

5

METODOLOGIA

Neste capítulo será apresentada a metodologia utilizada na investigação. Inicialmente o tipo de pesquisa será abordado, seguido do universo e da seleção da amostra. A coleta dos dados e seu tratamento serão apresentados a seguir, bem como as limitações do método.

5.1

Tipo de pesquisa

A presente pesquisa é do tipo explicativo segundo a abordagem tradicional positivista.

5.2

Universo e Amostra

O universo compreende as emissões de ações, públicas e privadas feitas pelas empresas listadas na Bovespa no período estudado (1995 a 2002, inclusive).

A amostra, no que diz respeito às atividades do mercado primário de ações no Brasil, consiste nas emissões primárias públicas mensalmente relacionadas pela CVM e pelas emissões particulares³⁴, levantadas nos demonstrativos da evolução do capital social das empresas publicados na Gazeta Mercantil pela Bolsa de Valores de São Paulo, entre janeiro de 1995 e dezembro de 2002.

A pesquisa inicial efetuada e apresentada na Tabela 1 aponta para uma população total de 776 emissões de ações, sendo 653 particulares e 123 públicas, entre janeiro de 1995 e dezembro de 2002. Da população original total foram retiradas 60 emissões correspondentes a 32 empresas com patrimônio líquido negativo, que inviabilizaria a análise de lucratividade através da razão LL/PL

³⁴ As emissões particulares não são registradas pela CVM, não havendo estatísticas sobre elas naquela Comissão.

(para prejuízo e PL negativo essa razão seria positiva). Além disso, o patrimônio líquido negativo também acarretaria distorções na variável grau de endividamento (exigível de longo prazo/patrimônio líquido). Dessas emissões, 2 foram públicas e 58 privadas. Da amostra foram, também, retiradas as 72 emissões realizadas por instituições financeiras, para evitar problemas em razão do elevado endividamento característico do setor. Dessas, 11 eram públicas e 61, privadas. Foram, ainda, retiradas, 9 emissões (sendo cinco públicas e quatro privadas) de empresas com falta de dados para algumas das variáveis utilizadas. Após os primeiros testes estatísticos, foram determinadas 6 emissões (5 públicas e uma privada) que correspondiam a empresas com endividamento elevadíssimo ou lucratividade extremamente negativa, que foram retiradas como *outliers*³⁵. A amostra final utilizada neste trabalho consistiu, portanto, de 629 emissões, sendo 101 públicas e 528, privadas. A Tabela 10 abaixo apresenta resumo da amostra utilizada neste trabalho.

Tabela 10 – Percentuais da amostra utilizada em número de observações e volume em relação à amostra original apresentada na Tabela 1

Emissões	Total		Públicas		Privadas	
	Volume R\$milhões	Número	Volume R\$milhões	Número	Volume R\$milhões	Número
Amostra original	96.250	776	25.747	123	70.046	653
Amostra Final	56.565	629	13524,1	101	43.040,8	528
% da amostra original	58,8%	81,1%	52,5%	82,1%	61,4%	80,9%

Importante ressaltar que a unidade observacional do presente trabalho é a emissão de ações. Portanto, uma empresa pode apresentar, em diferentes anos do período estudado, dois tipos de emissão de ações: pública ou privada.

³⁵ As estatísticas descritivas apresentavam distorções, muito embora os modelos discriminante e logístico acabassem por eliminá-las como outliers. Optou-se, porém pela sua retirada, o que acabou por não influenciar as conclusões obtidas.

A amostra utilizada apresenta, portanto, 81,1% das emissões realizadas entre janeiro de 1995 e dezembro de 2002, sendo 82,1% das emissões públicas e 80,9% das particulares.

5.3

Coleta dos Dados

Os dados da pesquisa bibliográfica foram coletados em teses, dissertações, livros, periódicos, documentos de trabalho universitário e revistas especializadas, todos eles citados no corpo e relacionados ao final deste trabalho.

Na pesquisa de campo foram obtidos os dados numéricos sobre as emissões, manualmente, através do cálculo do valor emitido a partir do número de ações anterior, percentual emitido e preço de cada ação emitida. Esses dados foram obtidos nos jornais Gazeta Mercantil e Valor Econômico, a partir dos demonstrativos de evolução do capital social das empresas divulgados pela BOVESPA. Ajustes foram feitos no caso de sobras, através do cancelamento de ações ou manutenção em tesouraria das empresas.

Para o cálculo do valor captado através de emissões privadas foi utilizada a seguinte fórmula:

$$Valor = P * \left(N_{ex} - \frac{N_{ex}}{1 + \frac{S}{100}} \right)$$

Onde P é o preço da ação emitida, N_{ex} o número de ações ex-subscrição e S o percentual da subscrição.

Os dados sobre as características das empresas, tais como seus balanços e indicadores financeiros³⁶, foram levantados nos Informativos Anuais da CVM, nos Guias Anuais das Empresas de Capital Aberto editados pela Editora IMF para

³⁶ Índices de endividamento, liquidez em bolsa e lucratividade entre outros.

os anos de 1994 a 2003³⁷, em informativos públicos das próprias empresas e/ou em outros meios de pesquisa (eletrônica via internet, por exemplo).

Dessa maneira foram obtidos dados sobre patrimônio líquido, lucro líquido, exigível de longo prazo (dívida), lucratividade (lucro líquido dividido pelo patrimônio líquido), endividamento (exigível de longo prazo dividido pelo patrimônio líquido) e concentração de propriedade. Essa última foi levantada a partir da propriedade direta dos três maiores acionistas ordinários listada nos Informativos Anuais da CVM, por empresa. Optou-se pela propriedade direta por esta sinalizar a concentração, embora se reconheça que estruturas indiretas de controle são relevantes, como utilizadas no trabalho de Da Silva (2002), entre outros. Como o objetivo aqui não era especificamente estudar o controle, utilizou-se essa simplificação. Os dados sobre o percentual das ações ordinárias de propriedade dos três maiores acionistas, foram obtidos e somados para cada empresa emissora para o exercício imediatamente anterior à emissão.

Os dados sobre o comportamento do mercado de capitais brasileiro foram obtidos a partir dos relatórios periódicos da CVM, da BOVESPA, da ABAMEC, da CNBV (Comissão Nacional de Bolsas de Valores), bem como de teses, dissertações e trabalhos acadêmicos. Nessas publicações foram obtidos os dados relativos à liquidez em bolsa medida em volume e número de negócios realizados nos anos em que ocorreram as emissões, para cada empresa emissora.

A classificação das empresas em Estatal, Privada ou Holding foi feita através dos informativos anuais da CVM e dos anuários editados pela Editora IMF.

A seguir, será apresentado o procedimento para tratamento dos dados.

5.4

Tratamento dos Dados

Para compreender as razões pelas quais empresas de capital aberto no Brasil decidem emitir ações pública ou privadamente, desenvolveu-se modelo que

³⁷ O Informativo do IMF- Companhias Abertas de 2003 apresenta os dados sobre as empresas para os anos de 2000, 2001 e 2002.

explique essa tomada de decisão. A construção do modelo, bem como sua validação são o objeto das próximas seções.

Em primeiro lugar foi definida a variável dependente. Criou-se uma variável categórica, do tipo binária, com valor zero para emissões públicas e valor um para emissões particulares. A seguir, definiu-se as variáveis explicativas relevantes para a obtenção do perfil das empresas emissoras, de modo a defini-lo. Este passo foi facilitado pela literatura revista, especialmente os trabalhos de Ness e Pereira (1980), Da Costa(1990), Furtado (1997), Siffert (1998), Leal (1993, 2000 e 2004), Leal, da Silva e Valadares((2000), Chachat (2000) e Da Silva (2002), no Brasil; Cronqvist e Nilsson (2001) e Mikkleson et al.(1997) no âmbito internacional. A seguir, são apresentadas as variáveis utilizadas e os métodos estatísticos aplicados.

5.4.1

Definição das variáveis utilizadas e apresentação dos métodos estatísticos utilizados na investigação

As variáveis utilizadas na investigação são³⁸:

i) Relativa às emissões:

- Tipo de emissão: pública (tipo 0) ou privada (tipo 1) :

Emissão Priv=1, Pub= 0

ii) Relativas à Estrutura de Capital referidas ao ano anterior à emissão.

- Endividamento relativo: exigível de longo prazo dividida pelo patrimônio líquido:

Endivid ELP/PL

³⁸ Os valores em dólares americanos se devem à eventual indisponibilidade de dados em reais nas fontes de pesquisa.

- Endividamento: exigível de longo prazo medida em dólares americanos (exigível de longo prazo):

Exigível L Prz Us\$

- Lucratividade relativa: lucro líquido contábil dividido pelo patrimônio líquido(LL/Pat.Líq.):

Lucrativ. LL/PL

- Lucratividade: lucro líquido medido em dólares americanos:

Lucro Líq. Us\$ 1000

iii) **Relativas à Eficiência de Mercado :**

- grau de assimetria de informação: variável ligada à concentração de controle (percentual de propriedade das ações com direito a voto dos três maiores acionistas imediatamente antes da emissão):

perc3Maciord

- Porte: medido pelo Patrimônio Líquido em dólares americanos:

Patrim. Líq.Us\$

- Porte: medido pelo Volume emitido, medido em reais:

Valor emit R\$

- Liquidez medida pelo volume negociado no ano da emissão, medido em reais:

Liquid Vol neg

- Liquidez medida pelo número de negócios realizados no ano da emissão:

Liquid n° neg

- Tipo de empresa: Estatal (tipo 0), Privada (tipo 1) ou *Holding*(tipo 2):

- i. Estatal é controlada pelo Estado (Federal, Estadual ou Municipal):

Tipoemp0

- ii. Privada é controlada por empresários ou pessoas jurídicas privadas brasileiras e estrangeiras:

Tipoemp1

- iii. *Holding* é uma empresa que controla o capital de outra(s).

Tipoemp2

O tratamento dos dados levantados a partir das variáveis selecionadas *a priori*, seguiu os passos apresentados no Quadro 5.

Quadro 5 – Proposta de tratamento dos dados, passos a serem seguidos, propósitos e os processos estatísticos utilizados

Passos	Procedimento	Objetivo	Método
1	Preparo da base de dados.	Verificar distorções que possam ocorrer, ou seja, se os dados atendem às premissas dos testes estatísticos	Médias e frequências das variáveis observáveis no período estudado.
2	Análise de correlação	Eliminar da amostra variáveis que tenham alta correlação entre si.	Matrizes de correlação com teste de hipótese de correlação linear <i>two-tailed</i>
3	Análise Discriminante	Classificar e agrupar as emissões de acordo com as características das empresas emissoras	Análise Discriminante
4	Regressão Logística com variável dependente binária.	Desenvolver regressão para prever a probabilidade de ocorrência da emissão privada ou pública	Regressão Logística

Dada a natureza categórica da variável dependente, emissões públicas ou privadas, o primeiro método escolhido para explicação e predição foi o da análise discriminante, apresentado a seguir. Optou-se, a seguir por realizar, também, uma regressão logística, da forma explicada na próxima seção. As estatísticas

descritivas do fenômeno, à luz das variáveis utilizadas, também serão relatadas na próxima seção.

5.5

Estatísticas Descritivas, Análise Discriminante e Regressão Logística

Um dos objetivos deste trabalho é classificar os dois grupos de dados dentro da amostra, separando as emissões particulares ou privadas das emissões públicas de ações segundo as características apresentadas pelas variáveis acima relacionadas³⁹. Esta seção está subdividida em três subseções. A primeira trata das estatísticas descritivas. A segunda apresenta o esquema teórico da análise discriminante e a terceira o esquema da regressão logística.

5.5.1

Determinação das Estatísticas Descritivas

Para iniciar a análise, inicialmente foi feito o levantamento das estatísticas descritivas. As médias, medianas, moda, desvios-padrão e variância foram levantados para se ter uma idéia inicial dos comportamentos das variáveis estudadas e do fenômeno investigado.

A Tabela 11 apresenta as estatísticas descritivas para as variáveis inicialmente definidas, já excluídas as empresas com patrimônio líquido negativo e as instituições financeiras.

³⁹ Características relativas às empresas emissoras.

Tabela 11 – Estatísticas Descritivas para as variáveis inicialmente definidas

		Média	Mediana	Desvio Padrão	Nº de observações válidas
Tip de Emissão					
Pública = 0	Variável <i>a priori</i>				
Privada = 1					
0					
	TIPOEMP0 Estatal	0,03	0,00	0,17	101
	TIPOEMP1 Privada	0,72	1,00	0,45	101
	TIPOEMP2 Holding	0,25	0,00	0,43	101
	Valor emit R\$	133.902.058,75	50.185.279,00	207.563.991,43	101
	perc3Maciord	83,26	89,00	19,14	101
	Liquid Vol neg	102.655.584,58	6.068.808,60	267.481.488,78	101
	Liquid nº neg	5.556	163	14.095,72	101
	Exigível L Prz Us\$	137.246,03	31.755,00	325.468,72	101
	Endivid ELP/PL	0,58	0,26	1,27	101
	Patrim Líq Us\$ 1000	307.539,68	130.426,00	593.955,12	101
	Lucro Líq Us\$1000	7.196,25	2.073,00	77.779,96	101
	Lucrativ.LL/PL	-0,11	0,018	1,05	101
1					
	TIPOEMP0 Estatal	0,13	0,00	0,34	528
	TIPOEMP1 Privada	0,58	1,00	0,49	528
	TIPOEMP2 Holding	0,29	0,00	0,45	528
	Valor emit R\$	81.516.590,06	17.272.164,00	235.028.704,32	528
	perc3Maciord	86,58	95,78	16,68	528
	Liquid Vol neg	4.163.427.020,19	108.940,51	82.012.471.901,23	528
	Liquid nº neg	9.084	36	35.651,73	528
	Exigível L Prz Us\$	224.325,66	25298,31	618.245,48	528
	Endivid ELP/PL	1,25	0,25	4,33	528
	Patrim Líq Us\$ 1000	669.853,66	81480,26	2.073.092,47	528
	Lucro Líq Us\$1000	21.375,49	-0,50	142.754,32	528
	Lucrativ.LL/PL	-0,50	-0,000042	2,28	528
Total					
	TIPOEMP0 Estatal	0,08	0,00	0,27	629
	TIPOEMP1 Privada	0,65	1,00	0,48	629
	TIPOEMP2 Holding	0,27	0,00	0,44	629
	Valor emit R\$	107.709.324,42	29.500.000,00	223.090.837,30	629
	perc3Maciord	84,92	92,61	18,02	629
	Liquid Vol neg	2.133.041.301,76	1.219.373,10	57.981.310.675,41	629
	Liquid nº neg	7.320	97	27.144,32	629
	Exigível L Prz Us\$	180.785,84	29.831,69	495.569,13	629
	Endivid ELP/PL	0,92	0,25	3,20	629
	Patrim Líq Us\$ 1000	488.696,67	103.746,66	1.534.410,32	629
	Lucro Líq Us\$1000	14.285,87	226,33	115.080,67	629
	Lucrativ.LL/PL	-0,31	0,0046	1,78	629

Pode-se observar, a partir das estatísticas descritivas preliminares: as emissões públicas de ações parecem ser, em média, maiores do que as privadas, estão relacionadas a empresas privadas, com maior liquidez em Bolsa, menos endividadas, de menor porte, que apresentam maior lucratividade (de fato, lucratividade menos negativa). Pode-se notar, também, a menor concentração de propriedade nas emissões públicas.

As emissões privadas parecem estar relacionadas a empresas do tipo *holding*, menos lucrativas (de fato, de maiores perdas), mais endividadas, de controle mais concentrado, com menor liquidez em Bolsa e de maior porte, o que é contraditório com o menor volume médio de emissões.

As diferenças de escala apresentadas pelas variáveis, bem como o atendimento às premissas de normalidade e linearidade levaram a transformações⁴⁰, objetivando-se obtenção de normalidade e linearidade de comportamento necessárias ao desenvolvimento do modelo discriminante. A seguir é apresentada a construção do modelo passo a passo.

5.5.2

Análise Discriminante – construção do modelo

O método da análise discriminante é apresentado a seguir, com a fundamentação teórica para a sua aplicação ao problema em questão. O processo de construção de modelo está subdividido em seis passos.

Passo 1: O Problema de Pesquisa.

São os seguintes os objetivos da análise discriminante múltipla:

- investigar diferenças entre grupos de observações. O pesquisador deve utilizar a análise discriminante quanto estiver interessado em explicar diferenças entre grupos de observações de um fenômeno ou classificar observações em grupos ou classes distintos. No contexto deste trabalho o

⁴⁰ As transformações processadas serão apresentadas na próxima seção, durante a construção do modelo discriminante.

que se pretendeu foi explicar a decisão de emissão pública ou privada segundo diferenças nas características apresentadas pelas variáveis observadas na amostra.

- determinar a forma mais parcimoniosa possível de distinção entre os grupos;
- eliminar as variáveis que não contribuam para a distinção entre os grupos;
- testar o modelo obtido comparando as observações empíricas às previsões.

Segundo Hair et. ali (1998), a técnica melhor se enquadra em situações onde há uma variável categórica dependente e várias variáveis métricas independentes. Para compreender diferenças entre grupos, o método permite apreender o papel de cada variável, bem como as combina de modo a obter a discriminação entre grupos. No capítulo de apresentação dos resultados serão relatadas as variáveis descartadas e a equação discriminante resultante.

A análise discriminante permite realizar classificações de observações empíricas que tenham valores para todas as variáveis independentes no modelo. Para efeitos de previsão, pode-se, dessa forma, inferir o comportamento futuro de empresas quanto ao processo decisório relativo às emissões públicas ou privadas, a partir de algumas de suas características (representadas nas variáveis independentes).

Passo 2: Elaboração da pesquisa – definição das variáveis, amostra e sub amostras.

A necessidade de utilização da análise discriminante múltipla (ADM)⁴¹ vem do fato de haver dicotomia na variável dependente do problema em análise. De fato, as emissões podem ser públicas ou privadas, assumindo, respectivamente, a variável dependente, os valores 0 ou 1. Esse tipo de variável tem a denominação

⁴¹ Traduzido Inglês Multiple Discriminant Analysis (MDA).

de categórica, já que apresenta duas categorias de emissões – pública e privada. Dessa forma, o que se almeja é classificar quaisquer emissões dentro da amostra estudada em uma das duas categorias: pública ou privada.

As variáveis independentes foram escolhidas da forma apresentada na seção 5.4.1, a partir da consulta à literatura existente e da disponibilidade de dados para levantamento. Houve um tratamento prévio no sentido de reduzir os problemas de escala, já que havia números muito diferentes para serem tratados no mesmo modelo. Além disso, para atender às premissas de normalidade e linearidade de relações do modelo, houve a necessidade de algumas transformações⁴². Sendo assim, a seguir serão apresentadas as novas variáveis utilizadas no modelo discriminante, em substituição às originalmente levantadas:

Lucr. Liq Us\$1000 (Log10) (loglucro) – logaritmo decimal do lucro líquido anual medido em milhares de dólares americanos da empresa emissora no exercício anterior à emissão.

Valor emit R\$ (Log 10) (vlremit) – logaritmo decimal do valor emitido em reais.

Patrim liq Us\$1000 (Log 10) – logaritmo decimal do patrimônio líquido medido em milhares de dólares americanos no exercício anterior à emissão.

Liquid nº neg (Ln)⁴³ – logaritmo neperiano do número de negócios realizados com as ações da empresa emissora no ano da emissão.

Liquid Vol neg (Log 10) – logaritmo decimal do volume negociado medido em reais no ano da emissão

⁴² De fato, as transformações realizadas melhoraram os *plots Normal Q-Q*, trazendo as variáveis para níveis aceitáveis de normalidade. Porém, por se reconhecer uma certa fragilidade na premissa de normalidade, optou-se por também construir o modelo logístico, menos restritivo quanto à normalidade.

⁴³ O tipo de logaritmo utilizado (Ln ou Log10) se deveu a melhoria na capacidade preditiva do modelo, após algumas simulações.

Exigível 1 prz Us\$ (Ln) - logaritmo neperiano do valor total da dívida de longo prazo, medida em dólares americanos no exercício anterior à emissão

Liquidez vol/nº negócios (Ln) – logaritmo neperiano da razão entre o volume negociado e o número de negócios realizados com as ações da empresa no ano da emissão⁴⁴

A amostra foi determinada da maneira descrita na seção 5.2. A utilização de sub-amostras para validação vai ser apresentada junto com os resultados no próximo capítulo.

Passo 3: Premissas do modelo discriminante.

A análise discriminante, segundo Hair(1998), envolve a busca de uma combinação linear de duas ou mais variáveis independentes que melhor farão a classificação entre grupos previamente definidos. A discriminação será obtida a partir da determinação dos fatores ou pesos para cada variável de modo que a variância entre os grupos seja maximizada em relação à variância intra-grupos. Este é um processo de classificação em que se busca dois grupos de características homogêneas de forma a distingui-los.

A função discriminante pode ser representada pela seguinte equação⁴⁵:

$$Z_{jk} = b + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk}$$

onde

Z_{jk} = valor discriminante Z da função discriminante j para a observação k .

b = intercepto

W_i = peso discriminante para a variável independente i .

X_{ik} = variável independente i para a observação k .

⁴⁴ Variável definida para dar maior estabilidade ao modelo. É o negócio médio realizado. No caso de ausência de liquidez, zero de volume e de negócios, atribuiu-se à variável o valor zero.

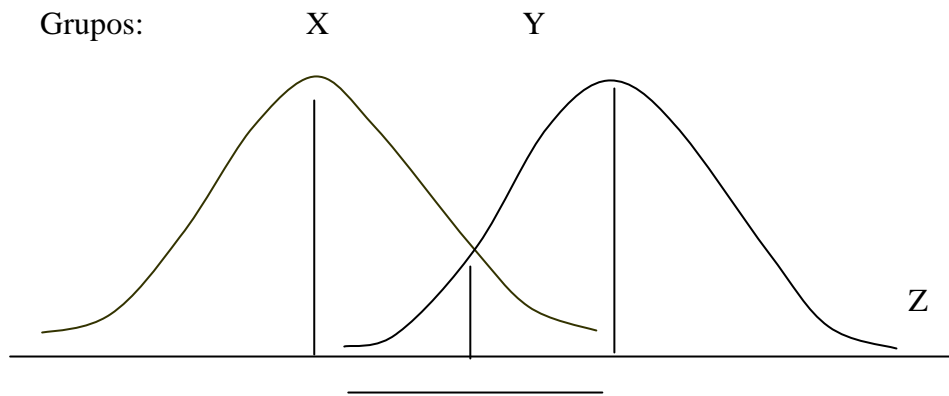
⁴⁵ A partir de Hair (1998) pp. 244.

O processo da análise discriminante consiste na multiplicação de cada variável independente por seu peso correspondente e a soma de todos os produtos obtidos a cada observação. O resultado é o valor Z para cada observação ou indivíduo na análise. A partir da média dos valores Z para todas as observações em um determinado grupo, obtém-se o valor Z para o grupo, também chamado de centróide. No caso deste trabalho serão obtidos dois centróides.

Para que se possa obter uma explicação sobre o comportamento das emissões na amostra, os centróides obtidos devem ter significância estatística. Deve-se obter a distribuição dos valores Z para cada um dos grupos. Sendo assim, duas distribuições surgirão, no caso de dois grupos. Se a interseção entre as distribuições for pequena, pode-se afirmar que houve discriminação melhor. Se for grande a superposição entre as distribuições, a discriminação será mais fraca, e os prognósticos de classificação menos confiáveis.

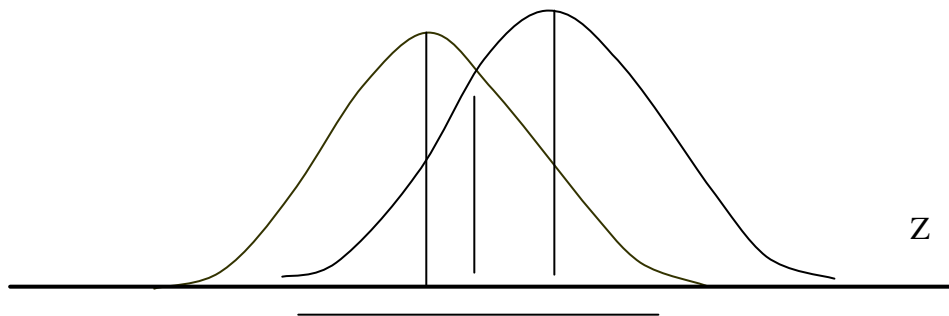
A Figura 7, a seguir, ilustra a situação:

Figura 7 – Distribuição univariada dos fatores de classificação Z



Função Discriminante Boa classificação – superposição pequena

Grupos: X Y



Função Discriminante Má classificação: grande superposição

A análise discriminante multivariada para mais de dois grupos apresenta $n-1$ funções discriminantes, onde n é o número de grupos. No caso em questão apenas uma função surgirá do processamento dos dados, já que existem somente dois grupos.

São premissas para a utilização do modelo discriminante:

- normalidade das distribuições de frequência das variáveis independentes;
- linearidade das relações;
- ausência de multicolinearidade das variáveis independentes (alta correlação entre duas ou mais variáveis);

- matrizes de dispersão iguais (homocedasticidade) – ausência de diferenças significativas das matrizes de variância e covariância dos resíduos para cada grupo classificado.

As transformações realizadas nas variáveis visaram atender às exigências do modelo. De fato entre as variáveis originais e as transformadas, utilizadas efetivamente na construção do modelo, chegou-se ao melhor conjunto possível. Optou-se por não manipular as variáveis além do já realizado de forma a não comprometer os resultados do modelo. A homocedasticidade não foi atingida, o que está explicado na apresentação dos resultados.

No próximo capítulo serão apresentadas e analisadas as estatísticas descritivas.

A seguir será apresentado o próximo passo, que trata da estimação do modelo discriminante.

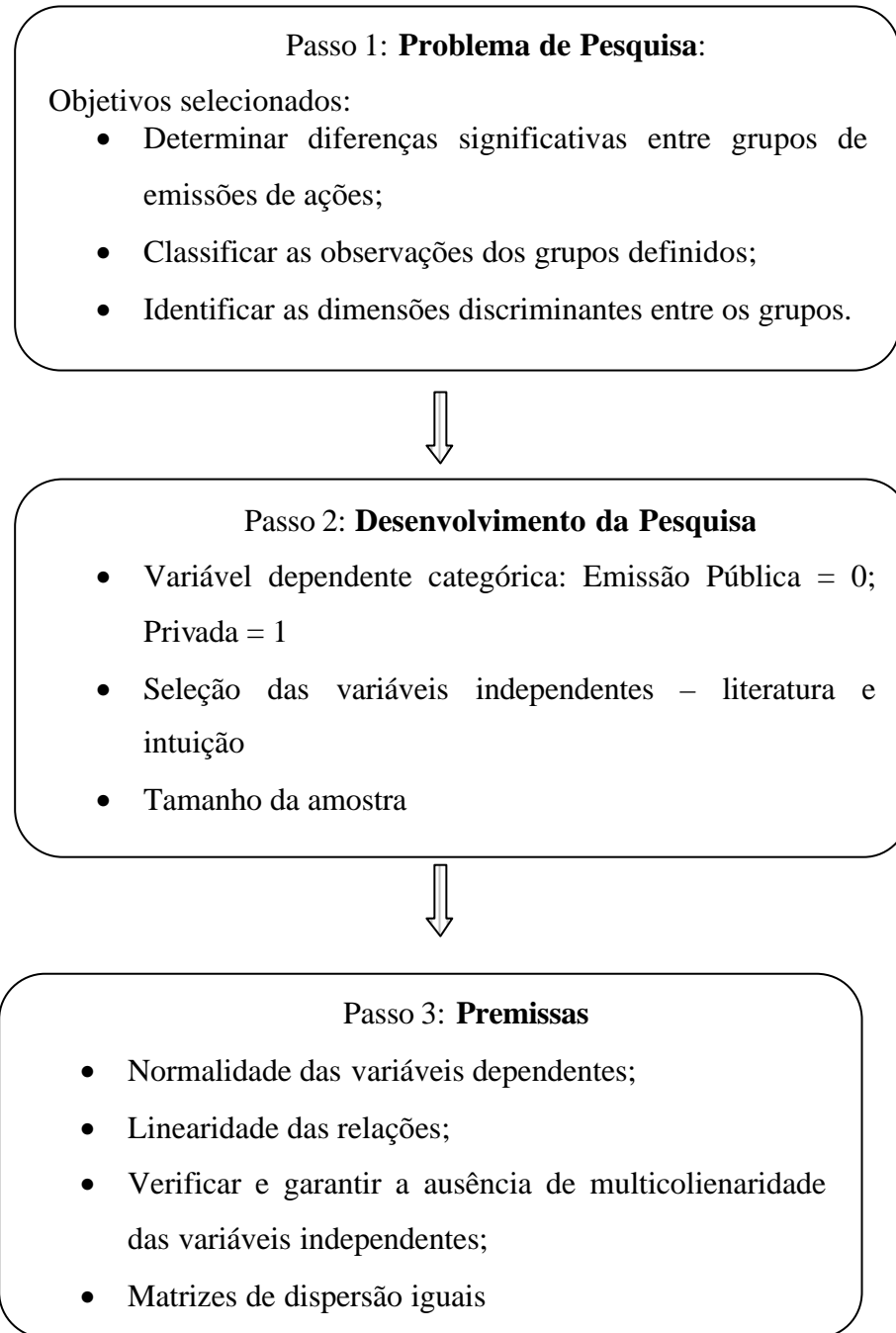
Passo 4: Estimação do modelo discriminante:

As variáveis foram introduzidas passo a passo, pelo método *stepwise* do programa estatístico SPSS. Assim, cada uma foi introduzida no modelo segundo seu poder discriminante. O processo se inicia com a escolha da melhor variável sob o ponto de vista da classificação categórica. A seguir, a próxima variável é escolhida dentre as outras de modo que, em conjunto com a primeira, ofereça a melhor capacidade discriminatória. O processo segue para todas as outras variáveis da mesma forma. Segundo Hair(1998), este processo se aplica melhor quando o número de variáveis é maior. À medida que são incluídas, uma por vez, as próximas variáveis com maior poder explanatório, vai se construindo um modelo robusto e parcimonioso.

As variáveis escolhidas, assim como o modelo obtido, serão apresentados no próximo capítulo, junto com os resultados.

Os passos 5 e 6, respectivamente, interpretação e validação dos resultados da análise discriminante serão apresentados junto com os resultados no próximo capítulo. A seguir, será apresentado um fluxograma do processo de formulação de modelo discriminante.

Figura 8 – Fluxograma teórico da análise discriminante empregada, baseado em Hair (1998)



Passo 4 a: Estimação do Modelo Discriminante

- Entrada de variáveis passo a passo ou em bloco



Passo 4b: Obter precisão preditiva através de Matrizes de Classificação

- Determinar o ponto de corte ótimo;
- Especificar o critério para obtenção da aderência do modelo;
- Determinar o nível de significância estatística da predição.



Passo 5: Interpretação da Função Discriminante

- Há apenas uma função a ser interpretada (dois grupos a serem discriminados)



Passo 5 a: Avaliação da Função Discriminante

- Pesos (coeficientes);
- Correlações estruturais;
- Valores F



Passo 6: Validação dos resultados discriminantes

- Sub amostras;
- Apresentação dos diferentes perfis de cada grupo.

A classificação dos grupos via análise discriminante possibilitará, conforme os resultados encontrados e apresentados a seguir, uma explicação do comportamento das emissões a partir das variáveis independentes. No entanto, para que este trabalho atinja o objetivo de obter uma explicação ampla sobre a decisão de emitir ações pública ou privadamente, procedeu-se também uma regressão logística com o mesmo conjunto de dados e variáveis. A coerência de resultados dos dois modelos embasará conclusões consistentes, e reduzirá os problemas relativos às limitações do modelo discriminante (exigência de normalidade e homocedasticidade).

A regressão logística pode ser interpretada, nesse contexto, como uma forma de validar os resultados obtidos com a análise discriminante. Similar à análise discriminante, o modelo de regressão logística apresenta algumas diferenças, ressaltadas na subseção seguinte.

5.5.3

O Modelo de Regressão Logística

O modelo de regressão logística atende aos casos em que a variável dependente é qualitativa em dois níveis (0,1). Neste trabalho, a decisão entre emitir ações de forma pública ou privada é a variável dependente qualitativa. Segundo Sincich(1995), o modelo logístico pode ser apresentado da seguinte forma:

$$E(y) = \frac{\exp(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{b}_2x_2 + \dots + \mathbf{b}_kx_k)}{1 + \exp(\mathbf{b}_0 + \mathbf{b}_1x_1 + \mathbf{b}_2x_2 + \dots + \mathbf{b}_kx_k)} \quad (4.5.3.1)$$

onde

$y = 1$ se emissão é privada;

$y = 0$ se emissão é pública;

$E(y)$ = Probabilidade (Categoria 1 ocorrer);

x_1, x_2, \dots, x_k são variáveis independentes quantitativas ou qualitativas

Pode-se, a partir do modelo logístico, estimar a chance de ocorrência (*odds*) do evento $y=1$, que é a razão entre as probabilidades de ocorrência do evento $y=1$:

$$\frac{P\{y = 1\}}{P\{y \neq 1\}} = \frac{P\{y = 1\}}{P\{y = 0\}} = \frac{P\{y = 1\}}{[1 - P\{y = 1\}]} \quad (4.5.3.2)$$

O logaritmo neperiano da chance, é, segundo Menard (1995), o logit de y :

$$\log \text{it}(Y) = \mathbf{b}_0 + \mathbf{b}_1 x_1 + \mathbf{b}_2 x_2 + \dots + \mathbf{b}_k x_k \quad (4.5.3.3)$$

Pode-se obter a chance(*odds*) novamente através da exponenciação, ou seja, pelo cálculo de:

$$\text{Chance } (Y = 1) = e^{\log \text{it}(y)} = e^{\mathbf{b}_0 + \mathbf{b}_1 x_1 + \mathbf{b}_2 x_2 + \dots + \mathbf{b}_k x_k} \quad (4.5.3.4)$$

Pode-se, então retornar à probabilidade de $Y=1$, através de:

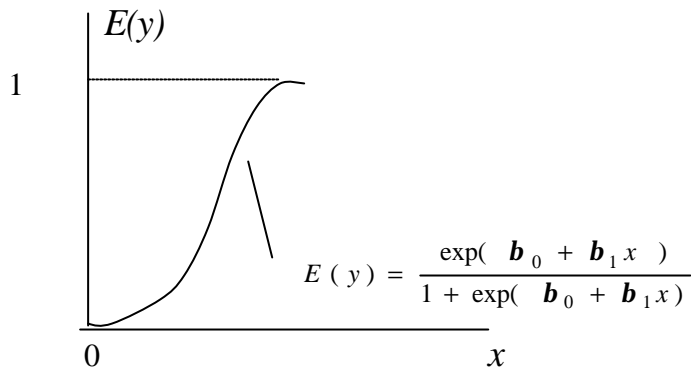
$$P(y = 1) = \frac{\text{chance } (y = 1)}{1 + \text{chance } (y = 1)} \quad (4.5.3.5), \text{ que produz novamente}$$

a equação (4.5.3.1).

A Figura 9 apresenta o gráfico do modelo logístico para uma única variável independente⁴⁶:

⁴⁶ Sincich (1995) pp1040

Figura 9 – Gráfico de Valor esperado de Y para o Modelo Logístico



Conforme ressalta Menard, (1995), a função logit, a chance e a probabilidade, acima apresentadas, são diferentes maneiras de expressar a mesma coisa.

A regressão logística não depende tanto quanto a análise discriminante da igualdade de variância e covariância entre os grupos, conforme Hair (1998)⁴⁷. Um exemplo disso é poder utilizar sem problemas variáveis independentes categóricas, que acarretam problemas de igualdade de variância e covariância. Além disso, existe a vantagem adicional da possibilidade de analogia com a regressão múltipla, tanto no exame dos resíduos como na interpretação dos resultados.

A primeira diferença importante está no estabelecimento de categorias para as variáveis independentes, de modo a que se tenha apenas variáveis categóricas no modelo. Isso se mostrou necessário devido à ocorrência de emissões muito pequenas, e de emissões muito grandes na mesma amostra. Além disso, variáveis como liquidez, lucratividade, endividamento e concentração de propriedade também apresentavam essas diferenças. Este procedimento pode reduzir significativamente a influência de comportamentos não lineares das variáveis. Dessa forma, esse procedimento tornará mais robustas os resultados da regressão logística.

⁴⁷ Por essa razão, optou-se por também construir o modelo logístico. A robustez dos resultados conjuntos das estatísticas descritivas, do modelo discriminante e do modelo logístico confere maior consistência às conclusões deste trabalho.

As variáveis foram categorizadas com igual percentual de observações, sempre que possível, sendo a primeira categoria a que concentra os menores valores e a última a que concentra os maiores valores. A variável patrimônio líquido foi excluída da análise por estar presente na fórmula da lucratividade ($\text{lucro líquido/patrimônio líquido}$) e endividamento ($\text{exigível de longo prazo/patrimônio líquido}$). Esse procedimento não comprometerá a análise, pois a variável valor emitido é, também, uma medida de porte da empresa e está incluída.

Estas categorias foram o ponto de partida para as faixas da regressão logística. A Tabela 12 apresenta os resultados da categorização feita no SPSS.

Tabela 12 – Categorização das Variáveis para a Regressão Logística

Variáveis e Categorias	N		Variáveis e Categorias	N	
VALOR_EM			Perc3maior		
1	101	Até 3.500.000	1	93	Até 60%
2	120	De 3.500.000 a 14.000.000	2	345	De 60% a 99%
3	139	De 14.000.000 a 44.000.000	3	191	Acima de 99%
4	137	De 44.000.000 a 120.000.000	Total	629	
5	132	Acima de 120.000.000			
Total	629				
LIQVOL/nº neg LN			ENDIVID		
1	192	Zero	1	135	zero
2	69	De 1 a 2K	2	142	Até 18%
3	99	De 2K a 10K	3	137	De 19% a 46%
4	133	De 10K a 30K	4	141	De 47% a 131%
5	136	Acima de 30K	5	74	Maior que 131%
Total	629		Total	629	
LUCRO LÍQ.			LUCRATIV		
1	114	Até -12.200	1	99	-25% ou menos
2	104	Entre -12.200 e -200	2	269	De -25% a 3%
3	142	Entre -200 e 2.350	3	141	De 3% a 10%
4	145	Entre 2.350 e 25.800	4	120	Acima de 10%
5	124	Acima de 25.800	Total	629	
Total	629				

A Tabela 13 apresenta os testes de linearidade para a relação entre as variável tipo de emissão (dependente) e as variáveis independentes.

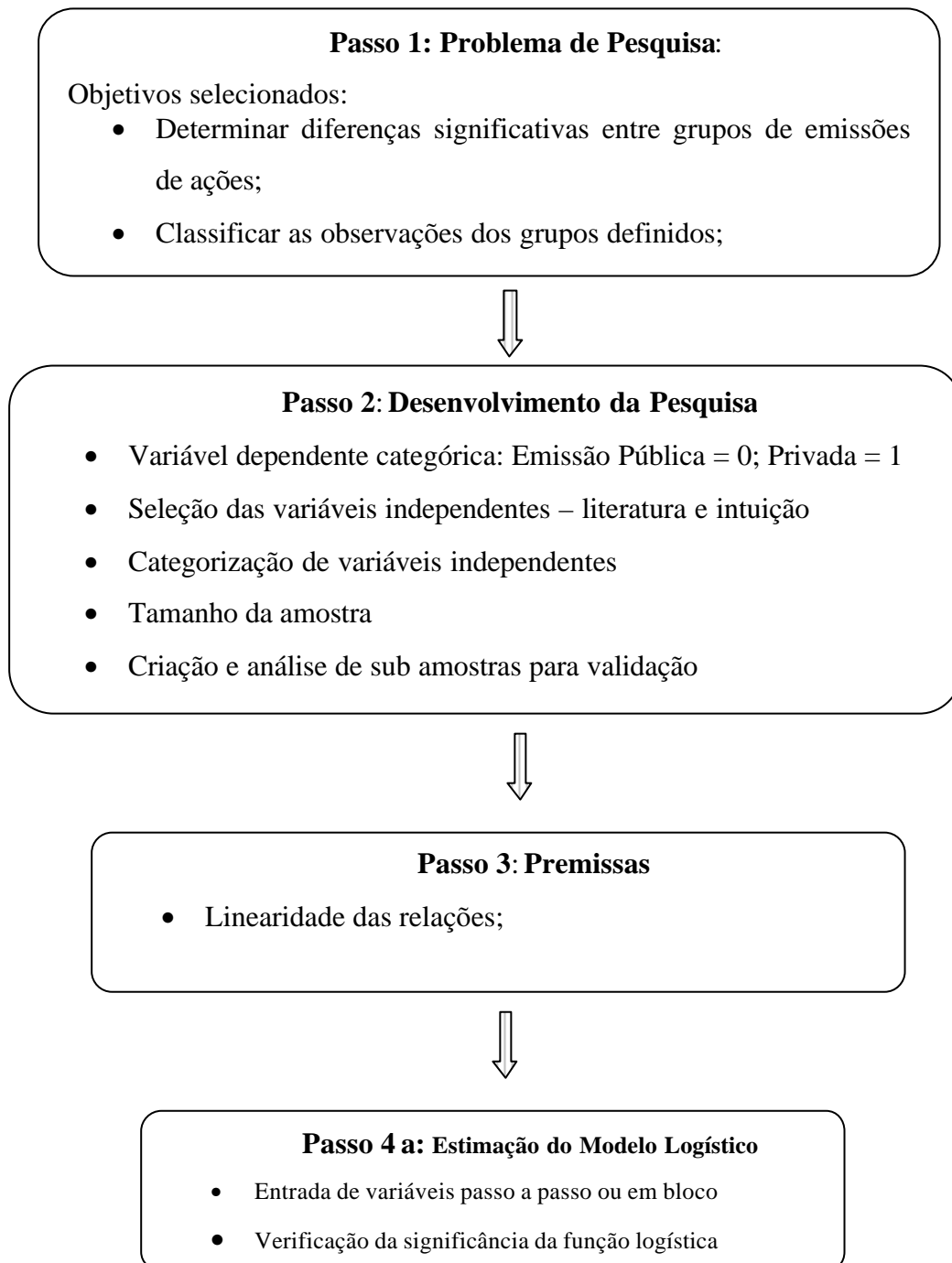
Tabela 13 – Testes de Linearidade para as relações entre a variável dependente e as independentes

Variáveis e testes		Erro padrão		Significância
VALOR_EM	Valor	assintótico	T aproximado	aproximada
Gamma	-,360	,051	-6,865	,000
Correlação de Spearman	-,258	,038	-6,700	,000
Perc3maior				
Gamma	,122	,070	1,735	0,083
Correlação de Spearman	,069	,040	1,731	,084
Liqvol/Nºneg LN				
Gamma	-,257	,055	-4,587	,000
Correlação de Spearman	-,181	,039	-4,589	,000
Endiv ELP/PL				
Gamma	,018	,057	,311	,756
Correlação de Spearman	,012	,040	,311	,756
Lucro Líq. Us\$1000				
Gamma	-,213	,055	-3,855	,000
Correlação de Spearman	-,152	,039	-3,844	,000
LUCRATIV.LL/PL				
Gamma	-,293	,058	-4,933	,000
Correlação de Spearman	-,191	,039	-4,862	,000
TIPOEMP 0				
Gamma	,671	,104	4,795	,000
Correlação de Spearman	,188	,034	4,783	,000
TIPOEMP 1				
Gamma	-,306	,077	-3,780	,000
Correlação de Spearman	-,149	,039	-3,773	,000

Para o teste de Correlação de *Spearman*, se o nível de significância for menor que 0,05, então a correlação é significativa e as duas variáveis são linearmente relacionadas. No caso da variável de concentração de controle (perc3maior), a significância apresentada é 0,084, o que dá um intervalo de

confiança um pouco menor para a linearidade, mas que foi aceito para a construção do modelo logístico. Para a variável endividamento (ELP/PL) o valor da significância é de 0,756, o que dá um intervalo de confiança muito pequeno para a linearidade. Essa variável será, portanto, desconsiderada no modelo logístico.

A premissa de linearidade das relações foi testada com sucesso, exceto para a variável relativa ao endividamento. A seguir, será apresentado o fluxograma do processo de formulação do modelo de regressão logística. Note-se que os primeiros três passos são muito similares aos da análise discriminante, e já foram apresentados nesta seção. Os passos 4 a 6 serão apresentados junto com os resultados no próximo capítulo.

Figura 10 – Fluxograma teórico da regressão logística⁴⁸

⁴⁸ baseado em Hair et. ali., 1998, pp. 314.

Passo 4b: Obter precisão preditiva através de Matrizes de Classificação

- Determinar o ponto de corte ótimo;
- Especificar o critério para obtenção da aderência do modelo;
- Determinar o nível de significância estatística da predição.
- Obtenção do estatístico Wald para aferir a significância



Passo 5: Interpretação dos resultados da Função Logística

- Há uma equação similar à função discriminante binária.
- Verifica-se a coerência dos resultados das duas análises



Passo 6: Validação dos resultados

- Sub amostras: duas amostras aleatórias de 50% são criadas para validação do modelo
- Apresentação de *outliers* para posterior explicação
- Verificação da capacidade preditiva obtida nas matrizes de classificação com e sem ponderação (o número de emissões particulares da amostra é muito maior que o das públicas)
- Comparação com a matriz de classificação da análise discriminante.

Como se pode notar, o modelo logístico é bastante similar ao discriminante. Trata-se de duas técnicas adequadas à classificação das emissões em públicas ou privadas segundo as características das empresas emissoras. O objetivo, como foi salientado antes, é dar consistência aos resultados obtidos com a análise discriminante. A seguir, serão apresentadas as limitações dos métodos utilizados.

5.6

Limitações do método

Dados:

O levantamento sistemático oficial, através da Comissão de Valores Mobiliários e das Bolsas de Valores, sobre as emissões particulares não existe. Foi necessária ampla pesquisa em documentos (informativos do mercado, jornais especializados, entre outros) de modo a estabelecer com confiança uma base empírica e o período de análise. Nesse sentido, os de levantamentos para os anos anteriores a 1995 poderiam ser úteis para uma análise comparativa entre os períodos pré e pós Plano Real. No entanto, esse levantamento se provou inviável, já que não se pode encontrar os dados completos das emissões nem os relativos às características das empresas emissoras. A falta de dados contábeis, tais como endividamento, lucratividade, causaria a eliminação pelo, modelo, de muitas observações. Técnicas de substituição de valores ausentes por médias ou outro valor poderiam comprometer a robustez das conclusões. De qualquer forma, a extensão para períodos anteriores a 1995 fica como sugestão para futuras pesquisas.

A utilização de dados referidos em dólares americanos, ocorrida em algumas variáveis, torna relevantes as datas e a forma da conversão. Essa limitação se deve à instabilidade da moeda nacional, em especial, antes do Plano Real. Ocorre que, mesmo após o Plano, várias fontes de dados continuaram a utilizar os dólares americanos como moeda para a denominação de variáveis contábeis. Isso pode acarretar, de algum modo, distorções. É uma limitação presente em trabalhos realizados no Brasil.

Modelo Teórico:

A questão da maximização do valor para os acionistas controladores, da forma colocada por Ferreira e Brooks (2000) pressupõe uma racionalidade que pode não estar presente no mercado de capitais brasileiro. A noção de maximização de valor para os controladores pode não ser perseguida como em um

mercado mais desenvolvido. Essa limitação, entretanto, não invalida o modelo teórico, já que este constitui um quadro conceitual aceito mundialmente. De fato, as conclusões do trabalho, poderão esclarecer sua adequação ao mercado brasileiro.

Pensou-se na construção de um modelo decisório do tipo árvore de decisão para a emissão privada ou pública de ações. Embora a base de dados seja ampla e as variáveis obtidas sejam explicativas do fenômeno, a relação de causalidade necessária para a obtenção de uma árvore de decisão é mais complexa do que o objetivo deste trabalho. Fica como sugestão para futuras pesquisas, a partir dos resultados aqui apresentados, a busca da relação de causalidade necessária para a construção de um modelo decisório mais amplo.

As variáveis relativas à lucratividade, endividamento, e patrimônio líquido, ficaram limitadas aos dados contábeis disponíveis. Dessa forma, ficaram de fora variáveis que seriam interessantes para dar maior robustez à análise. Pode-se citar variável ligada ao nível de preço das ações emitidas, em relação à média histórica, e o nível do IBOVESPA antes da emissão. Essas variáveis poderiam indicar o oportunismo das emissões públicas e/ou privadas, enriquecendo a análise. Entretanto, a liquidez muito baixa de grande parte das empresas da amostra tornaria essa análise frágil. Isso deveria à impossibilidade de cálculo confiável dos retornos obtidos pelos investidores nas emissões: a pouca liquidez tornaria o cálculo muito impreciso ou mesmo inviável. Fica, porém, como sugestão para futuros trabalhos.

Na medida de porte das empresas utilizou-se como *proxy* o patrimônio líquido e o valor emitido. Isso excluiu, ativo total, faturamento, geração de caixa, que poderiam ser variáveis relevantes. A escolha se deveu à maior disponibilidade do patrimônio líquido nas fontes pesquisadas. Uma outra limitação, é o fato do lucro líquido poder ser, também, considerado uma medida de porte. Nesse sentido, os modelos discriminante e logístico providenciam a eliminação de variáveis correlacionadas, mantendo apenas as que de fato têm poder classificatório.

A questão do risco percebido pelo investidor poderia ter sido abordada através de uma classificação das empresas por setores da economia. Além disso, seria interessante conhecer os tipos de investidor que adquiriram as emissões

públicas e privadas. Fica como sugestão para futuros trabalhos testar o processo decisório por setor e por tipo de investidor.

Os custos de emissão não se apresentam iguais para todas as empresas. De fato, empresas maiores, com maior prestígio, tendem a ter custos menores. Seria interessante abordá-los como uma variável, já que são relevantes dentro do quadro conceitual proposto. Entretanto, o levantamento dos custos das emissões públicas apenas para os últimos três anos do estudo, apresentado na Tabela 9, provou-se particularmente trabalhoso. Além da dificuldade de obtenção dos prospectos em anos de maior movimento de emissões públicas, o acesso aos custos das emissões particulares somente seria possível através de entrevistas com pessoas ligadas às empresas emissoras. Naturalmente, muitas empresas da amostra não mais existem, tornando impossível a obtenção de todos os dados necessários.

Modelos Empíricos:

A falta de normalidade nas observações das variáveis independentes prejudica a análise das estatísticas descritivas e da análise discriminante. Para contornar isso, foi desenvolvido o modelo logístico, menos restritivo quanto às premissas de normalidade e homocedasticidade da análise discriminante. Suas conclusões, em conjunto com as dos outros dois procedimentos, darão maior consistência a este trabalho. Há que se ressaltar que o modelo discriminante tem sua robustez ligada à capacidade de classificação. Um baixo percentual (inferior a 60%) indicaria fragilidade no modelo.

Além disso, variáveis foram manipuladas com transformações logarítmicas e razões para que o seu comportamento possibilitasse uma melhor classificação dos grupos. Procurou-se realizar as transformações com parcimônia de forma a não limitar demasiadamente o poder classificatório dos modelos. Nesse sentido, vários modelos alternativos, com variáveis transformadas de diversas formas foram testados, optando-se pelos de maior poder explanatório.

Os modelos discriminante e de regressão logística empregados partem de premissas quanto ao comportamento das variáveis. No capítulo de resultados serão abordadas as suas limitações e os procedimentos (além dos já relatados) para sua correção.

Neste capítulo foi apresentada a metodologia utilizada na investigação. Foram explicados o tipo de pesquisa, o universo e a seleção da amostra assim como a coleta dos dados e seu tratamento. Finalmente, foram apresentadas as limitações do método e a justificativa para sua utilização. O próximo capítulo fará a apresentação e a análise dos resultados.