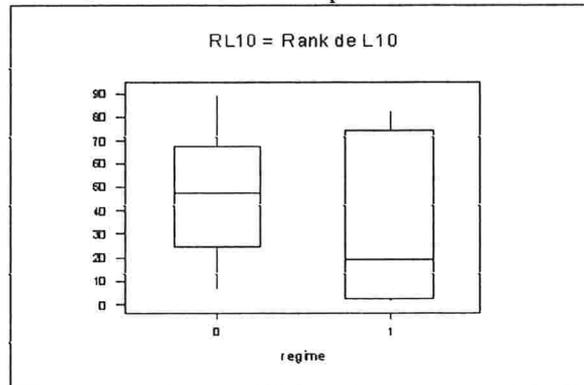


Gráfico A28 "Boxplot" de RL11

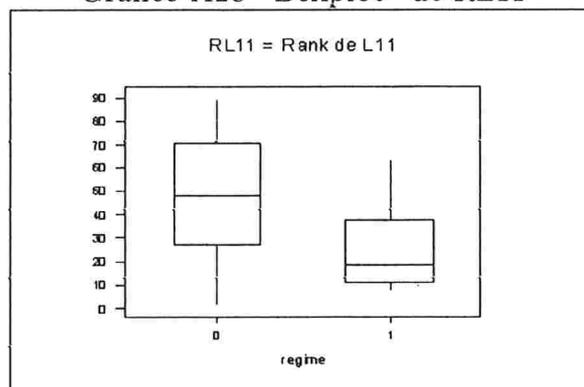


Gráfico A29 "Boxplot" de RL12

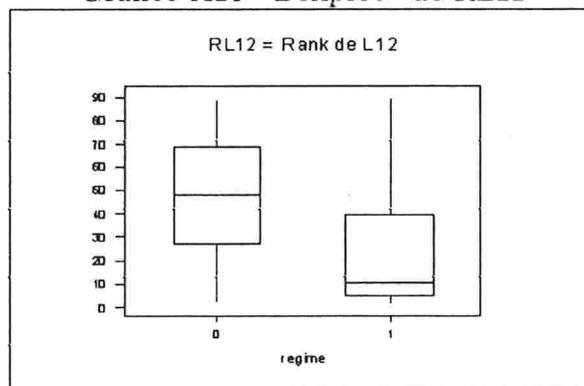


Gráfico A30 "Boxplot" de RL13

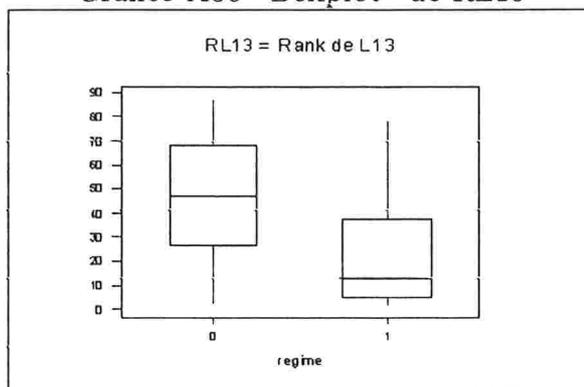


Gráfico A31 "Boxplot" de RL14

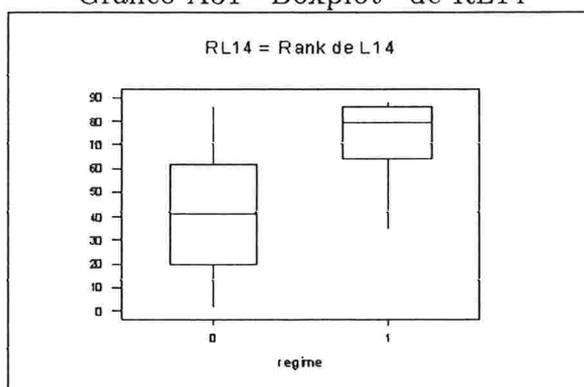


Gráfico A32 "Boxplot" de RL15

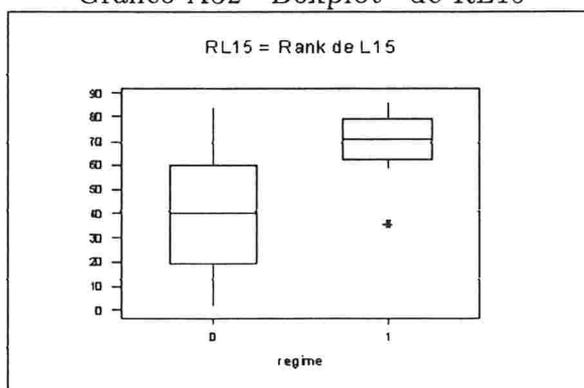


Gráfico A33 "Boxplot" de RL16

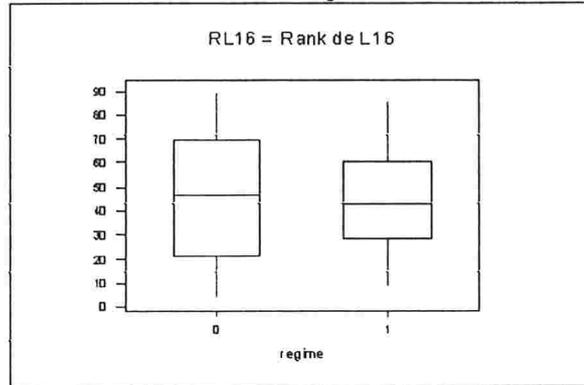
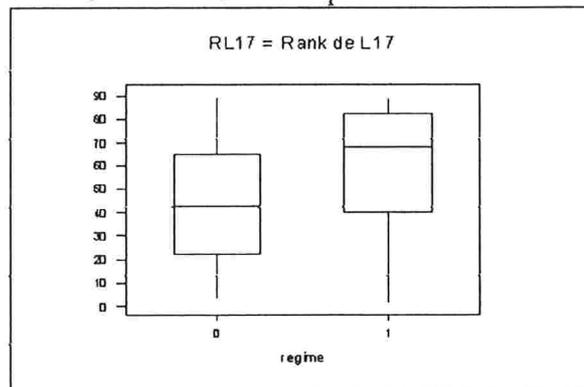


Gráfico A34 "Boxplot" de RL17



Apêndice B

Base de Dados

Tabela B1 Valores dos indicadores L1 até L6

Obs.	L1	L2	L3	L4	L5	L6
1	2,8503	0,9770	1,3713	18,0408	1,0771	0,0385
2	4,3348	1,0090	0,9273	8,0928	1,0108	0,0024
3	1,6192	1,0675	0,2678	11,9044	0,9451	0,3372
4	6,2694	1,0033	0,9224	23,5545	1,1529	0,0007
5	1,3316	1,0262	0,5695	16,8068	0,9934	0,0017
6	43,4806	1,0651	0,3587	9,8647	1,0544	0,0065
7	8,0090	0,9759	1,3591	14,8899	1,0290	0,0308
8	0,6439	0,9962	1,0428	11,2896	1,0352	0,6439
9	3,8196	1,0158	0,6107	26,0022	1,0128	0,0247
10	1,4419	1,0706	0,4329	8,2463	0,9423	0,0658
11	11,4785	1,0967	0,3410	7,0669	0,9489	0,0048
12	1,2116	1,0290	0,4926	17,6402	0,9786	0,0929
13	0,3076	0,4630	0,1908	10,4478	2,1969	0,3838
14	1,0868	0,9867	1,4512	39,4703	1,0479	0,0820
15	1,2444	1,0380	0,7250	7,9783	0,9635	0,0480
16	0,8888	0,9771	1,3683	18,3348	1,0255	0,7391
17	1,5818	1,0993	0,5352	4,8185	0,9170	0,1102

(continua)

Tabela B1 Valores dos indicadores L1 até L6

Obs.	L1	L2	L3	L4	L5	L6
18	3,5279	0,9850	5,1213	274,4857	1,0180	0,0972
19	0,0793	1,2006	0,2433	3,9732	0,8356	0,6986
20	.	14,1200	0,3546	0,0492	0,0853	.
21	1,0744	0,9930	1,1253	17,7187	1,0153	0,0845
22	1,2978	1,0069	0,9509	9,0858	0,9970	0,2470
23	1,0626	1,0176	0,5892	23,8217	0,9965	0,1905
24	1,7947	1,0133	0,8222	14,3231	0,9871	0,0837
25	1,1678	1,0141	0,8455	13,3504	0,9863	0,1531
26	2,1240	0,9928	1,0990	13,5285	1,0080	0,0563
27	1,6655	1,1390	0,0675	6,7506	0,8780	0,7268
28	2,2111	1,0540	0,4839	10,0591	0,9492	0,9166
29	2,1336	1,0751	0,2256	10,3416	0,9312	0,3248
30	1,6602	0,9630	1,2145	8,1367	1,0440	0,2318
31	0,7493	1,0122	0,8751	10,9029	0,9892	0,3576
32	9,6791	1,0306	0,1000	29,5557	0,9705	0,7789
33	1,4351	1,0443	0,2167	17,7832	0,9577	0,9133
34	1,1539	1,0085	0,7547	29,3571	0,9922	0,9183
35	2,5017	1,0089	0,9239	11,3433	1,0009	0,1619
36	1,5326	0,9600	1,7077	23,9514	1,0470	0,3811
37	1,2011	1,0133	0,9062	26,5175	0,9898	0,1912
38	2,1414	0,9843	1,1341	8,7256	1,0245	0,0020
39	1,1439	0,8245	1,5632	22,9414	1,2314	0,0267
40	1,0083	1,0077	0,6273	52,0697	0,9924	0,7312
41	11843,2646	1,6548	0,5289	0,7195	0,6067	0,2673
42	.	124052,1300	0,9995	0,0000	0,0000	.

(continua)

Tabela B1 Valores dos indicadores L1 até L6

Obs.	L1	L2	L3	L4	L5	L6
43	4,0810	0,9960	1,0556	15,9553	1,0066	0,0000
44	1,3242	1,0405	0,0788	23,1448	0,9611	0,6755
45	.	1,1870	0,2085	4,2319	0,8581	.
46	3,1026	1,0235	0,1255	37,5780	0,9771	0,6416
47	0,4906	1,9020	0,0220	1,0861	0,5580	0,8701
48	1,1694	1,0713	0,5218	7,1565	1,0698	0,0000
49	1,5651	0,9963	1,0684	20,7929	1,0053	0,1112
50	1,1749	1,0253	0,7420	11,5430	0,9778	0,4188
51	1,5892	1,0370	0,5459	13,9475	0,9681	0,3859
52	1,5773	1,0323	0,4232	18,8378	0,9690	0,0287
53	.	19,5696	0,0387	0,0518	0,0511	.
54	70,2539	1,0536	0,2547	14,1893	0,9498	0,0000
55	2,2090	1,0987	0,2301	8,0254	0,9117	0,0049
56	2,3709	1,1561	0,0576	6,0548	0,8650	0,0000
57	0,3670	1,3429	0,0000	2,9165	0,7447	1,0000
58	1,1920	0,9723	1,5286	20,1975	1,0340	0,4972
59	1,5895	0,9610	1,2848	7,9465	1,0453	0,2593
60	1,4412	0,9988	1,0092	7,0107	1,0088	0,0017
61	1,3272	0,9576	1,3276	9,8206	1,0448	0,4126
62	1,0994	0,9876	1,2027	20,1879	1,0159	0,1070
63	1,0321	1,0285	0,7257	12,9269	0,9752	0,1568
64	3,6639	0,9750	1,1135	4,6695	1,0279	0,0197
65	1,6088	0,9033	-5,7537	-72,0991	1,1209	0,0344
66	3,7136	1,0427	0,1576	19,8108	0,9602	0,0666
67	1,1228	1,0010	0,9917	8,5302	1,0108	0,6696

(continua)

Tabela B1 Valores dos indicadores L1 até L6

Obs.	L1	L2	L3	L4	L5	L6
68	1,5884	0,9414	4,1288	53,4227	1,0876	0,0343
69	1,2111	1,0332	0,0986	27,1863	0,9684	0,6404
70	2,7857	1,0562	0,5712	7,8033	0,9475	0,2953
71	1,8847	1,0161	0,8067	13,4160	0,9864	0,3880
72	0,5229	0,8953	1,6332	8,6826	1,1244	0,9473
73	1,0237	1,0148	0,7147	22,5669	0,9869	0,4226
74	1,6382	1,0119	0,6829	28,7708	0,9882	0,2144
75	1,0883	0,9868	1,1964	15,5212	1,0418	0,1620
76	1,3945	1,0619	0,4829	8,4022	0,9485	0,2946
77	13,3758	0,9889	1,1363	12,2779	1,0490	0,0000
78	14,3434	1,0107	0,6556	35,7962	0,9964	0,0275
79	.	1,4018	0,4034	1,4847	0,7134	0,0000
80	0,0140	0,8421	-0,1202	-7,0960	1,1922	1,0000
81	0,4922	0,9272	3,1690	29,7731	1,1127	0,3166
82	1,4427	1,1098	0,4251	5,2344	0,9019	0,0589
83	0,5438	0,9808	1,2235	11,5543	1,0300	0,5358
84	0,1904	1,0106	0,8807	11,5670	0,9917	0,7087
85	0,1102	0,7416	8,8131	46,6437	1,5280	0,7378
86	2,0006	0,9865	1,5616	46,8514	1,0321	0,0414
87	0,2321	1,0257	0,7324	11,3105	0,9753	0,9234
88	0,9019	0,8791	2,4650	34,2451	1,1414	0,1634
89	1,9757	1,0018	0,9822	14,6546	0,9997	0,3395
90	0,7276	0,9177	2,2657	15,3479	1,1077	0,3139

Tabela B2 Valores dos indicadores L7 até L12

Obs.	L7	L8	L9	L10	L11	L12
1	1,0142	0,8939	3,5370	0,0324	0,0197	-0,0073
2	0,9956	0,3352	1,0222	0,0277	-0,1459	-0,1142
3	0,9956	0,1372	0,5573	0,0167	0,0572	-0,0046
4	0,9892	0,2092	2,8817	0,0194	-0,0097	-0,0427
5	1,0158	0,0854	0,3293	0,0384	0,0040	0,0009
6	1,0061	0,3056	0,2365	0,0277	0,0013	-0,0354
7	1,0242	0,1419	0,8841	0,0247	-0,0997	-0,1523
8	0,9348	0,1725	0,6596	0,0341	-0,0710	-0,1532
9	0,9980	0,1136	0,8653	0,0341	-0,0074	-0,0097
10	0,9700	0,0859	0,3686	0,0262	-0,0009	-0,0062
11	1,0179	0,9970	4,9060	0,0317	0,0282	-0,0012
12	0,9116	0,0663	0,3375	0,0150	-0,0287	-0,0833
13	1,0078	0,3923	0,1867	0,0204	-1,2046	0,0167
14	0,9799	0,1656	2,2797	0,0274	-1,1126	-0,0836
15	0,8705	0,0640	0,0074	0,0556	-0,1267	-0,0968
16	1,0044	0,0382	0,1767	0,0219	0,0156	-0,0087
17	0,9915	0,1941	0,3563	0,0248	0,0055	0,0011
18	0,8206	0,3443	3,3727	0,0064	-0,6168	-0,1051
19	0,9586	0,1451	0,5443	0,0330	-0,0196	-0,0069
20	0,8450	12,9919	1,3875	0,0779	-2,2473	-0,0427
21	0,9899	0,1574	1,3815	0,0253	-0,1114	-0,0704
22	1,0241	0,0080	0,0238	0,0268	0,0135	-0,0203
23	1,0218	2,1250	0,3494	0,0237	-0,0366	-0,0210
24	0,9950	0,3014	0,0384	0,0293	-0,0077	-0,1628
25	0,9057	0,0000	0,0000	0,0220	-0,0698	-0,0463

(continua)

Tabela B2 Valores dos indicadores L7 até L12

Obs.	L7	L8	L9	L10	L11	L12
26	1,0061	0,0125	0,0258	0,0308	0,0467	0,0027
27	0,8748	0,0000	0,0000	0,0032	0,2291	0,0015
28	0,9754	1,4912	5,6326	0,0294	-0,1257	-0,0240
29	1,0172	0,1037	0,3084	0,0440	0,0065	-0,0106
30	0,7407	0,4296	0,0671	0,0434	-0,1152	-0,0432
31	1,0086	0,0878	0,3179	0,0281	0,0921	0,0054
32	1,0465	0,0022	0,0136	0,0160	0,0285	-0,0364
33	0,9806	0,1171	0,0731	0,0246	-0,0212	-0,0071
34	0,9466	0,0793	0,0874	0,0220	-0,0053	-0,0142
35	0,9592	0,1205	0,3953	0,0164	0,0435	-0,0042
36	0,9697	0,3316	1,4326	0,0475	-0,0989	-0,0573
37	0,9814	0,0733	0,1603	0,0184	-0,0124	-0,3612
38	1,0016	0,0281	0,0857	0,0348	0,0280	-0,0173
39	0,9093	0,2804	1,2708	0,0188	-1,2939	-3,0709
40	0,8463	0,2568	0,0382	0,0126	-0,1417	-0,0789
41	0,9956	0,0000	0,0000	0,0213	0,5420	0,0041
42	0,0560	0,0000	0,0000	0,0827	-0,5674	-0,2599
43	1,0102	0,9310	0,8381	0,0220	0,0220	-0,0096
44	0,9658	0,0000	0,0000	0,0464	-0,0809	-0,1685
45	0,6868	.	0,1787	0,0130	0,1422	-0,0214
46	1,0024	0,0187	0,0063	0,0255	-0,0045	-0,0018
47	1,0175	0,1069	0,1227	0,0514	0,0393	-0,0004
48	0,9155	10,0280	12,3552	0,6473	-0,0217	-0,1052
49	0,9804	0,0354	0,1111	0,0201	0,0146	-0,0086
50	0,9588	0,1363	0,2723	0,0268	0,0126	-0,0096

(continua)

Tabela B2 Valores dos indicadores L7 até L12

Obs.	L7	L8	L9	L10	L11	L12
51	1,0317	0,0235	0,0662	0,0168	0,0093	-0,0099
52	0,9990	0,0112	0,0096	0,0199	0,0040	-0,0097
53	0,9791	0,0000	0,0000	0,0215	-1,3770	-0,0331
54	0,9811	0,0308	0,2043	0,1223	0,0051	0,0022
55	1,0092	0,2944	0,8207	0,0151	0,0062	-0,0500
56	0,9966	0,0000	0,0000	0,0173	0,1286	-0,0109
57	0,8893	0,0000	0,0000	0,0178	0,4075	0,0055
58	0,9812	0,1061	0,6165	0,0314	0,0004	-0,4313
59	1,0012	0,3622	0,5978	0,0265	0,0529	0,0012
60	1,0075	0,0231	0,0661	0,0206	0,0006	-0,0460
61	0,9792	0,0344	0,1248	0,0242	0,0150	0,0006
62	0,9754	0,0944	0,4235	0,0189	0,0203	-0,2237
63	0,9781	0,0964	0,4191	0,0216	0,0258	-0,0146
64	1,0093	0,1251	0,1755	0,0263	0,0120	-0,0093
65	0,6888	0,3155	-5,9212	0,0327	-0,1698	-1,0549
66	0,9996	0,0631	0,1640	0,0253	-0,0415	-0,0386
67	0,9831	0,2091	0,2898	0,0221	-0,0320	0,0006
68	0,9571	0,4321	9,5195	0,0210	-0,2455	-0,5668
69	0,9785	0,6933	2,0255	0,0281	-0,0851	-0,1025
70	0,9889	0,2850	0,9262	0,0220	0,0064	0,0042
71	1,0493	0,0297	0,0983	0,0300	-0,0050	-0,0167
72	0,8771	0,1046	0,3400	0,0170	-0,0808	-0,0429
73	0,7943	4,2416	0,0145	0,0240	-0,0070	-0,0519
74	1,0372	0,0004	0,0011	0,0315	-0,0077	-0,0967
75	1,0176	0,0848	0,5930	0,0268	-0,1586	-0,0836

(continua)

Tabela B2 Valores dos indicadores L7 até L12

Obs.	L7	L8	L9	L10	L11	L12
76	0,9979	0,0818	0,2388	0,0308	0,0122	0,0027
77	0,7731	0,0777	0,2629	0,0313	-0,0532	-0,0303
78	1,0101	0,1250	0,7229	0,0300	-0,0340	-0,0312
79	0,9696	0,0000	0,0000	0,0179	0,1438	-0,0214
80	0,9980	12,2567	-0,7996	0,0159	-0,9363	-0,1623
81	0,9618	0,4023	5,8979	0,0327	-0,3264	-0,4396
82	0,0025	0,0905	0,3849	0,0001	-0,0957	-0,8617
83	0,9685	0,0314	0,1615	0,0388	-0,5591	-0,2040
84	0,9758	0,0802	0,5507	0,0466	0,0104	-0,0045
85	0,9562	0,8068	8,8885	0,0248	-0,7767	-3,2591
86	0,6179	0,2525	2,9770	0,0001	-0,1306	-0,6225
87	0,0035	0,0041	0,0089	0,0000	-0,1373	-0,0805
88	0,0005	0,1339	1,7439	0,0001	-0,0635	-0,5052
89	0,6072	8,3901	54,2255	0,0001	0,0031	0,0228
90	0,9299	0,1455	0,9174	0,0220	-0,3484	-0,1672

Tabela B3 Valores dos indicadores L13 até L17

Obs.	L13	L14	L15	L16	L17
1	0,0161	0,0256	0,0740	0,3224	0,0381
2	-0,0894	0,0231	0,0770	0,0854	0,0677
3	0,0192	0,0154	0,1208	0,0386	0,0107
4	-0,0072	0,0158	0,1014	0,3966	0,0231
5	0,0162	0,0328	0,0832	0,2495	0,0674
6	0,0007	0,0165	0,0617	0,5361	0,0631
7	-0,1651	0,0205	0,0515	0,0669	0,0902
8	-0,1392	0,0362	0,1426	0,0121	0,0310
9	-0,0060	0,0261	0,0825	0,1367	0,0680
10	-0,0004	0,0256	0,0922	0,0104	0,0122
11	0,0082	0,0187	0,0819	0,2640	0,0365
12	-0,0468	0,0240	0,0912	0,1086	0,0134
13	-0,1885	0,0134	0,0833	0,0272	0,7599
14	-4,3427	0,0252	0,1144	0,0894	0,1401
15	-0,0391	0,0599	0,3791	0,1606	0,0183
16	0,0100	0,0152	0,0626	0,3938	0,0441
17	0,0013	0,0274	0,0800	0,2732	0,0556
18	-1,0999	0,0091	0,0921	0,3384	0,0182
19	-0,0055	0,0296	0,1205	0,2941	0,0446
20	-0,5606	0,0616	.	0,0000	0,1380
21	-0,1788	0,0283	0,0760	0,2886	0,0655
22	0,0060	0,0230	0,1250	0,0798	0,0114
23	-0,0386	0,0104	0,3934	0,2921	0,0125
24	-0,0069	0,0240	0,1891	1,2009	0,0061
25	-0,0261	0,0219	0,2246	0,5430	0,0070

(continua)

Tabela B3 Valores dos indicadores L13 até L17

Obs.	L13	L14	L15	L16	L17
26	0,0375	0,0239	0,1590	0,0317	0,0046
27	0,0077	0,0036	0,0101	0,0004	0,0017
28	-0,1240	0,0173	0,1470	0,3420	0,0291
29	0,0059	0,0423	0,1971	0,0508	0,0289
30	-0,0506	0,0372	0,3829	0,2911	0,0190
31	0,0216	0,0291	0,1134	0,1307	0,0389
32	0,0161	0,0164	0,0684	0,0921	0,0049
33	-0,0071	0,0159	0,5016	0,2325	0,0082
34	-0,0047	0,0239	0,1881	0,4656	0,0046
35	0,0088	0,0168	0,1412	0,0620	0,0099
36	-0,2781	0,0244	0,1470	0,1618	0,0258
37	-0,0031	0,0184	0,1133	0,3531	0,0069
38	0,0110	0,0274	0,1322	0,3966	0,0443
39	-0,8816	0,0207	0,0758	0,1744	0,0327
40	-0,1450	0,0188	0,1138	0,4482	0,0015
41	0,0151	0,0121	0,1272	0,0000	0,0018
42	-0,3511	.	.	0,0292	0,0128
43	0,0232	0,0196	0,0718	0,1502	0,0291
44	-0,1507	0,0240	0,3636	0,0387	0,0052
45	0,0231	0,0186	0,2525	0,0000	0,0279
46	-0,0160	0,0243	0,1386	0,1193	0,0118
47	0,0104	0,0535	0,0594	0,0086	0,1353
48	-0,1378	0,0227	0,4811	0,1118	0,0711
49	0,0122	0,0157	0,1140	0,2040	0,0174
50	0,0052	0,0196	0,1039	0,1379	0,0209

(continua)

Tabela B3 Valores dos indicadores L13 até L17

Obs.	L13	L14	L15	L16	L17
51	0,0024	0,0157	0,0776	0,0383	0,0108
52	0,0018	0,0228	0,1301	0,4753	0,0026
53	-0,1512	0,0049	.	0,0000	0,1288
54	0,0085	0,1180	0,7328	0,0008	0,0082
55	0,0047	0,0154	0,0704	0,0756	0,0103
56	0,0140	0,0170	0,0774	0,0270	0,0091
57	0,0101	0,0138	0,0002	0,0031	0,0067
58	0,0007	0,0285	0,1041	0,3095	0,0440
59	0,0115	0,0217	0,0671	0,5570	0,0380
60	0,0003	0,0192	0,0644	0,3802	0,0422
61	0,0061	0,0197	0,1221	0,3482	0,0167
62	0,0137	0,0188	0,0961	0,2067	0,0184
63	0,0088	0,0192	0,1028	0,1724	0,0296
64	0,0017	0,0272	0,0970	0,2676	0,0388
65	.	0,0275	0,1055	0,2962	0,0652
66	-0,0341	0,0272	0,1556	0,0269	0,0072
67	-0,0205	0,0194	0,1189	0,5503	0,0053
68	-2,0107	0,0225	0,1067	0,2196	0,0558
69	-0,1101	0,0311	0,2477	0,0151	0,0082
70	0,0072	0,0218	0,1058	0,1235	0,0246
71	-0,0044	0,0250	0,1291	0,0337	0,0109
72	-0,0336	0,0184	0,1200	0,1180	0,0132
73	-0,0052	0,0244	0,1128	0,2649	0,0045
74	-0,0123	0,0290	0,1613	0,0558	0,0107
75	-0,1788	0,0263	0,0763	0,2514	0,0584

(continua)

Tabela B3 Valores dos indicadores L13 até L17

Obs.	L13	L14	L15	L16	L17
76	0,0032	0,0334	0,1026	0,3304	0,0412
77	-0,1479	0,0353	0,0962	0,1539	0,1304
78	-0,0782	0,0273	0,1032	0,3079	0,0261
79	0,0067	0,0223	0,0854	0,8199	0,0028
80	.	0,0249	0,0952	0,0694	0,0328
81	-1,7503	0,0344	0,1767	0,1628	0,0535
82	-0,1592	0,1115	0,1769	0,1355	0,1509
83	-0,2409	0,0375	0,2477	0,1481	0,0310
84	0,0142	0,0404	0,2220	0,2625	0,0434
85	-6,9240	0,0304	0,1340	0,0783	0,1252
86	-1,3960	0,1431	0,2128	0,0816	0,0002
87	-0,0718	0,0235	1,6176	0,0045	0,0499
88	-2,4335	0,1706	2,2665	0,5485	0,1222
89	0,0070	0,1026	0,1483	0,0751	0,0001
90	-0,2765	0,0209	0,1033	0,2957	0,0603

Tabela B4 Valores do rank dos indicadores L1 até L9

Obs.	RL1	RL2	RL3	RL4	RL5	RL6	RL7	RL8	RL9
1	69	18	78	62	79	23	78	80	83
2	76	41	56	24	57	11	57	71	72
3	53	72	20	44	17	55	58	52	60
4	77	36	54	72	86	7	52	61	80
5	39	56	35	58	46	9	79	35	48
6	83	71	23	33	77	14	69	68	41
7	78	17	76	54	65	20	86	53	69
8	12	31	62	38	69	69	24	58	64
9	74	50	38	75	58	16	62	45	68
10	43	73	27	26	16	28	36	36	53
11	80	76	21	18	20	12	83	82	84
12	34	58	30	59	35	33	21	28	49
13	6	1	13	36	90	59	72	74	39
14	22	24	79	85	75	30	44	57	79
15	35	63	43	22	26	25	15	27	14
16	15	19	77	63	63	77	68	25	37
17	48	78	33	13	14	36	54	59	52
18	71	22	89	90	61	34	12	72	82
19	2	83	18	10	8	72	28	54	58
20	.	88	22	4	3	.	13	89	75
21	21	29	67	60	59	32	53	56	74
22	36	37	57	31	49	47	85	13	19
23	20	52	37	73	48	43	84	84	51
24	57	47	49	52	39	31	55	67	22
25	28	48	50	47	36	38	19	5	7

(continua)

Tabela B4 Valores do rank dos indicadores L1 até L9

Obs.	RL1	RL2	RL3	RL4	RL5	RL6	RL7	RL8	RL9
26	61	28	65	49	54	26	70	15	20
27	56	80	7	16	11	74	16	5	7
28	65	68	29	34	21	81	38	83	85
29	62	75	16	35	15	54	80	41	46
30	55	14	72	25	71	46	9	76	25
31	14	45	51	37	41	57	73	37	47
32	79	59	10	80	30	78	89	11	17
33	41	66	15	61	23	80	46	46	26
34	27	39	47	79	44	82	25	31	28
35	67	40	55	40	51	40	30	47	55
36	45	12	84	74	74	58	35	70	76
37	32	46	53	76	42	44	49	29	33
38	63	21	68	30	62	10	66	19	27
39	26	3	82	70	88	17	20	64	73
40	17	38	39	88	45	75	14	63	21
41	85	86	32	6	5	49	56	5	7
42	.	90	60	3	1	.	4	5	7
43	75	30	63	57	53	4	77	81	67
44	37	64	8	71	25	71	32	5	7
45	.	82	14	11	9	.	7	.	38
46	70	53	11	84	33	68	67	16	13
47	8	87	4	7	4	79	81	44	31
48	29	74	31	19	78	4	22	87	89
49	46	32	64	68	52	37	45	24	30
50	30	54	46	41	34	63	29	51	44

(continua)

Tabela B4 Valores do rank dos indicadores L1 até L9

Obs.	RL1	RL2	RL3	RL4	RL5	RL6	RL7	RL8	RL9
51	50	62	34	50	27	60	87	18	24
52	47	60	25	64	29	19	63	14	16
53	.	89	5	5	2	.	42	5	7
54	84	67	19	51	22	4	47	21	40
55	64	77	17	23	13	13	74	66	66
56	66	81	6	15	10	4	59	5	7
57	7	84	3	9	7	86	18	5	7
58	31	15	80	67	68	65	48	43	63
59	51	13	74	21	73	48	65	73	62
60	42	33	61	17	55	8	71	17	23
61	38	11	75	32	72	62	43	23	32
62	24	26	71	66	60	35	37	39	57
63	19	57	44	46	31	39	40	40	56
64	72	16	66	12	64	15	75	49	36
65	52	7	1	1	83	22	8	69	1
66	73	65	12	5	24	29	64	26	35
67	25	34	59	28	6	0	50	60	45
68	49	10	88	89	80	21	27	77	88
69	33	61	9	77	28	67	41	78	78
70	68	69	36	20	18	51	51	65	71
71	58	51	48	48	37	61	90	20	29
72	10	6	83	29	84	84	17	42	50
73	18	49	42	69	38	64	11	85	18
74	54	44	41	78	40	45	88	10	12
75	23	25	70	56	70	41	82	34	61

(continua)

Tabela B4 Valores do rank dos indicadores L1 até L9

Obs.	RL1	RL2	RL3	RL4	RL5	RL6	RL7	RL8	RL9
76	40	70	28	27	19	50	60	33	42
77	81	27	69	45	76	4	10	30	43
78	82	43	40	83	47	18	76	48	65
79	.	85	24	8	6	4	34	5	7
80	1	4	2	2	87	86	61	88	2
81	9	9	87	81	82	53	31	75	86
82	44	79	26	14	12	27	2	38	54
83	11	20	73	42	66	66	33	22	34
84	4	42	52	43	43	73	39	32	59
85	3	2	90	86	89	76	26	79	87
86	60	23	81	87	67	24	6	62	81
87	5	55	45	39	32	83	3	12	15
88	16	5	86	82	85	42	1	50	77
89	59	35	58	53	50	56	5	86	90
90	13	8	85	55	81	52	23	55	70

Tabela B5 Valores do rank dos indicadores L10 até L17

Obs.	RL10	RL11	RL12	RL13	RL14	RL15	RL16	RL17
1	71	72	66	81	56	12	71	58
2	58	16	20	27	43	16	32	78
3	15	83	70	84	9	51	20	25
4	24	43	40	40	13	32	79	43
5	78	56	78	83	73	22	56	77
6	57	54	43	51	16	5	84	74
7	44	24	19	17	32	3	25	81
8	76	30	18	23	77	61	11	52
9	75	46	60	43	58	21	42	79
10	50	51	69	49	57	27	10	31
11	70	77	74	67	23	20	59	56
12	10	38	28	31	49	25	35	35
13	27	4	89	14	6	23	15	90
14	56	5	27	2	55	47	33	88
15	86	20	24	32	83	80	47	39
16	33	71	64	71	8	6	78	65
17	46	59	79	53	63	19	62	70
18	7	8	22	7	3	26	73	38
19	74	41	68	44	70	50	66	67
20	87	1	39	9	84	.	3	87
21	47	23	31	15	66	14	63	76
22	54	68	51	61	42	53	30	29
23	40	35	50	33	4	82	65	32
24	61	45	16	42	47	71	90	14
25	36	31	35	36	37	75	85	17

(continua)

Tabela B5 Valores do rank dos indicadores L10 até L17

Obs.	RL10	RL11	RL12	RL13	RL14	RL15	RL16	RL17
26	65	81	83	88	45	66	17	10
27	6	88	81	66	1	2	5	4
28	62	21	47	25	19	63	74	49
29	81	62	57	60	81	72	22	48
30	80	22	37	30	78	81	64	41
31	60	84	87	85	69	44	40	60
32	13	78	42	82	15	9	34	11
33	43	40	67	41	14	84	55	20
34	37	48	55	46	46	70	82	9
35	14	80	72	69	17	60	24	23
36	84	25	32	11	52	62	48	45
37	21	42	10	48	21	43	76	16
38	77	76	52	74	64	57	80	66
39	22	3	2	8	33	13	51	54
40	8	17	30	22	25	45	81	3
41	30	90	85	80	5	54	3	5
42	88	9	11	10	.	.	16	33
43	35	74	61	87	29	11	45	50
44	82	28	14	20	48	79	21	12
45	9	86	48	86	22	78	3	47
46	49	50	73	38	50	59	38	30
47	85	79	75	73	82	4	9	86
48	90	39	21	24	40	83	36	80
49	26	69	65	76	11	46	52	37
50	55	67	62	59	30	37	43	42

(continua)

Tabela B5 Valores do rank dos indicadores L10 até L17

Obs.	RL10	RL11	RL12	RL13	RL14	RL15	RL16	RL17
51	16	63	58	56	12	18	19	27
52	25	57	59	55	41	56	83	6
53	31	2	44	19	2	.	3	84
54	89	58	82	68	87	85	6	19
55	11	60	34	58	10	10	28	24
56	18	85	56	78	18	17	14	22
57	19	89	88	72	7	1	7	15
58	68	52	9	52	67	38	70	64
59	52	82	80	75	35	8	88	57
60	28	53	36	50	26	7	77	62
61	42	70	76	62	31	52	75	36
62	23	73	12	77	24	29	53	40
63	32	75	54	70	27	34	50	51
64	51	65	63	54	61	31	61	59
65	72	14	3	.	65	39	68	75
66	48	34	41	34	60	65	13	18
67	39	37	77	37	28	48	87	13
68	29	13	6	4	39	41	54	71
69	59	27	23	26	72	76	12	21
70	38	61	86	65	36	40	39	44
71	64	49	53	47	54	55	18	28
72	17	29	38	35	20	49	37	34
73	41	47	33	45	51	42	60	8
74	69	44	25	39	68	67	23	26
75	53	15	26	16	59	15	57	72

(continua)

Tabela B5 Valores do rank dos indicadores L10 até L17

Obs.	RL10	RL11	RL12	RL13	RL14	RL15	RL16	RL17
76	66	66	84	57	74	33	72	61
77	67	33	46	21	76	30	46	85
78	63	36	45	28	62	35	69	46
79	20	87	49	63	38	24	89	7
80	12	6	17	.	53	28	26	55
81	73	12	8	5	75	68	49	69
82	3	26	4	18	86	69	41	89
83	79	10	13	13	79	77	44	53
84	83	64	71	79	80	74	58	63
85	45	7	1	1	71	58	29	83
86	5	19	5	6	88	73	31	2
87	1	18	29	29	44	86	8	68
88	2	32	7	3	89	87	86	82
89	4	55	90	64	85	64	27	1
90	34	11	15	12	34	36	67	73

Apêndice C

Probabilidades Estimadas para os Modelos 1 a 6

Tabela C1 Probabilidades estimadas para os Modelos 1 a 6

Obs.	1	2	3	4	5	6
1	0,999	0,999	0,998	1,000	1,000	1,000
2	1,000	1,000	1,000	0,999	1,000	1,000
3	1,000	1,000	1,000	0,976	1,000	1,000
4	1,000	1,000	1,000	0,998	1,000	1,000
5	1,000	1,000	1,000	0,999	1,000	1,000
6	1,000	1,000	1,000	1,000	1,000	1,000
7	0,999	0,999	1,000	0,999	1,000	1,000
8	0,972	0,802	0,767	0,750	.	0,631
9	1,000	1,000	1,000	0,999	1,000	1,000
10	1,000	1,000	1,000	0,999	0,997	1,000
11	1,000	1,000	1,000	1,000	1,000	1,000
12	0,999	0,999	0,999	0,816	0,961	0,873
13	1,000	1,000	1,000	0,990	1,000	1,000
14	0,999	0,997	0,994	0,609	0,981	1,000
15	0,969	0,732	0,758	0,993	.	1,000

(continua)

Tabela C1 Probabilidades estimadas para os Modelos 1 a 6

Obs.	1	2	3	4	5	6
16	0,996	0,985	0,965	1,000	1,000	1,000
17	1,000	1,000	1,000	0,952	1,000	1,000
18	1,000	.	.	0,990	1,000	1,000
19	0,997	0,990	0,977	0,851	.	0,163
20
21	0,999	0,998	0,997	0,886	0,971	0,951
22	0,999	0,999	0,999	0,999	1,000	1,000
23	1,000	1,000	1,000	0,995	1,000	1,000
24	1,000	1,000	1,000	1,000	1,000	1,000
25	0,998	0,998	0,997	0,992	0,977	1,000
26	1,000	0,999	0,999	1,000	1,000	1,000
27	0,999	1,000	0,999	1,000	1,000	1,000
28	0,994	0,993	0,980	0,988	1,000	1,000
29	0,999	0,996	0,994	0,882	1,000	1,000
30	0,766	0,568	0,623	0,957	0,599	1,000
31	0,999	0,996	0,993	0,879	0,993	0,987
32	1,000	1,000	0,999	1,000	1,000	1,000
33	0,999	0,997	0,993	0,992	1,000	1,000
34	0,996	0,938	0,883	0,899	0,959	1,000
35	0,999	1,000	1,000	0,939	1,000	1,000
36	0,992	0,965	0,938	0,996	0,999	1,000
37	1,000	1,000	0,999	0,998	1,000	1,000
38	0,999	0,999	0,999	0,998	1,000	1,000

(continua)

Tabela C1 Probabilidades estimadas para os Modelos 1 a 6

Obs.	1	2	3	4	5	6
39	0,996	0,995	0,992	0,990	0,985	1,000
40	0,998	0,969	0,930	0,979	0,952	0,871
41	1,000	1,000	1,000	1,000	1,000	1,000
42
43	1,000	1,000	1,000	1,000	1,000	1,000
44	1,000	0,998	0,995	1,000	0,996	1,000
45
46	1,000	0,999	0,998	0,987	1,000	1,000
47	0,993	0,815	0,777	0,997	0,966	0,991
48	0,995	1,000	1,000	0,993	0,980	1,000
49	1,000	1,000	1,000	0,999	1,000	1,000
50	0,999	0,998	0,997	1,000	0,999	1,000
51	1,000	1,000	1,000	1,000	1,000	1,000
52	1,000	1,000	1,000	0,997	1,000	1,000
53
54	0,936	.	.	0,997	1,000	1,000
55	1,000	1,000	1,000	1,000	1,000	1,000
56	1,000	1,000	1,000	1,000	1,000	1,000
57	0,995	0,994	0,981	1,000	0,991	1,000
58	0,991	0,938	0,901	0,935	0,985	1,000
59	0,998	0,998	0,996	1,000	1,000	1,000
60	1,000	1,000	1,000	1,000	1,000	1,000
61	0,996	0,994	0,986	0,994	1,000	1,000

(continua)

Tabela C1 Probabilidades estimadas para os Modelos 1 a 6

Obs.	1	2	3	4	5	6
62	0,999	0,999	0,999	0,994	1,000	1,000
63	1,000	1,000	1,000	0,991	0,999	1,000
64	0,999	0,999	0,999	0,994	1,000	1,000
65	1,000	1,000	1,000	0,999	0,846	1,000
66	1,000	1,000	1,000	0,981	1,000	1,000
67	0,996	0,990	0,977	0,998	1,000	1,000
68	0,586	.	0,415	0,528	1,000	1,000
69	0,999	0,995	0,990	0,359	0,937	1,000
70	0,999	0,999	0,999	0,966	1,000	1,000
71	1,000	0,999	0,998	0,998	1,000	1,000
72	0,806	0,505	0,535	0,809	0,956	0,977
73	0,987	0,965	0,930	0,998	0,235	0,871
74	1,000	0,999	0,999	0,997	1,000	1,000
75	0,999	0,998	0,997	0,984	1,000	.
76	0,999	0,998	0,997	0,978	0,998	1,000
77	0,952	0,925	0,894	0,985	0,993	1,000
78	1,000	1,000	1,000	0,998	1,000	1,000
79	1,000	1,000	1,000	1,000	.	.
80	0,998	0,995	0,986	0,349	0,970	.
81	0,569	0,125	0,300	0,211	0,042	0,416
82	0,000	0,000	0,000	0,000	0,026	0,000
83	0,983	0,855	0,819	0,751	0,043	0,711

(continua)

Tabela C1 Probabilidades estimadas para os Modelos 1 a 6

Obs.	1	2	3	4	5	6
84	0,984	0,806	0,772	0,440	0,031	0,277
85	0,000	0,000	0,000	0,002	0,014	0,000
86	0,001	0,000	0,000	0,001	0,321	0,028
87	0,000	0,000	0,000	0,000	0,031	0,000
88	0,000	0,000	0,000	0,000	0,000	0,000
89	0,000	0,000	0,000	0,000	0,351	0,041
90	0,951	.	0,858	0,778	0,916	.