

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS SOCIAIS E HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

**CLASSIFICAÇÃO SETORIAL DAS EMPRESAS
BRASILEIRAS:
UMA APLICAÇÃO DA ANÁLISE DE *CLUSTER***

DISSERTAÇÃO DE MESTRADO

Vanderléia Leal Losekann

Santa Maria, RS, Brasil

2008

**CLASSIFICAÇÃO SETORIAL DAS EMPRESAS
BRASILEIRAS:
UMA APLICAÇÃO DA ANÁLISE DE *CLUSTER***

por

Vanderléia Leal Losekann

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Administração, Área de Concentração em Estratégia e Competitividade, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Administração.**

Orientadora: Prof^a. Dr^a Kelmara Mendes Vieira

Santa Maria, RS, Brasil

2008

**UNIVERSIDADE FEDERAL DE SANTA MARIA CENTRO DE
CIÊNCIAS SOCIAIS E HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**CLASSIFICAÇÃO SETORIAL DAS EMPRESAS
BRASILEIRAS:
UMA APLICAÇÃO DA ANÁLISE DE *CLUSTER***

elaborada por
Vanderléia Leal Losekann

**COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE
Mestre em Administração**

COMISSÃO EXAMINADORA:

Kelmara Mendes Vieira, Dr^a.
(Presidente/Orientador)

Paulo Sérgio Ceretta, Dr. (UFSM)

Roselaine Ruviaro Zanini, Dr^a. (UFSM)

Santa Maria, 17 de dezembro de 2008.

Ao meu filho Vinicius,
Ao meu marido Milton,
Aos meus pais - Abigail e Vanderli (in memoriam)
A minha irmã Vânia,
Aos meus colegas técnicos administrativos da UFSM
Dedico este trabalho.

AGRADECIMENTOS

Uma dissertação não é um trabalho individual. É algo que só se conclui com a contribuição de muitas outras pessoas. Lembrar de todas, no fim da jornada, nem sempre é tarefa fácil, mas é justo que se façam sinceros agradecimentos:

- a Prof^ª. Kelmara Mendes Vieira, minha orientadora, que foi mestra e amiga, dedicando seu tempo e atenção durante todo este caminho de construção desta dissertação, a quem eu agradeço de todo coração pela confiança depositada e pelo carinho; e acima de tudo, agradeço essa grande mestra, pois aprendi que o “ensinar” não é univariado, pelo contrário, é multivariado e estabelece-se de várias formas quando se conjugam comportamentos;

- a minha mãe e ao meu marido, que me deram todo suporte possível para que eu pudesse alcançar mais este objetivo;

- ao Professor Breno Diniz Pereira, pelas valiosas contribuições a este trabalho e em especial quando coordenador, por ter acolhido de forma generosa e persistente todo o processo da abertura de vagas para os técnicos administrativos neste programa de pós-graduação. Obrigada por concretizar a “**1ª turma de técnicos administrativos**”;

- ao Professor Paulo Sérgio Ceretta, por ter colocado sua inteligência, conhecimento e boa vontade à disposição de minha formação na área de finanças;

- ao Magnífico Reitor Professor Clóvis Lima, Diretor do Centro de Ciências Sociais e Humanas Professor Rogério Koff, ao Pró-Reitor de Pós-Graduação Professor Helio Leães Hey, ao Pró-Reitor de Recursos Humanos Téc. Adm. João Pillar Pacheco de Campos e ao Chefe do laboratório de Informática/CCSH Téc. Adm. Nilson Amaury Siqueira, por esta valiosa oportunidade;

- a Maria Medianeira, minha querida amiga e colega, secretária do curso de pós-graduação de administração e aos meus queridos colegas de trabalho do curso de Direito, Maria Ester, Lilia, Beloni e Claiton;

- aos demais professores do programa de pós-graduação;

Agradeço em especial aos meus colegas de mestrado, Marlon, Diego, Ney, Alexandre, Lisi, Kati e Débora pelo companheirismo e amizade. E, em especial a Dolores e a Ana.

As pedras do caminho são árduas, mas a chegada é magnífica.

A todos, muito obrigada!

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Administração
Universidade Federal de Santa Maria

CLASSIFICAÇÃO SETORIAL DAS EMPRESAS BRASILEIRAS: UMA APLICAÇÃO DA ANÁLISE DE *CLUSTER*

Autora: Vanderléia Leal Losekann

Orientadora: Kelmara Mendes Vieira

Data e Local da Defesa: Santa Maria, 17 de dezembro de 2008.

O processo de tomada de decisão numa empresa requer conhecimento externo do ambiente em que ela atua. O agrupamento de empresas, de acordo com o ramo de atividade ou com o comportamento de certas variáveis independentemente do tipo de produto ou serviço oferecido, serve como base para a compreensão das características e relações das empresas nos mercados em que vendem ou compram. Esta dissertação busca avaliar a eficiência da classificação setorial a partir do método análise de *cluster*. A utilização da classificação setorial nos estudos na área de administração tem, por base, o pressuposto de que empresas de um mesmo setor são semelhantes. Portanto, se a classificação setorial é eficiente, espera-se que empresas de um mesmo setor formem um *cluster* e que empresas de setores diferentes sejam agrupadas em *clusters* díspares. Os resultados, de um modo geral, demonstram uma alta heterogeneidade setorial nos *clusters* quer quando o número de *clusters* é sugerido pela regra de parada ou critério BOVESPA. E esta heterogeneidade é estável, uma vez que se manteve ao longo do período estudado. Portanto, os resultados apontam para a ineficiência da análise setorial. O estudo do agrupamento de empresas a partir de dados econômicos e financeiros serviu para identificar a não existência de homogeneidade dos setores do ponto de vista dos indicadores quando sujeitos a uma classificação baseada unicamente no tipo de bem ou serviço produzido. Para tomadas de decisões mais precisas os setores podem ser reestruturados de modo a incorporar, em cada agrupamento, empresas que possam ser razoavelmente semelhantes.

Palavras-chave: agrupamento de empresas; classificação setorial; análise de *cluster*

ABSTRACT

The decision-making in a company requires knowledge of the external environment in which it operates. The group of companies, according to an activity or with the behavior of certain variables regardless of the type of product or service, serves as a basis for understanding the relationships and characteristics of businesses in markets that sell or buy. This thesis seeks to evaluate the efficiency of the sector classification from the *cluster* analysis method. The use of rating sector in studies in the administration area has been based on the assumption that companies within the same industry are similar. So if the classification sector is efficient, it is expected that companies within the same industry form a *cluster* of industries and companies that are grouped into different *clusters* mixed. The results, in general, show a high heterogeneity in sectoral *clusters* or when the number of *clusters* is suggested by the rule of stopping or BOVESPA criterion. And this heterogeneity is stable, as it remained throughout the study period. Therefore, the results point to the inefficiency of sectoral analysis. The study group of companies from financial and economic data served to identify the lack of homogeneity of the sectors in terms of the indicators when subjected to a classification based solely on the type of goods or services produced. For more precise decisions taken in those sectors can be restructured to incorporate in each group, companies that can be reasonably similar.

Keywords: group of companies; sectoral classification; cluster analysis

LISTA DE TABELAS

TABELA 01 - Quantidade de percentual médio, mediana, desvio padrão de dados omissos para cada variável, no período de 2000-2007	74
TABELA 02 - Quantidade de <i>outliers</i> para os <i>clusters</i> índices e mercados, no período de 2000-2007.....	75
TABELA 03 - Classificação setorial segundo a BOVESPA para cada ano das ações em negociação no período de 2000-2007.....	76
TABELA 04 - Classificação segundo a BOVESPA considerando todas as ações em negociação no período de 2000–2007.....	77
TABELA 05 - Classificação setorial, segundo a NAICS para cada ano das ações em negociação, no período de 2000-2007.....	77
TABELA 06 - Classificação segundo a NAICS considerando todas as ações em negociação no período de 2000–2007.....	78
TABELA 07 - Número de casos válidos, valores mínimo e máximo, média, mediana, desvio-padrão, assimetria e curtose das variáveis médias, no período de 2000-2007.....	80
TABELA 08 - Teste Kruskal Wallis para as variáveis médias segundo as classificações setoriais da NAICS e BOVESPA	81
TABELA 09 - Frequência percentual e total de casos, segundo a classificação BOVESPA, no ano 2000.....	86
TABELA10 - Classificação BOVESPA, com 4 <i>clusters</i> para <i>cluster</i> índice e 5 <i>clusters</i> para <i>cluster</i> mercado, no ano 2001.....	87
TABELA 11 - Classificação BOVESPA, com 5 <i>clusters</i> para o <i>cluster</i> índice e 2 <i>clusters</i> para o <i>cluster</i> mercado, no ano 2002	88
TABELA 12 - Classificação BOVESPA, com 4 <i>clusters</i> para o <i>cluster</i> índice e 9 <i>clusters</i> para <i>cluster</i> mercado no ano 2003	88

TABELA 13 - Classificação BOVESPA, num total de 4 <i>clusters</i> para o <i>cluster</i> índice e 4 <i>clusters</i> para o <i>cluster</i> mercado, no ano 2004	89
TABELA 14- Classificação BOVESPA, com 6 <i>clusters</i> para o <i>cluster</i> índice e 2 <i>clusters</i> para o <i>cluster</i> mercado, no ano 2005	90
TABELA 15 - Classificação BOVESPA, com 8 <i>clusters</i> para o <i>cluster</i> índice e 7 <i>clusters</i> para o <i>cluster</i> mercado, no ano 2006	90
TABELA 16 - Classificação BOVESPA, com 10 <i>clusters</i> para o <i>cluster</i> índice e 4 <i>clusters</i> para CM, no ano 2007	91
TABELA 17 - Resumo da classificação setorial BOVESPA, para o <i>cluster</i> índice no período 2000–2007.....	91
TABELA 18 - Resumo da classificação setorial BOVESPA, para o <i>cluster</i> mercado no período 2000–2007.....	92
TABELA19 - Classificação BOVESPA, com 10 <i>clusters</i> para o <i>cluster</i> índice, no ano 2000.....	93
TABELA 20 - Classificação BOVESPA, com 10 <i>clusters</i> para o <i>cluster</i> mercado no ano 2000.....	94
TABELA 21 - Classificação BOVESPA,10 <i>clusters</i> para o <i>cluster</i> índice, ano 2001.....	95
TABELA 22 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2001.....	95
TABELA 23 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2002.....	96
TABELA 24 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2002.....	96
TABELA 25 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2003.....	97
TABELA 26 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2003.....	97
TABELA 27 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2004.....	98
TABELA 28 - Classificação BOVESPA, 10 <i>clusters</i> para <i>cluster</i> mercado, ano 2004.....	99
TABELA 29 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2005.....	99

TABELA 30 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2005.....	100
TABELA 31- Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2006.....	101
TABELA 32 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2006.....	101
TABELA 33 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2007.....	102
TABELA 34 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2007.....	102
TABELA 35 - Resumo da classificação setorial BOVESPA, para o <i>cluster</i> índice para o período 2000–2007.....	103
TABELA 36 - Resumo da classificação setorial BOVESPA, para o <i>cluster</i> mercado para o período 2000–2007.....	104
TABELA 37 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> índice, ano 2001.....	106
TABELA 38 - Classificação BOVESPA, 10 <i>clusters</i> para o <i>cluster</i> mercado, ano 2001.....	106
TABELA 39 - Valores de significância para o Teste de Wilcoxon, com 4 <i>clusters</i> para o <i>cluster</i> índice e 2 <i>clusters</i> para o <i>cluster</i> mercado, com critério de regra de parada, para os anos do período de 2001-2007.....	107
TABELA 40 - Valores de significância para o teste de Wilcoxon, com 10 <i>clusters</i> para os <i>clusters</i> índice e mercado, com critério da classificação BOVESPA, para os anos do período de 2001-2007.....	109

LISTA DE QUADROS

QUADRO 01 - Indicadores com suas medidas e atributos	44
QUADRO 02 - Variáveis e suas respectivas medidas, dispostas segundo os indicadores de liquidez, rentabilidade, estrutura de capital, dividendos, mercado e tamanho.....	50
QUADRO 03 - Parâmetros e ajustes dos dados das variáveis coletadas.....	51
QUADRO 04 - Quantidade de ações em negociação nos respectivos anos, no período de 2000-2007	62
QUADRO 05 - Variáveis dos <i>clusters</i> índice, mercado e variáveis de controle.....	82
QUADRO 06 - Dados da matriz de aglomeração para o <i>cluster</i> mercado sem as variáveis de tamanho, para o ano de 2000.....	83
QUADRO 07 - Definição do número de <i>clusters</i> considerando ou não as variáveis de tamanho na construção dos <i>clusters</i> no período de 2000–2007.....	84
QUADRO 08 - Definição do número de clusters segundo o mesmo número de subdivisões da classificação da BOVESPA, ou seja, 10, no período de 2000–2007.....	85
QUADRO 09 - Levantamento do número de <i>clusters</i> quando se impõe o número de <i>clusters</i> iguais a 20, o mesmo número de subdivisões da classificação NAICS, no período de 2000–2007.....	105

LISTA DE ANEXOS

ANEXO A - Propriedades do operador valor esperado (<i>E</i>).....	119
---	-----

LISTA DE APÊNDICES

APÊNDICE A - Quantidade de dados omissos com percentuais para cada variável nos respectivos anos do período 2000-2007.....	121
APÊNDICE B – Estatística descritivas das variáveis para cada ano do período de 2000 - 2007, apresentados nas respectivas tabelas de 01 a 07.....	122
APÊNDICE C – Teste Kruskal Wallis aplicado para as classificações setoriais NAICS, e BOVESPA, segundo cada ano do período de 2000-2007.....	126
APÊNDICE D – Número de <i>clusters</i> com mais de um elemento, considerando o mesmo número de subdivisões da classificação NAIC nível 1, no período de 2000–2007.....	127
APÊNDICE E – Classificação NAICS com frequência percentual nos setores, com número de clusters definidos conforme a regra de parada, de 2000–2007.....	128
APÊNDICE F - Critério classificação BOVESPA, com variáveis de tamanho, para cada ano do período 2000-2007.....	133

SUMÁRIO

INTRODUÇÃO	13
1.1 Problema de pesquisa.....	16
1.2 Objetivos.....	16
1.3 Estrutura do estudo	17
2. REFERENCIAL TEÓRICO	18
2.1 Classificação setorial	18
2.2 Carteiras de investimento	23
2.3 Diversificação internacional	26
2.4 Análise econômico-financeira.....	41
2.5 Grupos de controle	45
3. TIPO E MÉTODO DE ESTUDO	48
3.1 Abordagens da pesquisa.....	48
3.2 População amostral e período de estudo	48
3.3 Coleta de dados e definição das variáveis.....	49
3.4 Dados omissos e outliers	52
3.4.1 Dados omissos	52
3.4.2 Outliers	53
3.5 Estatísticas descritivas.....	54
3.6 Técnica estatística	55
3.6.1 Estágios de aplicação da técnica análise de <i>cluster</i>	58
3.7 Definição da análise de <i>cluster</i> via tratamento vetorial	63
3.7.1 Noções de distância	65
3.7.2 Algoritmo de ligação	67
3.7.3 Dendograma	69
3.7.4 Critério de seleção do número de <i>clusters</i>	70
3.7.5 Homogeneidade dos <i>clusters</i>	70
3.7.6 Estabilidade dos <i>clusters</i>	71
4. ANÁLISE DOS RESULTADOS	73

4.1 Procedimento de amostragem	73
4.2 Tratamento dos dados	74
4.2.1 Dados omissos	74
4.2.2 Dados <i>outliers</i>	75
4.3 Classificações setoriais	76
4.4 Estatísticas descritivas.....	79
4.5 Análise dos <i>clusters</i>	82
4.5.1 Definição do número de <i>clusters</i>	83
4.6 Construção dos <i>clusters</i>	85
4.6.1 Análise dos <i>clusters</i> segundo a regra de parada	86
4.6.2 Análise dos <i>clusters</i> índices e mercados formados com o mesmo número de setores das classificações BOVESPA (10) e NAICS (20)	93
4.7 Análise dos <i>clusters</i> índice e mercado com a inclusão das variáveis de tamanho, segundo a classificação BOVESPA	105
5. ESTABILIDADE DOS <i>CLUSTERS</i>	108
6. CONCLUSÕES.....	110
7. REFERÊNCIAS BIBLIOGRÁFICAS	113
ANEXOS	119
ANEXO A - Propriedades do operador valor esperado (<i>E</i>)	120
APÊNDICE A - Quantidade de dados omissos com percentuais para cada variável nos respectivos anos do período 2000-2007	121
APÊNDICE B – Estatística descritivas das variáveis para cada ano do período de 2000 - 2007	122
APÊNDICE C – Teste Kruskal Wallis aplicado para as classificações setoriais NAICS e BOVESPA, segundo cada ano do período de 2000-2007	126
APÊNDICE D – Número de <i>clusters</i> com mais de um elemento, considerando o mesmo número de subdivisões da classificação NAICS, no período de 2000 – 2007	127
APÊNDICE E – Classificação NAICS com frequência percentual nos setores, com número de <i>clusters</i> definidos conforme a regra de parada, no período de 2000 – 2007	128
APÊNDICE F - Critério classificação BOVESPA, com variáveis de tamanho, para cada ano do período 2000-2007	133

INTRODUÇÃO

O problema econômico pode ser sintetizado em três questões básicas, o que, quanto e para quem produzir e as respostas para estas indagações encontram-se na economia de mercado (ANGELO e SANVICENTE, 1990). Com isso, do ponto de vista econômico, o agrupamento de empresas de acordo com o ramo de atividade ou com o comportamento de certas variáveis independentemente do tipo de produto ou serviço oferecido, serve como base para a compreensão das características e relações das empresas nos mercados em que vendem ou compram. Nesse contexto, é possível identificar dois enfoques para o agrupamento de empresas. Um deles, a abordagem clássica, a "análise da indústria" ou do "posicionamento estratégico", tem em Michael Porter seu maior protagonista. Esta abordagem prioriza a análise dos mercados e da competição e o entendimento da posição relativa de cada empresa em sua indústria ou segmento produtivo, como elementos primordiais no processo de formulação da estratégia (PORTER, 1996). Os principais focos de análise são produtos, consumidores e competidores, e a estratégia da empresa deve ser resultante da identificação de tendências e de oportunidades. Nesse sentido, é considerada uma abordagem "de fora para dentro", e reconhece nas atividades da empresa a base da vantagem competitiva (PROENÇA, 1999). Nesse caso o agrupamento de empresas relaciona-se à necessidade que as empresas têm de conhecer o ambiente de mercado e dele extraírem elementos que facilitem o planejamento de suas atividades, sendo o setor previamente definido em função do ramo de atividade em que a empresa está inserida.

Uma abordagem alternativa é construída a partir da "visão da empresa baseada em recursos" (*Resources Based View of the Firm*). Essa abordagem procura ampliar e refinar o quadro de referência dos tomadores de decisão. Considera que toda empresa possui um *portfolio* de recursos físicos, financeiros, intangíveis (marca, imagem), organizacionais (cultura organizacional, sistemas administrativos) e recursos humanos, sendo a partir desse *portfolio* que a empresa pode criar vantagens competitivas. Para os defensores dessa abordagem (PRAHALAD e HAMEL, 1990; KROGH e ROSS, 1995), a definição das estratégias competitivas devem partir de uma perfeita compreensão das possibilidades estratégicas passíveis de serem operacionalizadas e sustentadas por tais recursos. Isto caracteriza essa abordagem como primordialmente "de dentro para fora". Mais precisamente, esta abordagem parte da concepção de setor formado de acordo com o comportamento de certas variáveis, independentemente do tipo de produto ou serviço oferecido. Com isso,

evidenciar e avaliar a extensão desses paradigmas é uma importante contribuição para a análise e formulação de estratégias e de políticas mais eficazes para as organizações e países.

Sob a ótica das duas vertentes citadas à segmentação estratégica das empresas, torna-se uma importante estratégia de investimento, uma vez que o processo de decisão por parte dos investidores está cada vez mais complexo, graças à crescente internacionalização dos mercados financeiros, que proporciona uma ampla gama de escolha entre os milhares de produtos financeiros disponíveis em qualquer parte do mundo. Neste contexto, a unificação das informações comerciais e econômicas, tanto do ponto de vista estatístico quanto do ponto de vista conceitual dos países torna-se ponto chave para um melhor entendimento por parte do investidor, sendo a classificação setorial neste viés um poderoso instrumento de análise.

A classificação setorial de empresas pode também ser aplicada em diversos outros contextos, como por exemplo, a construção de *portfolio*, a análise econômico-financeira e a formação de grupos de controle. Cabe, no entanto salientar que a classificação setorial se apresenta dependendo do tipo de problema a ser tratado, por exemplo, em estudos de desenvolvimento, dentro da visão econômica no qual se pode classificar as indústrias em setores primário, secundário e terciário. Entretanto, neste tipo de classificação o setor secundário agrega “indústrias” distintas, e, em outros tipos de análise, como a análise da organização industrial, as indústrias constituem uma unidade de agregação, que é o setor de atividade. Neste estudo as classificações utilizadas focam no setor de atividade.

Dentre as diversas aplicabilidades da classificação setorial mostra-se relevante a construção de *portfolios* de ativos globais (internacionais) no que tange a diversificação internacional que, por sua vez, representa uma estratégia de investimento. Na prática internacional de gerenciamento de ativos, Heston e Rouwenhorst (1995) argumentam que administradores de *portfolios* usam duas estratégias de seleção de ativos com relação ao tipo de diversificação de um *portfolio*.

A primeira estratégia é aquela que o administrador acredita que o retorno do *portfolio* internacional é guiado pela diversificação por setor, e apresentam para esta estratégia dois estágios. O primeiro estágio é aquele no qual o administrador aloca recursos em muitas indústrias para formação do *portfolio*, o segundo aquele no qual o administrador usa a análise por setor para seleção dos ativos mais promissores provenientes daquele setor.

A segunda estratégia é aquela que o administrador acredita que o fator mercado doméstico é mais importante para o retorno do *portfolio* internacional, assim, decide primeiro pela alocação por país e seleciona, posteriormente, as ações mais atrativas de cada país.

Tradicionalmente o efeito país tem dominado o efeito setor (indústria), tal como mostram os estudos de Solnik (1974), Heston e Rouwenhorst (1995), Griffin e Karolyi (1998), Rouwenhorst (1999), Hamelink *et al.* (2001) e Clifford (2002). Por outro lado autores como Roll (1992), Cavaglia *et al.* (2000), Baca *et al.* (2000), Boillat *et al.* (2002) e Flavin (2004) mostram que o efeito setor tem nivelado ou superado o efeito país nos últimos anos. Estes se baseiam no argumento que a diversificação internacional por setor oportuniza grandes reduções de risco, então tal estratégia é melhor do que a estratégia tradicional de diversificação do *portfolio* por país. Entretanto, autores como Ferreira (2006), Phylaktis e Kia (2006), Steliaros e Thomas (2006) mostram evidências da importância de ambas as estratégias de diversificação.

Quanto à aplicabilidade da classificação setorial sob o ponto de vista da análise econômico-financeira mostra-se relevante à medida que proporciona uma visão abrangente da empresa dentro do segmento de atuação. Para autores como Padovese e Benedicto (2007) e Matarazzo (2003) a utilização de padrões setoriais para a análise comparativa é um instrumento importante para uma melhor avaliação da empresa, pois permite aferir os diversos indicadores dentro de padrões referendados pela realidade de mercado.

Outro ponto relevante de aplicabilidade da classificação setorial está na sua utilização em pesquisas acadêmicas. De acordo com Sampieri *et al.* (2006) deve-se escolher ou desenvolver um modelo específico para cada pesquisa com vistas a responder as questões de pesquisa propostas. Nos modelos experimentais, geralmente usados sob o enfoque quantitativo é que a classificação setorial, em geral, é utilizada. Esses modelos usam a manipulação de uma ou mais variáveis independentes (causas) como forma de analisar suas conseqüências sobre uma ou mais variáveis dependentes (efeitos), onde cada nível de manipulação possui o grupo exposto à presença da variável independente (grupo experimental) e o não exposto (grupo de controle). Dentro dessa idéia, com foco específico na área de finanças, salienta-se que a formação do grupo de controle utiliza a classificação setorial para seleção de uma empresa similar àquela do objeto de pesquisa, mas sem a característica em estudo. E, juntamente com a construção de *portfolio*, que é preponderante para o mercado de capitais, assim como à análise econômico-financeira das empresas, que representa uma importante ferramenta para a prospecção de oportunidades de investimento oriundas de análises técnicas corretas, um estudo a respeito das peculiaridades da classificação setorial mostra-se fundamental.

1.1 Problema de pesquisa

Nestes três contextos, ora citados, geralmente, a classificação setorial é utilizada sob o pressuposto de que as empresas pertencentes a um mesmo setor de atividade apresentam características similares. Por outro lado, empresas de setores distintos apresentam características dissimilares, isto é, os setores apresentam diferenças. Este pressuposto está diretamente associado ao grau de informação disponível sobre as variáveis que condicionam o comportamento das empresas em geral, uma vez que o agrupamento de empresas procura estruturar as informações de um segmento específico da atividade econômica, visando orientar a ação dos agentes que atuam motivados por múltiplos interesses.

Portanto, para que a classificação setorial seja um instrumento eficiente utilizado na construção de *portfolios*, na análise econômico-financeira das empresas e na formação de grupos de controle, é preciso que este pressuposto seja observado. Assim, o desafio apresentado à academia e, especificamente, o foco deste estudo, é a resposta à seguinte questão de pesquisa:

A classificação setorial mostra-se eficiente?

1.2 Objetivos

Diante da questão de pesquisa, este estudo tem como objetivo geral avaliar a eficiência da classificação setorial a partir do método análise de *cluster*.

Quanto aos objetivos específicos destacam-se:

- i) identificar características relevantes para a análise dos setores;
- ii) utilizar a análise de *clusters* para a formação dos grupos ou *clusters*;
- iii) identificar a composição setorial dos *clusters*;
- iv) avaliar a “estabilidade” dos *clusters* ao longo do período de estudo.

1.3 Estrutura do estudo

O estudo é apresentado em cinco capítulos. O capítulo um, introduz o estudo e engloba a contextualização do tema, a justificativa, a formulação do problema, o objetivo geral e os objetivos específicos.

No capítulo dois é feita uma revisão bibliográfica sobre o tema dentro da literatura existente. De início aborda uma visão teórica de conceitos e princípios fundamentais da classificação setorial, da diversificação de formação de *portfolio*, da análise econômico-financeira e da formação de grupos de controle.

O capítulo três trata dos aspectos metodológicos necessários para a consecução do estudo, ilustra o método a ser utilizado, por meio de sua descrição, explicita os critérios de seleção dos elementos da amostra a forma de coleta de dados e os procedimentos adotados.

O capítulo quatro mostra a análise dos resultados e o desenvolvimento das atividades o capítulo cinco trata da estabilidade.

Por fim o capítulo seis apresenta a conclusão da pesquisa juntamente com as limitações do método.

2. REFERENCIAL TEÓRICO

O objetivo desse capítulo é identificar o papel da classificação setorial nas pesquisas em finanças, com ênfase na construção de *portfolio*, análise econômico-financeira e grupos de controle.

O capítulo está dividido em quatro seções, sendo que as mesmas referem-se à aplicabilidade da classificação setorial e mostram os pressupostos teóricos básicos. A primeira seção trata da classificação setorial elencando conceitos básicos e estudos de aplicabilidades. A segunda seção mostra os resultados básicos para a construção de *portfolios* com foco no debate crucial da diversificação por país versus setor de um *portfolio* internacional. A terceira seção pontua questões básicas na análise econômico-financeira e mostra estudos de aplicabilidades referentes ao tema. E, finalmente a quarta seção traz estudos de aplicabilidade com ênfase em grupos de controle.

2.1 Classificação setorial

O agrupamento de empresas consiste numa abordagem metodológica que tem como objetivo a organização de dados e informações de tal sorte a permitir a realização de estudos analíticos que orientem o processo de tomada de decisão (ANGELO e SANVICENTE,1990). Dessa forma, a eficácia de uma organização, traduz-se pelo seu resultado econômico à medida que todas as decisões tomadas levem a sua otimização (SANTOS e PONTE, 1998). A otimização de resultados se relaciona diretamente aos caminhos ou ações para alcançá-los. Chandler (1962) definiu estratégia como “a determinação de metas básicas a longo prazo e dos objetivos de uma empresa e a adoção da linha de ação para alcançar essas metas”. E nessa linha, ressalta-se a vantagem competitiva que é conquistada através de um amplo processo de estratégias, o que proporciona a empresa sustentar uma situação de destaque no mercado em relação aos concorrentes. Uma das primeiras referências à expressão “vantagem competitiva” na literatura de estratégia pode ser encontrada em Ansoff (1965) que a definiu como a vantagem de perceber, de forma pró-ativa, tendências de mercado à frente dos concorrentes e de ajustar a oferta em função dessa antecipação. O que significa que pode ser vista como o objetivo das ações da empresa e para explicar a diversidade entre elas.

A visão baseada em recursos (*Resource Based View* – RBV) oferece um arcabouço próprio para o estudo da vantagem competitiva ao enfatizar que os recursos específicos da empresa são determinantes para a geração de lucro econômico (BARNEY, 1991; PETERAF, 1993). Já as teorias de posicionamento estratégico sustentam que a vantagem competitiva pode derivar não apenas de recursos específicos, mas também de posições de mercado privilegiadas (PORTER, 1979; CAVES, 1982). O que implica em ser explicada por fatores externos (mercados, estrutura das indústrias).

A estratégia competitiva, segundo Porter (1986) preocupa-se em estabelecer uma abordagem para a competição de uma unidade empresarial em sua indústria, que seja, ao mesmo tempo, lucrativa e sustentável. Suas pesquisas objetivavam analisar a estrutura de um determinado setor e a capacidade que tem uma unidade empresarial de ser lucrativa e sustentável nesse setor.

Por outro lado na Visão Baseada em Recursos (VBR), que considera as empresas como um feixe de recursos produtivos e estratégicos que permitem alcançar diferentes estágios de desempenho, concebem que as empresas utilizam esses diferentes recursos como elementos de sustentação de suas estratégias competitivas. Assim, a principal consequência dos pressupostos da VBR é que a diferença de desempenho e a consequente obtenção da vantagem competitiva entre as empresas são explicadas pela heterogeneidade dos recursos. Isso significa que a fonte básica para a vantagem competitiva são os recursos estratégicos desenvolvidos e controlados pelas firmas (AMIT e SCHOEMAKER, 1993). Portanto exprime a idéia de que o desempenho competitivo está muito mais relacionado aos recursos que a empresa detém e administra do que às características da indústria ou do setor ao qual está relacionada (BARNEY, 1991; AMIT e SCHOEMAKER, 1993).

Logo, diante do exposto e partindo do princípio que as classificações apóiam as análises setoriais da economia e tratamento de dados qualquer tipo de estudo que envolva setores industriais deve mostrar com clareza as definições preestabelecidas das classificações. Com o objetivo de elas expressarem a forma mais fidedigna possível.

A classificação setorial pode ser realizada segundo diferentes critérios como, por exemplo:

- i) Ross *et al.* (2007) destacam o comportamento cíclico e não cíclico das empresas, mostrando que existem empresas que produzem um bom desempenho na fase de expansão do ciclo econômico e possui um mau desempenho na fase de contração. Em complemento, Damodaran (2002) afirma que setores classificados como cíclicos,

possuem tendência a apresentar risco de mercado mais alto do que empresas com comportamento não cíclico, pois estas últimas possuem uma tendência a apresentar baixos riscos.

ii) a taxonomia de classificação apresentada por Pavitt (1984) posteriormente adaptada por Dosi *et al.* (1990) baseia-se na *performance* tecnológica, partindo do princípio que diferentes trajetórias tecnológicas das indústrias podem ser explicadas pelas condições de oportunidade, cumulatividade e apropriação das tecnologias, assim classificaram os setores de alguns países desenvolvidos. O modelo desenvolvido classifica os setores em dominados por fornecedores (setores tradicionais na manufatura, na agricultura, habitação, produção familiar informal e diversos serviços profissionais, financeiros e comerciais), baseados em ciência (indústrias químicas e eletroeletrônicas) e intensivo em produção. Sendo por sua vez o ramo intensivo em produção dividido em intensivo em escala (fabricantes de matérias padronizados aço e ferro e de bens de consumo duráveis e veículos) e fornecedores especializados (firmas pequenas que fornecem insumos de alta *performance* na forma de máquinas, componentes, instrumentos ou software).

iii) a classificação setorial das empresas listadas na Bolsa de Valores de São Paulo (BOVESPA) é proveniente de uma estrutura de classificação setorial que considera, principalmente, os tipos e os usos dos produtos ou serviços desenvolvidos pelas empresas com o propósito de fornecer uma identificação mais objetiva dos setores de atuação das empresas. Bem como se, aproximar de critérios utilizados pelo mercado financeiro nacional e internacional. Para a classificação das empresas, são analisados os produtos ou serviços que mais contribuem para a formação das receitas das companhias, considerando-se, ainda, as receitas geradas no âmbito de empresas investidas de forma proporcional às participações acionárias detidas. No caso de companhias de participação, é considerada a contribuição de cada setor na formação das receitas consolidadas, sendo que, se algum setor representa participação maior ou igual a dois terços das receitas, a empresa de participação fica classificada nesse setor; caso contrário, a empresa de participação fica classificada como *holding* diversificado. A estrutura setorial e a classificação das empresas negociadas são objetos de revisões periódicas. No caso de haver alterações nos produtos ou serviços com maior contribuição para a formação das receitas, é avaliado se essas alterações

se mantêm ao longo do tempo ou constituem uma tendência, antes que seja feita a reclassificação setorial da empresa (BOVESPA, 2008).

Outras classificações setoriais podem ser destacadas, a seguir elenca-se a Classificação Nacional de Empresas (CNAE), que é a classificação oficialmente adotada pelo Sistema Estatístico Nacional na produção de estatísticas por tipo de atividade econômica, e pela Administração Pública, na identificação da atividade econômica em cadastros e registros de pessoa jurídica.

A CNAE permite ampliar a comparabilidade entre as estatísticas econômicas provenientes de distintas fontes nacionais, e das estatísticas do país no plano internacional, ao prover uma base padronizada para a coleta, análise e disseminação das estatísticas relativas à atividade econômica. E tem como principal propósito ser uma classificação estandardizada das atividades econômicas produtivas, provendo um conjunto de categorias para serem usadas na coleta e divulgação de estatísticas por tipo de atividade econômica (IBGE, 2007). Cujo modelo de codificação adotado é formado de um código alfabético (uma letra) para indicar o primeiro nível de grupamento da classificação, a *Seção*, e de códigos numéricos para os demais níveis de agregação, *Divisão*, *Grupo*, *Classe* e *Subclasse*. O código de quatro dígitos das classes CNAE é acompanhado de um dígito verificador (DV) e a maneira como as atividades são combinadas nos estabelecimentos produtivos é um critério central na definição das classes (4 dígitos).

Os setores da indústria de transformação, por exemplo, foram agrupados conforme a classificação setorial de Dosi *et al.* (1990), onde cada um dos setores a dois dígitos da Classificação Nacional das Atividades Econômicas (CNAE) foi incluído em um dos grupos dessa classificação.

O Instituto Brasileiro de Geografia e Empresas (IBGE), órgão governamental, enfatiza a importância da classificação com base nas atividades econômicas, para organizar as informações das unidades produtivas e institucionais, com vistas a produzir as estatísticas dos fenômenos derivados da sua participação no processo econômico. Ou seja, servem para classificar as unidades de produção de acordo com a atividade que desenvolvem, em categorias definidas como segmentos homogêneos, quanto à similaridade de funções produtivas. Dessa forma, é possível definir cada atividade da economia, tornando comparáveis suas informações (IBGE, 2007). Em relação ao conceito de setor, complementa, este reflete uma visão da economia que privilegia a complexidade, a diferença entre as partes

que compõem o sistema econômico, cuja função é reunir empresas ou atividades econômicas que apresentam elementos comuns.

Contudo, ainda existem divergências entre os limites estabelecidos para os setores em que uma economia está dividida, dificultando o processo de análise dos dados. Sendo assim, persiste uma busca contínua por um sistema de classificações de atividades econômicas numa mesma base conceitual e comparável em todo o mundo.

Nessa linha, ressalta-se a prerrogativa dos dados serem agrupados em categorias com características semelhantes (setor), proporcionando informações agregadas necessárias para descrever e analisar a estrutura, funcionamento e inter-relações setoriais da economia como um todo.

Assim, o desenvolvimento e aplicação de modelos de classificações nos diversos âmbitos da indústria, nos sistemas tecnológicos e até nas políticas científicas e tecnológicas constituem elementos de grande importância dentro das tendências recentes de estudos setoriais (PORCARO, 2000).

Alguns estudos tais como o de Dalhstedt *et al.* (1994) e Gutiérrez e Sorensen (2006) evidenciam a importância das classificações no âmbito do setor bancário, demonstrando que é problemático comparar as empresas através da razão financeira, entretanto enfatizam que mesmo que as empresas pertencem a uma mesma categoria de classificação oficial é preciso ter cuidado. Ambos usam como método estatístico de segmentação das empresas a análise de *cluster*, e usam na amostra setores bancários pertencentes a países que participam da união européia.

No Brasil, Ângelo e Sanvicente (1990), com vistas a oferecer uma alternativa metodológica à análise setorial, mostram através de um exemplo do mercado de valores mobiliários a fragilidade para descobrir a existência de correlação entre os dados usuais de classificação industrial e o grau sistemático associado ao investimento em atividades em diversas indústrias. Usam como principal método estatístico para atender os objetivos propostos a análise de *cluster* e a amostra é constituída de 6 setores com 10 empresas, listadas na BOVESPA, no período de 1982-1987. E concluem que os dados publicados, de classificação industrial, embora representem uma fonte de informação, devem ser adaptados ao problema que se deseja estudar.

2.2 Carteiras de investimento

A seguir elenca-se alguns pressupostos básicos da moderna teoria de carteiras com vistas a elucidar a diversificação internacional (estratégia de investimentos). A publicação do célebre artigo *Portfolio Selection* de Harry Markowitz, em 1952, culminou no surgimento da moderna teoria de *portfolios* (carteiras). O artigo chama atenção para a prática da diversificação de ativos na composição de um *portfólio*, argumentando que através da combinação de ativos, que não variam em conjunto, um investidor pode reduzir o impacto do risco e obter um maior retorno. Sua grande inovação está em relacionar o retorno, dado pelo retorno esperado, e o risco, quantificado pelo desvio padrão do retorno dos ativos. Através da consideração de um número finito de combinações possíveis de ativos que comporiam uma carteira, Markowitz (1952) contraria a teoria da maximização do retorno (todos os recursos alocados no ativo de maior retorno), através da constatação do fenômeno da diversificação.

Na construção do seu modelo teórico, esse autor, parte do princípio que o investidor racional toma suas decisões baseadas não apenas no retorno, mas leva em consideração o risco, assim, numa carteira diversificada a satisfação do investidor está na maximização da relação risco – retorno. E, através de um conjunto de pontos otimizados da relação risco-retorno constrói a fronteira eficiente sobre a qual está a carteira eficiente. Então, define carteira eficiente, como aquela que proporciona o maior retorno esperado para certo risco percebido ou, alternativamente, mínimo risco para dado nível de retorno esperado.

De maneira similar, Elton *et al.* (2004) definem carteira eficiente a partir da análise de combinações de todos os ativos com risco, sendo a carteira eficiente a preferida por todos os investidores que têm aversão ao risco e gostam de retornos mais altos. Estes autores definem retorno como uma referência absoluta ou relativa associada à diferença entre a meta de consumo alcançada no futuro e o consumo abdicado no presente.

A idéia de risco, atualmente, é utilizada diariamente na maioria das operações financeiras. Podem-se citar vários exemplos, um operador da bolsa de valores trabalha com esse conceito durante várias horas por dia, grande empresas utilizam o risco para avaliar seus investimentos e até mesmo as pequenas lojas de varejo utilizam a idéia do risco para calcular prestações, preocupadas com o risco de inadimplência. Markowitz (1952) define risco, na área financeira, como o desvio-padrão (σ) em relação a uma média de um conjunto de retornos.

De acordo com Damodaran (2001), as questões referentes ao modo como o risco é medido, como é recompensado e quanto risco assumir são fundamentais na decisão de investimento, desde a alocação de ativos até a sua avaliação. Para Elton *et al.* (2004) a idéia de

risco está diretamente associada às probabilidades de ocorrência de determinados resultados em relação a um valor médio esperado no futuro, revelando uma possibilidade de perda para um dado investidor. A análise do nível de risco medida pelo desvio padrão indica se o valor médio esperado é representativo do comportamento observado. Assim, ao se tomarem decisões de investimento com base num resultado médio esperado, o desvio padrão passa a revelar o risco da operação, ou seja, a dispersão dos resultados em relação à média.

De um modo geral, a existência de risco significa que o investidor não pode mais associar um único número ou qualquer resultado num investimento de ativos. O resultado precisa ser descrito por um conjunto de valores e suas probabilidades de ocorrência, ou seja, uma distribuição de retornos.

Assim, considerando a distribuição de retornos começa-se o exame dos dois atributos mais freqüentemente empregados: uma medida de tendência central chamada de retorno esperado, e uma medida de dispersão em torno da média chamada de desvio padrão.

O retorno de um *portfolio p* de ativos é dado por uma média ponderada dos retornos dos ativos individuais (ELTON *et al.*, 2004). O peso aplicado a cada retorno corresponde à fração do valor do *portfolio* aplicado naquele ativo. O cálculo do retorno de um dado *portfolio p* se expressa por:

$$R_{pj} = \sum_{i=1}^n (X_i R_{ij}) \quad (1)$$

onde:

R_{pj} = *j*-ésimo retorno do *portfolio*

X_i = fração dos fundos do investidor aplicado no *i*-ésimo ativo

n = número de ativos do *portfolio*

De modo generalizado a fórmula (1) pode ser reescrita da seguinte forma:

$$\bar{R}_p = \sum_{i=1}^n X_i \bar{R}_i \quad (2)$$

onde:

\bar{R}_p = retorno médio do *portfolio*

\bar{R}_i = retorno médio do ativo

A seguir, a partir de (1), demonstra-se a fórmula generalizada (2) com base em Elton *et al.* (2004, p. 66-67). Por definição, o retorno esperado é uma média ponderada dos retornos dos ativos individuais. Então o valor esperado do retorno de um *portfolio* dado por (1) fica:

$$R_p = E(R_p) = E\left(\sum_{i=1}^n (X_i R_{ij})\right) \quad (3)$$

Como o valor esperado da soma dos vários retornos é igual à soma dos valores esperados, conforme Anexo A, então:

$$R_p = E(R_p) = E\left(\sum_{i=1}^n (X_i R_{ij})\right) = \sum_{i=1}^n E(X_i R_{ij}) \quad (4)$$

E, usando $E[C(R_{ij})] = CR_1$, isto é, o valor esperado de uma constante vezes o retorno é igual a constante vezes o retorno esperado, Anexo A, chega-se a (2).

A variância (σ^2) de um dado *portfolio* p por definição é o valor esperado dos quadrados dos desvios do retorno do *portfolio* em relação ao retorno médio do *portfolio* e depende tanto das variâncias dos retornos dos ativos que a compõem quanto da covariância entre os ativos. Mais precisamente, a variância mede a variabilidade do retorno de um dado ativo e a covariância mede a relação entre os ativos ou mede como os retornos dos ativos variam em conjunto.

Segundo Ross *et al.* (2007), considerando os níveis de variâncias dos ativos individuais, uma covariância positiva entre dois ativos indica um aumento da variância do retorno do *portfolio* e, uma covariância negativa, reduz a variância de retorno do *portfolio*.

Este importante resultado parece confirmar o que diz o senso comum. Se um dos ativos tende a se valorizar o outro tende a cair e vice-versa, logo os ativos estão contrabalanceando um ao outro.

Além disso, salienta-se que a padronização da covariância pode ser obtida dividindo-se a covariância entre os dois ativos pelo produto dos desvios padrão dos dois ativos, sendo essa medida denominada de coeficiente de correlação. Essa medida tem as mesmas propriedades da covariância, mas dentro do intervalo entre -1 e +1. O coeficiente de correlação é dado por:

$$\rho_{ik} = \frac{\delta_{ik}}{\delta_i \delta_k} \quad (5)$$

onde:

δ_{ik} = covariância entre dois ativos

δ = desvio padrão de cada ativo

O cálculo da variância de um dado *portfolio p* se expressa por:

$$\delta_p^2 = \sum_{j=1}^n (x_j^2 \delta_j^2) + \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n (x_j x_k \delta_{jk}) \quad (6)$$

onde:

δ_p^2 = variância do *portfolio*

δ_{jk} = covariância entre dois ativos

δ_j^2 = variância de cada ativo

X_j^2 = o quadrado da proporção

As equações (4) e (6) definem os dados necessários para a realização de uma análise do *portfolio* eficiente.

2.3 Diversificação internacional

A diversificação é um fator determinante e complexo num investimento financeiro. Dada à complexidade dessa tarefa, muitos investidores confiam em agentes profissionais como os administradores de *portfolios* para selecionar e atualizar os *portfolios* conforme seus interesses.

Para Levy (1999) um investidor é um indivíduo que está disposto a abdicar do consumo hoje para alcançar a meta de um nível maior de consumo no futuro, assim nessa concepção o ato de investir remete-se a busca por veículos e estratégias de investimentos adequados para o alcance de tais metas.

Dentre o grande número de oportunidades de investimentos, a estratégia de diversificação internacional, na qual o investidor aplica parte do recurso em ativos emitidos em outros países, apresenta-se como um determinante para o retorno do investidor. No entanto para se descobrir a importância da diversificação internacional no *portfolio* do investidor, Elton *et al.* (2004) mostram a relevância da análise de correlação entre os mercados, bem como o risco e o retorno de cada mercado.

Em geral, para o cálculo do coeficiente de correlação usam-se os retornos mensais do índice de mercado. Os índices calculados nos EUA, por exemplo, pela *Morgan Stanley Capital Internacional*, proporciona uma análise de correlação sob o ponto de vista de um investidor americano. Os índices são ponderados pelo valor de mercado, sendo a proporção de cada ativo no índice determinado pelo valor de mercado dividido pelo valor de mercado agregado de todos os ativos em seu mercado (ELTON *et al.*, 2004). Os índices incluem ativos que representam aproximadamente 60% do valor total de mercado em cada país.

Assim a análise de correlação pode ser feita observando que as correlações negativas e positivas, respectivamente, indicam baixa e alta correlação, e quando zero indica ausência de correlação.

O retorno do investimento em um ativo no exterior é afetado pelos retornos dos ativos dentro do mercado e pela variação da taxa de câmbio entre a moeda do ativo e a moeda do país de origem, o que poderá gerar divergência no valor do retorno do ativo em seu próprio mercado e no domicílio do investidor (ELTON *et al.*, 2004).

O retorno esperado para um dado investimento no exterior pode ser expresso como:

$$\bar{R}_{US} = \bar{R}_x + \bar{R}_H \quad (7)$$

onde:

\bar{R}_{US} = retorno estrangeiro

\bar{R}_x = retorno cambial

\bar{R}_h = retorno doméstico ou local

O desvio padrão do retorno se expressa por:

$$\sigma_{US} = \sqrt{[\sigma_X^2 + \sigma_H^2 + 2\sigma_{HX}]} \quad (8)$$

onde:

σ_{US} = desvio padrão do retorno dos títulos estrangeiros

σ_H = desvio padrão do retorno dos títulos no mercado doméstico

σ_X = desvio padrão dos ganhos e perdas cambiais

σ_{HX} = covariância entre ganhos ou perdas cambiais e retornos em um país.

O benefício da diversificação internacional foi identificado primeiramente por Grubel (1968). Contudo, Lessard (1974) foi o primeiro a considerar a importância da composição setorial nas diferenças da variância do retorno do *portfolio* global.

Autores como Roll (1992), Heston e Rouwenhorst (1995) reavivaram essa questão, isto é, na construção do *portfolio* global (internacional) qual a melhor forma de diversificá-lo por país ou por setor (indústria)?

Heston e Rouwenhorst (1995) apresentam um modelo para medir o efeito país e o efeito setor no retorno de um *portfolio* internacional e fornecem um *framework* quantitativo para a seleção de ações analisando as duas estratégias, ou seja, a crença de que o retorno do *portfolio* é guiado pela diversificação por setor e aquela em que o retorno do *portfolio* é guiado pela diversificação por país.

Eles mostram três razões para os administradores terem mais atenção por país do que por indústria na construção do *portfolio* internacional.

Quais sejam:

i) em média a diversificação por país exige um maior número de variáveis para monitorar os erros do que a composição por indústria;

- ii) as ações do mercado doméstico alocadas em diferentes indústrias são substitutas das ações alocadas em indústrias, mas em países diferentes;
- iii) os benefícios da diversificação internacional estão baseados na diversificação por país e não por indústria.

A amostra do estudo constitui-se dos retornos mensais do índice europeu *Morgan Stanley Capital Internacional Contry* (MSCI), que inclui 829 indústrias que pertencem a 12 países da Europa, no período de 1978 a 1992.

Na sua forma generalizada o modelo apresentado é expresso do seguinte modo:

$$R_{it} = \alpha_t + \beta_{jt} + \gamma_{kt} + e_{it} \quad (9)$$

onde:

R_i = retorno do *portfolio*

α = parte comum de todos os retornos

β_j = é o efeito indústria para cada indústria j

γ_k = é o efeito país para cada país k

e_i = é um componente específico do retorno das firmas num período t

De forma mais detalhada apresenta-se o modelo como segue abaixo, isto é, para cada mês da amostra estimam α e β e γ rodando regressões *cross-section* dos retornos das 829 firmas, tomando uma variável *dummy* conjunta para indústria e país.

$$R_i = \alpha + \beta_1 I_{i1} + \beta_2 I_{i2} + \dots + \beta_7 I_{i7} + \gamma_1 C_{i1} + \gamma_2 C_{i2} + \dots + \gamma_{12} C_{i12} + e_i \quad (10)$$

onde:

R_i = retorno de cada indústria

$I_{ij} = 1$ se a firma i pertencer à indústria j (zero se for diferente);

$C_{ik} = 1$ se a firma i pertence ao país k (zero se for diferente).

Os autores identificam no modelo um problema de multicolinearidade perfeita entre os regressores. Este problema origina-se do fato das firmas pertencerem ao setor e ao país ao mesmo tempo e, os autores contornam esse problema, escolhendo como medida do efeito país e do efeito indústria, um fator comum dado pelo índice *European equally weighted* (EW). Assim, cada efeito indústria estimado, B_j , pode ser interpretado como excesso de retorno do *portfolio* em relação a uma determinada indústria j , sem a cobertura de outras indústrias num *portfolio* diversificado geograficamente, e como o *portfolio* tem na composição os mesmos países do índice *European EW*, então se trata de um objeto puro de indústria. Similarmente, o excesso de retorno γ_k de um investimento diversificado industrialmente num país k , representa um país puro.

Portanto a influência dos objetos indústria e país no *portfolio* dependem somente da indústria e do país designado pela ação, o que são inicialmente conhecidos. Assim os excessos de retornos estão dentro das possíveis estratégias de investimentos.

Quanto ao *framework* quantitativo para analisar as duas estratégias acima descritas para a seleção de ações num *portfolio*, argumentam que para a seleção de ativos internacionais, um administrador de *portfolio* reduz o risco através dos benefícios da diversificação por país e setor. Entretanto, sob o ponto de vista do risco e retorno, concluem que é mais importante a diversificação por país do que a diversificação por setor.

Por fim apontam como resultado do estudo que a diversificação via país se apresenta mais significativa do que à diversificação por setor em regiões geograficamente concentradas e economicamente integradas, tal como a Europa Ocidental.

A questão levantada por Heston e Rouwenhorst (1995), da diversificação por país ser mais significativa do que por indústria em regiões geograficamente concentradas e economicamente integradas, é reexaminada por Griffin e Karolyi (1998), os quais especulam se esse efeito ocorre em países separados geograficamente (desconcentrados). Esses autores investigam se a extensão dos ganhos com a diversificação por país, é devido às diferenças na estrutura industrial, ou seja, reexaminam as origens específicas da variação do retorno do *portfolio* global segundo as estratégias de diversificação por país e por setor. Para tanto usam o índice mundial de ações da *Dow Jones*, que é composto por 25 países e 66 classificações industriais, sendo o período do estudo de janeiro de 1992 a abril de 1995. Usam um modelo similar ao de Heston e Rouwenhorst (1995), com o retorno dos ativos decomposto em componentes país e indústria através de uma variável *dummy*.

Estes autores mostram em respostas aos testes padrões, uma importante variação da correlação nos diferentes grupos de indústrias (setores) e afirmam que esta variação é reflexo dos fatores econômicos subjacentes, assim essa correlação é vista como uma importante influência no retorno do *portfolio* global. Contudo, mesmo usando nove agregados de setores industriais bem como uma refinada classificação industrial do índice mundial da *Dow Jones*, os autores confirmam os resultados de Heston e Rouwenhorst (1995), isto é, que uma pequena parte (menos que 4%) da variância do índice país pode ser explicada por sua composição industrial. Entretanto, devido a uma maior divisão da classificação industrial eles obtiveram interessante diferença na variância da *cross-section* do efeito indústria, proveniente do índice setorial, como por exemplo, as indústrias que apresentam boa base de produção internacional que mostram uma variância relativamente grande dos fatores industriais.

Por conseguinte, concluem que administradores de *portfolios* focados em negócios domésticos e nas indústrias de bom comércio internacional devem continuar com essa estratégia de investimento. E para os administradores que investem fora do país, com fortes tendências para as indústrias de bons negócios, o programa de diversificação pode ser enfraquecido por meio da composição industrial.

Por fim, encontram que a diversificação por setor explica pouco da variação do retorno do *portfolio*, enquanto a diversificação por país explica uma grande parcela, e, conseqüentemente, para estes autores, a melhor *performance* de diversificar um *portfolio* internacional é guiada por país.

Os estudos de Heston e Rouwenhorst (1995) e de Griffin e Karolyi (1998) mostram que a diversificação por país é mais importante do que a diversificação por setor, tanto para países concentrados geograficamente quanto para países distantes.

A diversificação internacional pode ser estudada em diferentes contextos. Alguns autores avaliam o efeito geográfico para países geograficamente concentrados (HESTON E ROUWENHORST, 1995) ou países distantes (GRIFFIN E KAROLYI, 1998). Outros analisam a influência de efeitos econômicos dentro do efeito geográfico como é o caso dos países pertencentes à união europeia (ROUWENHORST, 1999). E autores como Hamelink *et al.* (2001) mostram que os fatores crescimento/valor e tamanho provocam diferenças na variância do retorno do *portfolio* quando diversificado por país e setor.

Em relação à influência dos fatores econômicos na diversificação por país ou por setor mostra-se o estudo de Rouwenhorst (1999). O foco desse estudo é analisar o comportamento da diversificação na construção do *portfolio* constituído por ativos dos países que fazem parte da União Monetária Europeia (EMU). A amostra de estudo é formada por 952 retornos das

ações contidos no índice da Morgan Stanley Capital International (MSCI) de 12 países europeus entre 1978 a 1992, com exceção da Noruega e da Suécia todos os países da amostra fazem parte da EMU. Ele usa um modelo de regressão similar ao do modelo de Heston e Rouwenhorst (1995).

Este autor argumenta que, apesar da coordenação fiscal e monetária de muitos estados europeus que seguem o *Tratado de Maastricht*, o efeito país continua como a melhor estratégia de diversificação. Este tratado prevê uma convergência de interesses, de taxas e de harmonização das políticas fiscal e monetária. Também observa que antes e depois do *Tratado de Maastricht* o retorno do *portfolio* diversificado por país continua a dominar o efeito setor.

Por fim, conclui que as forças relativas ao efeito país não são afetados pelo tempo e pela integração econômica, pois apesar da formação da EMU o efeito país não teve nenhuma tendência em desaparecer do mercado europeu. E, como implicação prática para os administradores de *portfolios* enfatiza que tanto gerentes passivos como ativos que buscam o melhor desempenho do *portfolio* no mercado europeu, a composição por país é melhor do que a composição por setor.

Em relação aos fatores que influenciam o retorno do *portfolio* internacional diversificado por setor e por país, apresenta-se o estudo de Hamelink *et al.* (2001) os quais investigam a influência de outros fatores além dos econômicos e que provocam diferenças na variância do retorno do *portfolio* diversificado por país e setor. Tais fatores como, crescimento/valor e tamanho são denominados fator estilo. Os autores argumentam que esses fatores merecem monitoramento cuidadoso na diversificação internacional, pois os mesmos geram implicações importantes no policiamento da alocação do ativo.

Os autores argumentam que o melhor desempenho do retorno proveniente da diversificação por país pode ser reflexo da composição setorial do mercado do país de domicílio, o que torna difícil desassociar os efeitos de ambas as diversificações. Os autores postam as seguintes questões iniciais:

Deve um administrador estruturar seu *portfolio* através de setor, país ou estilo?

Estratégias baseada na diversificação por setor são definitivamente superiores às baseadas na diversificação por estilos?

O método de pesquisa utilizado é a análise multi-fatorial seguida da análise de regressão. Tal método é usado com o objetivo de isolar as várias influências através da eliminação de interações existentes. Salientam que isto não é alcançado quando o índice país é usado como variável para o fator país, índice de setor para fator indústria e índice de estilo

para fator de estilo. Partem do fato de que se a composição industrial difere através de países então o índice país agrega efeito do índice setor e o índice setor agrega efeito do índice país.

O mesmo é verdadeiro entre efeito país e efeito estilo, ou efeito setor e efeito estilo, assim, o modelo multi-fatorial aplicado em ativos individuais satisfaz esse quesito, sendo os efeitos país, setor e estilo facilmente separado, logo, pode-se estimar simultaneamente o retorno pelos fatores “puros” através da técnica de regressão. Com este método, o efeito país pode ser interpretado como a *autoperformance* da diversificação do *portfolio* por país relativo ao índice mundial. O estudo pauta-se sobre os constituintes do índice *SSB World Primary Market (PMI)*, esse índice é um sub-índice do índice *SSB World Broad Market Index (BMI)*, que é um índice de desenvolvimento constituído de índices de mercado de 23 países.

Os autores encontram evidências empíricas da importância dos fatores setor, país e estilo na construção do *portfolio* global. Argumentam que os benefícios potenciais do retorno provenientes da diversificação no modelo baseado em setor tiveram dramático aumento na década de 90 e foi comparável ao efeito país no primeiro semestre de 2001. Outro resultado importante é que o efeito estilo permanece substancial no país enquanto o efeito setor foi reduzido. A dramática divergência entre a *performance* dos estilos (crescimento e valor) desde 1999 a 2001 não pode ser explicada pelo efeito setor. Isto é visivelmente independente dos efeitos valor e crescimento. O mesmo é verdade para o efeito tamanho.

Quanto ao modelo e significância dos fatores puros no retorno investigado encontram claras evidências:

- i) significância estatística dos fatores das quatro categorias: país, setor, crescimento/valor e tamanho;
- ii) um aumento da significância do fator setor e decréscimo da significância do fator país.

Por fim, concluem que o retorno proveniente da diversificação via setor é guiado por momentos, enquanto retorno da diversificação via país não, sendo isto um forte indício da superioridade da alocação por setor sobre a alocação por país.

Portanto, diante desses resultados essenciais, os administradores de ativos não devem estruturar seus *portfolios* tomando por base setores. E salientam que somente um modelo multidimensional é o modelo apropriado, pois todos os quatro fatores: país, setor, crescimento/valor e tamanho, representam fatores significantes de risco e devem ser explicitamente tomados no cálculo, monitorados e administrados no *portfolio* global.

Cavaglia *et al.* (2000) buscam verificar se o efeito país que historicamente foi dominante em explicar a variância do retorno do *portfolio* global em mercados desenvolvidos, ainda persiste. Cavaglia *et al.* (2000) analisam os dados do universo de 21 países que compõem o índice Europa *Morgan Stanley's Capital International (MSCI)*. Para o cálculo do retorno utilizam o índice *Financial Times / Standard & Poors (FT/S& P)* num nível de trinta e seis indústrias nacionais para medir o desempenho do *portfolio* seguro composto das mesmas indústrias dentro do país. Os autores argumentam que o estudo de Rouwenhorst (1999) fornece estimativas imprecisas do efeito setor no *portfolio* global, devida a sua análise restrita a países da Europa e da sua classificação industrial adotada.

Esses autores argumentam que a diversificação por país sugerida por Rouwenhorst (1999) pode ser consequência da crescente integração, nos modelos regionais dos mercados na Europa. Utilizam o índice *FT/S & P* e justificam seu uso por este apresentar um nível de trinta e seis indústrias, ou seja, apresenta uma classificação mais homogênea do que a do índice escolhido por Rouwenhorst (1999), o qual apresenta uma forma mais abrangente de classificação, sete grandes categorias industriais. Por exemplo, na classificação utilizada por Rouwenhorst (1999), "bens de consumo e serviços" agrega categorias de indústria como automóvel, saúde e cuidados pessoais, computadores e softwares indústrias. Enquanto na análise de Cavaglia *et al.* (2000), estas atividades econômicas distintas são tratadas como empresas em setores separados.

Cavaglia *et al.* (2000) usam o modelo fatorial num período de abrangência de dezembro de 1985 a novembro de 1999. O estudo sugere que o fator indústria é economicamente mais importante do que o fator país. Mostram evidências empíricas da crescente importância do fator indústria relativas à capacidade econômica em relação ao fator país na determinação do retorno seguro.

Dentre as implicações para a gestão de *portfolios* fazem a seguinte argumentação: quanto à análise do desempenho de um administrador de *portfolios com foco na alocação* em indústrias deve ser observada a tendência do *benchmark* no comprometimento do mercado doméstico, uma vez que poderá resultar na ineficácia do desempenho do administrador quando da alocação em ativos globais, como por exemplo, o mercado britânico. Esse mercado tem um pequeno comprometimento com a indústria de tecnologia da informação (1,5%) em comparação com o mercado mundial (11,3%).

Uma tendência para análise do desempenho do administrador de *portfolio* em relação à alocação de ativos no mercado doméstico britânico seria explicitamente em direção oposta à análise do desempenho do administrador de *portfolio* em relação à alocação global das

indústrias de tecnologia da informação. Assim, evidências empíricas do estudo sugerem que uma tendência materialmente impacta o risco/retorno do investidor.

Logo, administradores ativos em relação ao investimento de ativos globais terão cada vez mais necessidade de equilibrar o risco/retorno na alocação da indústria global, além da alocação por país. Por último, a seleção do ativo, cada vez mais, deve residir nas comparações através de países, porém no âmbito de indústrias globais.

Com o objetivo de verificar se, em mercados desenvolvidos, o efeito país ainda permanece dominante Baca *et al.* (2000) analisam a variância do retorno do *portfolio* global dos ativos nos setes maiores mercados de ações (G7, sete maiores mercados como definido pelo Mercado de Capitalização em 31 de março de 1999). Esses autores argumentam que na maioria das estratégias de investimentos globais, os investidores usam como nível de investimento o limite nacional ou regional para alocação de recursos, supondo implicitamente que o benefício da diversificação internacional vem do investimento em mercados nacionais, para tanto alocam os recursos globais por país. Entretanto, essa análise é secundária na composição por setor do *portfolio* global. Argumentam também que a diversificação por setor raramente é vista nos termos de sua habilidade de adicionar valor ou controlar o risco dentro de um *portfolio* global. E, à medida que a economia global e o mercado de capital se mostram integrados, esse fato passa a interferir na relação dos efeitos país e setor com o retorno do *portfolio* global, pois altos níveis de integração obscurecem as fronteiras nacionais, diminuem o significado do efeito país e aumentam o significado do efeito setor. E uma maior segmentação entre os mercados aumenta o impacto do efeito país no retorno do *portfolio* à custa do efeito setor.

A amostra do estudo é constituída dos retornos mensais (dólar/EUA), no período de 31 de março 1979 a 31 de março de 1999, sendo usado o índice global das ações da *Datastream*, foram estudados 10 índices setoriais dentro de cada um dos países do G7. No método de estudo, com o intuito de fazer uma determinação precisa da importância relativa dos fatores país e setor, comparam os índices países sem as polarizações do setor e vice-versa. Assim, os componentes do índice de retorno foram decomposto em país e setor, usando a mesma estrutura de Heston e Rouwenhorst (1995) e como a estendida por Griffin e Karolyi (1998) aplicando o seguinte modelo da regressão:

$$R_i = \alpha + \beta_1 I_{i1} + \beta_2 I_{i2} + \dots + \beta_{10} I_{i16} + \gamma_1 C_{i1} + \gamma_2 C_{i2} + \dots + \gamma_7 C_{i7} + e_i \quad (11)$$

onde:

R_i = retorno das observações

α = uma constante

β = componente “puro” de setor

γ = componente “puro” de país

C = variável *dummy*, 1 se for país ou setor, 0 se não

I = variável *dummy*, 1 se for país ou setor, 0 se não

Devido ao problema desse modelo matemático já apontado anteriormente, no estudo de Heston e Rouwenhorst (1995), cada ativo pertence a ambos, país e setor, o que torna indeterminada a regressão. Propuseram então, medir os retornos em relação a uma marca de nível, dada pelo termo constante, representado pelo índice ponderado de capitalização mundial (EW) e adicionam as seguintes limitações ao modelo:

$$\begin{aligned} \sum_{i=1}^{10} w_i \beta_i &= 0 \\ \sum_{k=1}^7 v_k \gamma_k &= 0 \end{aligned} \tag{12}$$

onde:

w_i = ponderação do setor no *portfolio* de mercado mundial

v_k = ponderação de país no *portfolio* de mercado mundial

Estes autores concluíram que o efeito país já não domina o efeito setor para explicar a variância do retorno do *portfolio* global nos sete principais mercados acionários do mundo, pois nos vinte anos analisados a influência do fator país declinou enquanto o impacto do fator setor aumentou ou permaneceu constante. Sugerem diversas razões para suas conclusões, tais como, o número e o tamanho dos mercados de ações analisados, a extensão da classificação das indústrias, o período de estudo e a tendência da continuação da integração global dos importantes mercados econômicos.

Apontam alguns pontos que justificam o deslocamento para o efeito setor tais como convergências políticas, monetárias, fiscais, e a expansão iminente de grandes companhias

multinacionais, redução gradual, mas implacável das barreiras do comércio internacional e investimento.

Apesar do modelo de regressão do método de Heston e Rouwenhorst (1995) ser base de modelo para vários estudos (GRIFFIN e KAROLYI, 1998; ROUWENHORST, 1999; BACA *et al.*, 2000; FLAVIN, 2004), De Moor e Sercu (2006) levantam considerações críticas em relação ao método de Heston e Rouwenhorst (1995).

De Moor e Sercu (2006) argumentam que o método em questão parte da suposição básica da decomposição do retorno dos ativos em fator de mercado, fator “puro” país, fator “puro” setor e fator idiossincrático, todos com estimação única. Fator puro de país (setor) num *portfolio* pode ser interpretado como contendo somente ações específicas de país ou setor na composição setorial (ou geográfica) como no *portfolio* de mercado.

Esses autores argumentam que devido ao progresso tecnológico, acordos de negócios, enfraquecimento das fronteiras econômicas e políticas, os mercados financeiros internacionais tendem a ser mais integrados. Citam como exemplo a EMU, a qual é amplamente vista como tendo enfraquecido a importância relativa do efeito país na diversificação do *portfolio* global de ações em relação ao retorno, o que tem aguçado o debate em relação ao benefício da diversificação geográfica versus diversificação industrial. Argumentam também, que muitos administradores de *portfolios* passaram a defender primeiramente a alocação por setor.

Neste estudo os autores atacam quatro questões remanescentes do método de Heston e Rouwenhorst (1995), o tamanho da capitalização da firma, o modelo empírico especificado, a variação de o retorno estimado ser parcialmente devida à variância de estimação do erro e da inexistência da relação entre os resultados de seus procedimentos e o benefício da diversificação em relação ao risco.

Neste estudo o banco de dados é composto pela lista internacional de pesquisa da *Datastream* que contém todas as cotações de títulos especificadas por país e a lista da *Datastream* “versão morta” contendo todas as ações deslistadas de títulos especificados por país. Eles juntam ambas as listas com o objetivo de compor uma base de dados internacional. Foram considerados 21 países, no período de 1990 a 1999. Usam uma proposta semelhante à de Fama e Mac Beth (1973).

Mais precisamente, comparam os modelos de regressão de Fama e Mac Beth (1973) e o de Heston e Rouwenhorst (1995) estimados por *Ordinary Least Squares* (OLS). Concluem que firmas de pequeno capital têm alta variabilidade média e são menos expostas ao índice de setor global do que firmas de grande capital, ou seja, firmas pequenas são mais voláteis do que as grandes mesmo depois de controlados os efeitos país e setor. Logo, firmas pequenas

reduzem a importância relativa do efeito setor. Tomando por base a própria variância dos fatores, eles classificam os fatores segundo ordem crescente de importância: fatores mercado, país e setor. Porém tomando por base a variância explicada do retorno do *portfólio* global, com estimações dissimilarmente distribuídas através dos fatores, a ordem de importância se altera.

Eles generalizam a metodologia de Heston e Rouwenhorst (1995) adicionando a captura na estimação desigual da distribuição ao longo dos efeitos país e setor. Justificam a importância de a estimação ser desigual em relação aos efeitos setor e país devido à variância explicada por país e setor ser desigualmente transferida para o retorno das ações. Mostram também, que, nas estimações, o impacto relativo ao efeito país aumentou, ou seja, ignorando as desigualdades das estimações país e setor fica subestimado a importância relativa do efeito país. Em relação à variância da variável estimada ser parte da variância estimada do erro eles mostram que a correção do erro aumenta significativamente a importância relativa ao efeito país, e ignorando a estimação do erro o impacto relativo ao efeito país é depreciado.

Por fim, mostram que não existe necessariamente uma ligação entre os resultados do procedimento de Heston e Rouwenhorst (1995) e os benefícios da diversificação no risco. E apesar do efeito país ter maior impacto do que o efeito setor na variância do retorno individual do ativo, a diversificação por setor é uma ferramenta mais efetiva na redução do risco do que diversificar através de país.

A seguir mostra-se esta mesma questão discutida até agora, porém sob a ótica do método análise de *cluster*. A literatura mostra que o método análise de *cluster* para este tipo de discussão foi usado, entre outros, por King (1966), Meyers (1973) e Livingston (1977). Estes autores usam a análise de *cluster* nos retornos dos ativos das companhias dos EUA e encontram que companhias escolhidas numa limitada categoria de indústrias agrupam-se naturalmente via indústria.

Farrel (1974, 1975) e Arnott (1980) também usam a análise de *cluster* em um amplo grupo de companhias dos EUA, não agregadas a uma limitada categoria, sendo as indústrias agrupadas dentro de quatro ou cinco clusters que foram interpretados como amplos grupos setoriais.

Estes estudos mostram que a análise de *cluster* é capaz de recuperar grupos significativos de companhias, e expandem a estrutura da análise de *cluster* na arena internacional com vistas à questão sobre a influência da diversificação por país versus setor no retorno do *portfolio* global.

O estudo de Clifford (2002) trata dessa questão, com foco no seguinte problema de pesquisa: qual a maior influência no retorno do *portfolio* global o país de incorporação das indústrias ou o setor de sua atividade?

O autor justifica o uso do método análise de *cluster* devido ao fato de que a maioria dos métodos estatísticos tem características pré-determinadas para a formação dos grupos de retornos das companhias, porém a análise de *cluster* agrupa as companhias naturalmente, sem estatísticas prévias. O retorno das ações das companhias é usado para determinar matematicamente grupos de companhias semelhantes, descritos em relação ao país de incorporação e ao setor da principal atividade da companhia.

A amostra de estudo é constituída de dados de 4748 companhias, incorporadas em 18 países e os setores de atividade principal classificado segundo o esquema *Financial Times/Goldman Sachs* (FTGS), nos períodos de 1989 a 1993, 1992 a 1996 e de 1995 a 1999.

O autor apresenta quatro passos para processar o método da análise de *cluster*:

- i) cálculo e normalização dos retornos mensais ajustados (*splits* e dividendos);
- ii) cálculo da distância entre os *clusters*;
- iii) recuperação dos *clusters* pela matriz de ligação;
- iv) avaliação dos resultados da classificação.

A medida de distância usada é a de *Pearson*, a qual será detalhada na seção 3.7.1 e, para a aglomeração, o algoritmo *ward*. Os resultados indicam que os três clusters formados contêm companhias provenientes de muitos países e setores. Entretanto, na análise via setor, a maior parte das companhias ficou dispersas dentro dos três clusters, enquanto na análise via país à maioria das companhias caíram dentro de um mesmo *cluster* (*cluster1*). Estas evidências foram constatadas durante os três períodos do estudo, sugerindo um agrupamento naturalmente melhor das companhias por país do que por setor.

O autor enfatiza que os investidores internacionais devem continuar a focar na alocação de ativos por país e exemplifica que mesmo os países da EMU agruparam-se num percentual crescente no *cluster 1*, nos períodos da amostra, na análise via país. As companhias japonesas, no entanto, agruparam-se dentro do *cluster* (3) diferente das demais companhias dos outros países que nas 52 das 54 possibilidades de combinações agruparam-se no *cluster 1*, então chamam a atenção para o fato de que os resultados para um mesmo país podem se alterar com o decorrer do tempo.

Por outro lado, Boillat *et al.* (2002) objetivam agrupar setores como bases de um *potfolio*, dentro de um pequeno número de *clusters* que mostrem flutuações similares de retorno. É também objetivo do estudo validar os resultados sob o ponto de vista financeiro lançando duas questões: em qual extensão o método análise de *clusters* é melhor do que qualquer outro? E como sua efetividade pode ser demonstrada em termos financeiros?

A amostra é constituída dos retornos mensais dos índices de setor da *Datastream* E.U., no período de 1990 até o começo do ano 2002. Os autores apresentam duas análises distintas, na primeira mostram os benefícios do método, através do critério média-variância, na redução do tamanho do universo de investimento para a seleção dos melhores representantes do mercado. A segunda mostra o grau de estabilidade, em termos do *trade-off* entre risco e retorno, dos índices criados para cada *cluster*, ou seja, pela estabilidade dos *clusters* estimam quanto relevante pode ser a segmentação via análise de *cluster* quando aplicada à amostra de observações que são distintas das usadas para se obter estes *clusters*.

Dentre as medidas de distância eles escolheram a medida euclidiana. Os autores argumentam que a escolha da medida da distância tem um grande efeito na formação do *cluster*. Como técnica de ligação eles usaram o método *ward*, o qual é apropriado para manipulação de série temporal relativamente fechada, tal como o retorno de dados dos vários setores econômicos.

Como resultados encontram que o critério de média-variância se mostra relevante do ponto de vista econômico para segmentar os objetos via análise de *cluster* produzindo grupos homogêneos. Entretanto, enfatizam que os resultados dependem fortemente da maneira como os dados são padronizados, do período de observação, frequência dos dados, da noção de distância e do método de ligação. Mostram também que houve uma transição do *cluster* geográfico para o *cluster* de setor nos últimos anos, isto é, o efeito geográfico foi substituído pelo efeito setor nos últimos anos.

Finalmente, eles enfatizam que a técnica análise de *cluster* desmente o fato de se fazer uma efetiva classificação dos objetos com a necessidade a priori de hipóteses estatísticas. Outro ponto que salientam é que o método permite selecionar um objeto representante de uma população, mesmo não fornecendo uma solução acabada, pois as informações adicionais obtidas da análise econômica tradicional devem ser consideradas lado a lado. Mas de qualquer modo, a técnica análise de *cluster* se presta como uma técnica fácil de alocação de ativos pela criação de grupos homogêneos.

Na linha de formação de carteiras de investimento apresenta-se o estudo de Silva *et al.* (2007) cujo objetivo é construir carteiras hipotéticas de ações com características homogêneas de risco versus retorno pautando-se na técnica análise de *cluster*.

A amostra é constituída de dados históricos das ações negociadas na BOVESPA no período de 1998 a 2002, sendo a estabilidade dos *clusters* verificada no período de 2002 a 2006. As variáveis integrantes do estudo foram subdivididas nas categorias: retorno, risco, tamanho da empresa, *cheapness* (relação entre o valor atribuído pelo mercado e a variável financeira da empresa), estrutura de capital e setor. Com relação ao método, a medida de distância escolhida foi à euclidiana e dentre os processos de aglomeração usaram o método hierárquico aglomerativo com regra *ward*; como medida de correlação usaram o coeficiente de *Pearson*. Os resultados mostram que as carteiras formadas mantiveram um desempenho percentual em linha com o esperado, no entanto, no período de 2002 a 2006 as taxas de retornos esperadas eram muito inferiores às realizadas, mas segundo os autores este aumento de diferença no retorno esperado justifica-se pelo cenário econômico brasileiro e mundial.

2.4 Análise econômico-financeira

Quanto à análise das empresas, de um modo geral, pode-se dizer que devido às atuais tendências do mundo econômico impostas pela globalização, ganham importância a valorização do empreendimento e a competência em atuar em mercados mais concorridos e exigentes. Assim, tanto as decisões empresariais com o objetivo de identificar as melhores oportunidades de investimento e financiamento quanto os investidores em geral são guiados essencialmente pelos agentes de mercado, onde a informação sobre as posições passadas, presentes e futuras (projetada) da empresa tornam-se fundamentais.

Para Assaf Neto (2002) a análise de balanços (ou demonstrativos financeiros), que é baseada em informações contábeis, visa relatar a posição econômico-financeira da empresa, sendo essa análise, bem mais consistente quando interpretada dentro das características do setor de atividade no qual a empresa está inserida. Esse autor enfatiza que a propriedade comparativa da análise de balanços é essencial, pois determinado índice, quando avaliado isoladamente, não produz informações suficientes para uma correta conclusão, sendo indispensável o conhecimento do nível em que a empresa se situa em relação as concorrentes e aos padrões do mercado.

Existem vários agentes econômicos, acionistas, fornecedores, administradores, banqueiros, governo, concorrentes e funcionários interessados nas demonstrações financeiras

da empresa, sendo esses interesses distintos. Por exemplo, a taxa de retorno interessa mais especificadamente aos acionistas, o grau de liquidez (curto-prazo) interessa aos fornecedores, o pagamento dos empréstimos preocupa mais aos banqueiros e os pontos fortes e fracos da empresa, aos concorrentes.

Dentre as técnicas de análise de balanços, Assaf Neto (2002) apresenta a análise horizontal a qual se relaciona com o crescimento, avaliação da evolução das vendas, aumento dos investimentos realizados nos itens ativos; a análise vertical que relaciona basicamente as tendências da empresa; e os indicadores econômico-financeiros que relacionam elementos afins das demonstrações contábeis de modo a extrair conclusões sobre a empresa, usando índices padrões do setor de atividade e do mercado para compará-la com outras empresas.

Em relação aos índices-padrão, Matarazzo (2003) argumenta que eles permitem comparar as empresas com outras semelhantes. Entretanto, aponta questionamentos como de que maneira se pode separar as empresas em ramos de atividades e o que é exatamente um ramo de atividade, sendo que estas questões referem-se à construção dos índices-padrão. Complementando, que para a construção dos índices-padrão não deve ser tomado simplesmente os ramos tradicionais de atividades sendo necessária a escolha de ramos que envolvam empresas semelhantes também do ponto de vista financeiro, pois dessa forma, se pode comparar os diversos ramos de atividades, já que os mesmos farão parte de um conjunto (setor) que abrangerá todos os ramos industriais, conseqüentemente, se pode comparar o desempenho da indústria, comércio e serviços entre si.

Para Assaf Neto (2002), Matarazzo (2003), Ludícibus (1998) os índices padrões classificam-se, basicamente, em índices de liquidez; rentabilidade; estrutura/endividamento; e indicadores de preço e retorno das ações.

A idéia central de criar medidas de liquidez está na necessidade de medir a capacidade de pagamento de uma empresa, ou seja, sua habilidade em cumprir corretamente as obrigações passivas assumidas. Uma grande restrição que se atribui a esses indicadores é a posição de liquidez estática que revelam, isto é, não refletem a magnitude e a época em que ocorreram as diversas entradas e saídas circulantes, pois são extraídos apenas do balanço patrimonial. Na realidade, os indicadores tradicionais de liquidez exprimem uma posição em dado momento de tempo (na data do levantamento dos valores), mas, na realidade, os diversos valores considerados são continuamente alterados em função da dinamicidade natural dos negócios da empresa.

Indicadores de rentabilidade visam avaliar os resultados auferidos por uma empresa em relação a determinados parâmetros que melhor revelem suas dimensões. Uma análise

baseada exclusivamente no valor absoluto do lucro líquido traz normalmente sério viés de interpretação ao não refletir se o resultado gerado foi condizente ou não com o potencial econômico da empresa.

As principais bases de comparação adotadas para o estudo dos resultados empresariais são o ativo total, o patrimônio líquido e as receitas de vendas. Os resultados normalmente utilizados são os lucros operacionais (lucro gerado pelos ativos) e líquidos (após o imposto de renda). Ressalte-se, ainda, que os analistas em geral atribuem grande valor aos indicadores de rentabilidade, os quais costumam exercer significativas influências sobre as decisões que envolvem a empresa em análise, tomada tanto no mercado de crédito, como no mercado acionário.

A finalidade básica dos indicadores de endividamento e estrutura é aferir a composição (estrutura) das fontes passivas de recursos de uma empresa bem como mensurar percentualmente sua relação com o capital próprio. Ilustram, com isso, a forma como o recurso de terceiros é usado pela empresa e sua participação em relação ao capital próprio. Fornecem, ainda, elementos de avaliação do grau de comprometimento financeiro de uma empresa perante seus credores e a sua capacidade de cumprir os compromissos financeiros assumidos a longo prazo.

Conforme Assaf Neto (2002, p.413), “as empresas normalmente procuram não se afastar muito do padrão de endividamento do setor. Com isso, a relação $P/(P+PL)$ é constituída em geral, pelos diferentes segmentos de atividade”.

Para as empresas constituídas juridicamente como sociedade por ações, é possível construir indicadores de avaliação do preço das ações e de sua rentabilidade. De modo geral, essa análise é aplicada para as ações de sociedade anônima de capital aberto, ou seja, aquelas que têm suas ações negociadas na bolsa de valores.

Esta análise das ações objetiva avaliar os reflexos do desempenho da empresa sobre suas ações e são de grande utilidade para os analistas de mercado e acionistas (potenciais e atuais), como parâmetros de apoio às suas decisões de investimentos.

De forma idêntica, as cotações de mercado são freqüentemente sensíveis aos resultados desses indicadores, podendo-se estabelecer inclusive tendências futuras sobre a sua maximização.

A seguir apresentam-se os principais indicadores com os respectivos índices de medidas e atributos.

INDICADORES	MEDIDAS	ATRIBUTOS
Liquidez	Índice de liquidez corrente	Relaciona todos os ativos realizáveis no curto prazo com todos os passivos que deverão ser pagos no curto prazo
	Índice de liquidez seca	É uma liquidez de curto prazo excluindo-se os estoques
Rentabilidade	Margem operacional	Valor financeiro da conta do lucro operacional dividido pelo valor financeiro da conta receita líquida operacional
	Margem ebitida	Indica o valor financeiro do ebitida dividido pelo valor financeiro da receita bruta
	Rentabilidade do ativo	Valor financeiro do lucro líquido dividido pelo valor financeiro do ativo total
	Rentabilidade patrimônio líquido	Valor financeiro do lucro líquido dividido pelo patrimônio líquido
Estrutura de capital	Exigível total/ativo total	Indica a porcentagem do ativo total financiada com capitais de terceiros a curto e longo prazo
	Endividamento geral	Verificar a garantia de pagamento das dívidas com recursos próprios
	Endividamento financeiro	Faz a relação entre capital de terceiros e capital próprio
Preço e Retorno das ações	Valor patrimonial da ação	O valor é representado por uma avaliação a valores contábeis
	Taxa de retorno	A taxa de retorno é uma forma de padronização considerando a quantia de retorno por unidade de investimento
Dividendo	Dividendo por ação <i>Payout</i>	Soma dos dividendos em dinheiro pagos por ação no ano Valor em dinheiro dos dividendos propostos na demonstração das origens e aplicações dos recursos dividido pelo lucro líquido

Quadro 01- Indicadores com suas medidas e atributos respectivamente

Outro tema que traz de forma latente a classificação setorial, dentro da análise econômico-financeira refere-se à avaliação econômica e medida de desempenho das empresas, que visam identificar a criação de riqueza pelas empresas. Por exemplo, o estudo de Assaf Neto (2004) cujo objetivo é pesquisar o desempenho das empresas brasileiras, e avaliar se as mesmas foram competentes para atender ao objetivo básico da teoria de finanças, maximizar a riqueza de seus acionistas. Esse estudo desenvolve e aplica as principais medidas de criação de valor nas companhias brasileiras com destaque para os setores de atividade de melhores desempenhos, o autor propõe um método de aplicação dos instrumentos de avaliação econômica a partir da classificação setorial e dos demonstrativos financeiros, para o estabelecimento dos padrões de desempenho das empresas brasileiras.

O estudo centrou-se no universo das companhias brasileiras de capital aberto, conforme disponibilizadas pela ECONOMÁTICA, a amostra é constituída por companhias de capital aberto atuantes no período de 1996-2002. As medidas calculadas para cada empresa foram agrupadas por setor para cada um dos sete anos considerados do estudo. Dentre os resultados, foram apurados, em cada um dos anos delimitados do estudo, para cada empresa e setor de atividade, importantes medidas financeiras como: resultados operacional e líquido, índice de endividamento oneroso, fluxo de caixa operacional disponível e retorno sobre o patrimônio líquido. O autor destaca o comportamento dos índices apresentados por cada setor de atividade, além de apontar e discutir suas principais causas.

2.5 Grupos de controle

De acordo com Sampieri *et al.* (2006) com o objetivo de responder às questões de pesquisa propostas, deve-se escolher ou desenvolver um modelo específico para cada pesquisa, ou seja, desenvolver um plano ou estratégia para se obter informações necessárias para uma determinada pesquisa. Os modelos de pesquisa podem ser experimentais ou não-experimentais. Os experimentais geralmente são usados sob o enfoque quantitativo, uma vez que sob o enfoque qualitativo a pesquisa “foge” de estratégias prévias no sentido de que podem contaminar os dados puros dos indivíduos da pesquisa em seu contexto. Os não-experimentais consideram o fator tempo ou número de vezes de coleta dos dados.

Considerando o modelo experimental, define-se experimento como “uma situação de controle na qual se manipulam, de maneira intencional, uma ou mais variáveis independentes/causas para analisar as conseqüências dessa manipulação sobre uma ou mais variáveis dependentes/efeitos” (SAMPIERI *et al.*, 2006, p. 159).

Os experimentos podem ser realizados em dois ou mais graus de manipulação da variável independente, sendo que cada grau requer um grupo no experimento. O nível mínimo de manipulação da variável independente é dois, presença-ausência dessa variável. Na presença-ausência um grupo é exposto à presença da variável independente (grupo experimental) e o outro não (grupo de controle), a seguir compara-se os dois grupos verificando se existem diferenças entre os grupos. O termo controle tem diversas conotações dentro da experimentação, mas sua acepção mais comum é quando se observa no experimento que uma ou mais variáveis independentes provocam variações nas variáveis dependentes. A variação nas variáveis dependentes deve-se à manipulação, e não a outros fatores ou causas (SAMPIERI *et al.*, 2006).

Em termos mais coloquiais, ter controle significa saber o que está ocorrendo realmente com a relação entre as variáveis independentes e dependentes. Na estratégia da pesquisa experimental, o pesquisador não manipula uma variável apenas para comprovar o que ocorre com a outra, assim, ao realizar um experimento, é necessário fazer uma observação controlada (VAN DALEN E MEYER, 1994).

Na formação de grupos de controle, a classificação setorial é utilizada para a seleção de uma empresa similar aquela do objeto de pesquisa, mas sem a característica em estudo. Como exemplos de utilização da classificação setorial em grupos de controle apresentam-se os estudos de Beaver (1966), Famá *et al.* (2003).

Beaver (1966), com o objetivo de identificar os índices financeiros que predizem a probabilidade de falência de uma empresa, utiliza dados de 158 empresas, classificadas como 79 falidas e 79 como não falidas, através de janelas de cinco exercícios contábeis antes da falência, para cada empresa falida foi selecionada uma empresa não falida do mesmo setor e de porte equivalente. As empresas do estudo possuíam ações negociadas no mercado da bolsa norte americana no início da segunda metade do século passado, num período de estudo de 10 anos, de 1954 a 1964. Os principais índices utilizados na análise foram de fluxo de caixa, rentabilidade, endividamento, liquidez e atividade (giro), totalizando 30 índices. Dentre os resultados mostra que os índices de fluxo de caixa sobre dívida, retorno sobre ativos e endividamento sobre ativos, foram os que melhor previram a falência.

No Brasil, Famá *et al.* (2003) usam a mesma abordagem de Beaver (1966) com o objetivo de avaliar o potencial preditivo dos índices financeiros sobre a situação da empresa em honrar seus compromissos. A amostra é formada por 48 empresas brasileiras, exceto financeiras, de capital aberto com demonstrativos financeiros publicados, de acordo com as normas da Comissão de Valores Mobiliários (CVM) e com ações negociadas em bolsa, no período de 1990 a 2001. Para cada empresa considerada como falida foi selecionada uma não falida do mesmo setor econômico e de porte equivalente, formando pares de falidas e não falidas, da mesma maneira como foi feita na abordagem de Beaver (1966). Os autores justificam que as empresas foram agrupadas por setor e porte, porque os índices financeiros são determinados pelos fatores específicos de um setor e um mesmo valor de índice pode significar probabilidades diferentes de falência para empresas que atuam em diferentes setores. Nesta abordagem, diferentemente da de Beaver (1966), os índices que melhor prediziam a falência foram liquidez corrente, liquidez seca, investimentos sobre patrimônio líquido, liquidez geral e rentabilidade do ativo.

Num exemplo de segmentação da amostra em subgrupos apresenta-se o estudo de e Procianoy e Heineberg (2003) cujo objetivo é identificar os aspectos determinantes da política de proventos das empresas brasileiras com ações negociadas na BOVESPA no período entre 1994 e 2000. Argumentam que setores diferentes possuem diferentes necessidades de recursos e possuem distintas possibilidades de financiamento para seus novos investimentos e capital de giro. Portanto, setores com maior necessidade de recursos, com novos investimentos e com ciclos de giro com sazonalidades, podem utilizar os lucros gerados para realizar o financiamento necessário, o que poderá impactar na política de distribuição dos proventos aos acionistas.

Como técnica estatística eles usam a análise de regressão com dados em painel, abrangendo uma amostra final de 196 empresas, inseridas em 27 setores industriais. A análise dos resultados via segmentação por setor permitiu uma comparação entre os resultados obtidos para cada setor industrial, especificando mais precisamente as diferenças existentes entre as políticas de distribuição de proventos dos setores mais representativos na pesquisa. O objetivo dessa análise é verificar para quais setores as variáveis independentes trabalhadas explicam melhor a determinação dos proventos, e se existe uma homogeneidade no comportamento da amostra estudada. Assim, a partir da análise desta segmentação, segundo os autores, ficou evidente a existência de diferentes padrões da política de proventos entre os setores industriais mais representativos da amostra do estudo.

3. TIPO E MÉTODO DE ESTUDO

Neste capítulo são apresentados os pressupostos e as opções metodológicas que norteiam o presente estudo.

3.1 Abordagens da pesquisa

Esta pesquisa fundamenta-se no pensamento de Seltiz *et al.* (1959) quando argumentam que a finalidade da pesquisa é descobrir respostas para as perguntas por meio da aplicação de procedimentos científicos, sendo mais provável disso acontecer assim, do que por qualquer outro método conhecido pelo homem, uma vez que estes procedimentos foram desenvolvidos para aumentar a probabilidade de confiança da informação obtida, sem distinção às perguntas formuladas.

De acordo com os objetivos propostos para este estudo, percebe-se seu caráter descritivo, pois “os fatos são observados, registrados, analisados, classificados e interpretados, sem que o pesquisador interfira neles” (ANDRADE, 1997, p.104). Ainda nessa linha tem-se a visão de Sampieri *et al.* (2006) que definem os estudos descritivos como aqueles que buscam especificar propriedades importantes de pessoas, grupos, comunidades ou qualquer outro fenômeno que seja submetido à análise. Quanto a escolha do procedimento sistemático para a descrição e explicação de fenômenos que estão sendo pesquisados percebe-se seu caráter quantitativo caracterizado por Richardson *et al.* (1985) pela quantificação na coleta e no tratamento das informações, lançando mão basicamente das técnicas estatísticas, desde as mais simples até as mais complexas. Complementando esta visão, Horna (1994) diz que os projetos de pesquisa quantitativos são caracterizados pela suposição de que o comportamento humano pode ser explicado, pelos “fatos sociais” que podem ser investigados pelas metodologias que utilizam a lógica dedutiva das ciências naturais. Assim, esta é uma pesquisa descritiva, eminentemente quantitativa e objetivamente incrementalista, na medida em que busca relações entre as variáveis com a finalidade de analisar uma observação empírica.

3.2 População amostral e período de estudo

Para a amostra desse estudo foram consideradas todas as empresas de capital aberto que tenham ações negociadas na BOVESPA, no período compreendido entre 02 de

janeiro/2000 e 31 de dezembro/2007, com alguns critérios de seleção, tais como balanços publicados no período do estudo e dados disponíveis no banco de dados da empresa ECONOMÁTICA Software para Investimentos Ltda, sendo que a classificação NAICS nível 1 apresenta 20 setores e a classificação BOVESPA segmenta-se em 10 setores.

A escolha da bolsa em referência decorre do fato de ser o local na atualidade, que concentra uma maior quantidade de empresas e um maior volume diário de transações no Brasil. Com relação ao período de tempo escolhido, salienta-se que em, tal período, não houve fatos relevantes que pudessem provocar alterações consideráveis no mercado brasileiro, apesar do período de transição governamental (2002), que segundo o IBGE (2007) não provocou alterações significativas que pudessem afetar a estabilidade econômica vigente.

3.3 Coleta de dados e definição das variáveis

Os dados da pesquisa são secundários e foram coletados a partir do banco de dados da ECONOMÁTICA, declarados pelas companhias brasileiras entre 02/01/2000 e 31/12/2007.

Dentre os indicadores usados no presente estudo, salienta-se que, para os indicadores de tamanho, optou-se por utilizar as variáveis logaritimizadas, assim, as variáveis valor de mercado, receita líquida e ativo total tornaram-se \ln (valor de mercado), \ln (receita líquida) e \ln (ativo total). As variáveis médias, por sua vez, são expressas pela média aritmética simples de todos os dados da variável num dado período. Por exemplo, a variável média retorno é expressa por:

$$V_m RET = \frac{RET_{i,00} + \dots + RET_{i,07}}{8}, i = \text{ação} \quad (13)$$

A seguir mostra-se as variáveis usadas com suas respectivas medidas dispostas segundo os indicadores de liquidez, rentabilidade, estrutura de capital, mercado e tamanho.

Indicadores	Medidas	Fórmula
Liquidez	Índice de liquidez corrente (<i>LC</i>)	$LC = \frac{\text{ativo circulante (AC)}}{\text{passivo circulante (PC)}}$
	Índice de liquidez seca (<i>LS</i>)	$LS = \frac{\text{ativo circulante} - \text{estoque (E)}}{\text{passivo circulante}}$
Rentabilidade	Margem operacional (<i>MO</i>)	$MO = \frac{\text{lucro operacional (LO)}}{\text{vendas}}$
	Margem ebitida (<i>ME</i>)	$ME = \frac{\text{ebitida}}{\text{receita operacional líquida (ROL)}}$
	Rentabilidade do ativo (<i>RAT</i>)	$RAT = \frac{\text{lucro operacional}}{\text{ativo operacional (AO)}}$
	Rent. patrim.líquido(<i>RPATL</i>)	$RPATL = \frac{\text{lucro líquido exercício (LLE)}}{\text{patrimônio líquido (PL)}}$
Estrutura de capital	Exigível total/ativo total (<i>ETAT</i>)	$ETAT = \frac{\text{exigível total (ET)}}{\text{ativo total (AT)}}$
	Endividamento geral (<i>EG</i>)	$EG = \frac{\text{passivo circ} + \text{exigível longo prazo (ELP)}}{\text{patrimônio líquido}}$
	Endividamento financeiro (<i>EF</i>)	$EF = \frac{\text{empréstimo} + \text{financiamento (F)}}{\text{patrimônio líquido}}$
Mercado	Valor patrimonial da ação (<i>VPA</i>)	$VPA = \frac{\text{patrimônio líquido}}{\text{número de ações em circulação (NAC)}}$
	Liquidez em bolsa (<i>LIQB</i>)	$LIQB = 100 \cdot \frac{P}{P} \cdot \sqrt{\frac{n \cdot v}{N \cdot V}}$ <p><i>p</i> = número de dias em que houve pelo menos um negócio com a ação; <i>P</i> = nº total de dias do período escolhido; <i>n</i> = nº negócios com a ação dentro do período escolhido; <i>v</i> = volume em dinheiro com a ação; <i>V</i> = volume em dinheiro com todas as ações dentro do período escolhido.</p>
	Retorno	$R_{it} = \frac{P_{t+1} - P_t}{P_t}$ <i>P</i> _{t+1} = preço da ação em t+1; <i>P</i> _t = preço da ação no período t
Tamanho	Valor de mercado (<i>VMER</i>)	$VMER = PMA \cdot N$ <i>PMA</i> = preço de mercado das ações; <i>N</i> = quantidade total de ações
	Receita líquida (<i>RL</i>)	<i>RL</i> = valor financeiro da conta receita líquida
	Ativo total (<i>AT</i>)	<i>AT</i> = valor financeiro da conta ativo total

Quadro 02 - Variáveis e as suas respectivas medidas, dispostas segundo os indicadores de liquidez, rentabilidade, estrutura de capital, mercado e tamanho

Os dados das variáveis foram coletados do Economática, conforme as considerações abaixo.

INDICADORES	VARIÁVEIS	PARÂMETROS E AJUSTES
Liquidez	Índice de liquidez corrente Índice de liquidez seca	Dados do balanço consolidado referente ao 4º trimestre do ano calendário (entre 8 de outubro a 7 de janeiro do ano seguinte). Entretanto, se os dados consolidados não estavam disponíveis para o período em questão foram tomados os dados não consolidados.
Rentabilidade	Margem operacional Margem ebitida Rentabilidade do ativo Rent. patrim. líquido	
Estrutura de capital	Exigível total/ativo total Endividamento geral Endividamento financeiro	
Dividendos	Dividendo <i>yield</i> <i>Payout</i>	Dados coletados em 31/12 de cada ano do estudo, em moeda original. Dados do balanço consolidado referente ao 4º trimestre do ano calendário
Mercado	Valor patrimonial da ação Liquidez em Bolsa Retorno	Dados do balanço consolidado referente ao 4º trimestre do ano calendário. Dados ajustados por desdobramentos e coletados na moeda original. Na janela de 02/01/ à 31/12 de cada ano e coletados na moeda original. Cotações de fechamento em 02/01 e 31/12 para cada ano, na moeda original, ajustada por proventos inclusive dividendos. Sendo a variável retorno resultante de cada ano dada pela fórmula mencionada no método.
Tamanho	Valor de mercado	Dados coletados em 31/12 de cada ano do estudo, em moeda original.
	Receita líquida Ativo total	Dados do balanço consolidado referente ao 4º trimestre do ano calendário. Dados coletados na moeda original.

Quadro 03 - Parâmetros e ajustes dos dados das variáveis coletadas

No presente estudo as empresas selecionadas são classificadas nos setores conforme a classificação setorial da BOVESPA e da NAICS nível 1. Tal escolha deve-se ao fato destas classificações levarem em consideração que embora a empresa exerça atividades diferentes, ela atua em estágios similares da cadeia produtiva, ou seja, com produtos e serviços relacionados.

3.4 Dados omissos e *outliers*

Inicialmente mostra-se no item 3.4.1 uma explanação sobre dados omissos e na seqüência no item 3.4.2 apresenta-se, com base na literatura, procedimentos para identificar *outliers* sob o ponto de vista multivariado.

3.4.1 Dados omissos

O tratamento para os dados omissos depende basicamente das características dos dados omissos e do método de pesquisa. A utilização do procedimento *listwise*, eliminando todos os casos com pelo menos uma variável sem informação, apresenta a desvantagem da diminuição do tamanho amostral (KLINE, 1998). No procedimento *pairwise* os casos são retirados apenas quando não oferecem informações suficientes para o cálculo de determinado momento amostral. Entretanto, quando o modelo tem três ou mais variáveis a utilização do *pairwise* pode gerar matrizes de covariância e correlação que não sejam positivas definidas, o que significa que certas operações matemáticas com a matriz não poderão ser computadas, prejudicando a análise. Outra opção é a utilização de um modelo de imputação dos dados. Kline (1998) apresenta três caminhos possíveis:

- i*) substituição pela média, que tem a vantagem da simplicidade, não sendo, entretanto, sensível à resposta dos sujeitos nas demais variáveis;
- ii*) regressão, que leva em conta as respostas dos sujeitos nas demais variáveis, não sendo, entretanto, factível se a variável omissa não tem covariância pelo menos moderada com as demais variáveis; e
- iii*) ajuste padrão, que é a substituição pelo valor da variável em um caso com escores semelhantes nas demais variáveis.

Portanto, sendo o *listwise* e o *pairwise* considerados inadequados para o propósito deste estudo, pois reduziria acentuadamente os casos quando da aplicação da técnica análise de cluster, optou-se por utilizar um modelo de imputação dos dados. O número de variáveis com dados omissos e a sua aleatoriedade contribuíram para que se definisse a imputação pela substituição pela média.

3.4.2 Outliers

De um modo geral, pode-se concluir que um *outlier* é caracterizado pela sua relação com as observações restantes que fazem parte da amostra, assim, seu distanciamento em relação a essas observações é fundamental para fazer a sua caracterização. A preocupação com observações *outliers* é antiga, grande partes dos autores que estudavam este assunto reportam-se aos comentários de Bernoulli datados de 1777 (BARNETT e LEWIS, 1994) os quais justificavam a rejeição dos *outliers*. Entretanto, as opiniões não eram unânimes, ou seja, enquanto uns defendiam a rejeição de tais observações outros afirmavam que elas nunca deveriam ser retiradas da amostra. Atualmente, autores como Corrar *et al.* (2007) argumentam que a retirada de *outlier* pode melhorar a análise multivariada, mas limita sua generalidade.

Mais especificadamente a identificação de *outliers* sob o ponto de vista univariado é relativamente de fácil determinação. Conforme sugerem Corrar *et al.* (2007) deve-se a partir da padronização das variáveis considerar como *outliers*, para grandes amostras ($n > 80$) dados com *score* maior ou igual a três ao passo que para pequenas amostras ($n < 80$) dados com *score* maior ou igual a 2,5. No entanto, para o conjunto multivariado de dados, como é o caso, há um acréscimo de dificuldades, uma vez que uma observação pode não ser um *outlier* em nenhuma das variáveis originais estudadas isoladamente e sê-lo na análise multivariada. Assim, conforme sugerido por Johnson e Wichern (2002) para a identificação de *outliers* multivariados utiliza-se a distância de *Mahalanobis* (D^2), a qual tem aproximadamente a distribuição do qui-quadrado com p graus de liberdade.

Assim, a partir da relação cumulativa da probabilidade do valor da distribuição do qui-quadrado com grau de liberdade igual ao número de variáveis incluídas no cálculo, considera-se como um *outlier* multivariado aquele cuja respectiva probabilidade associada com sua D^2 , for menor ou igual 0,001. Mais precisamente, a distância generalizada de *Mahalanobis* pode ser entendida como a versão multidimensional do *z-score*, e pode ser usada como uma técnica de comparação quanto à separação entre diversos grupos permitindo avaliar a extensão e a direção dos afastamentos entre os valores médios das variáveis usadas na discriminação (TABACHNICK e FIDELL, 2000), ou, uma medida de distância entre os casos proveniente do centróide médio, dada uma distribuição de covariância (JOHNSON E WICHERN, 2002).

Nesta pesquisa são identificados os *outliers* do ponto de vista multivariado.

3.5 Estatísticas descritivas

Usa-se na descrição da base de dados, uma vez que as estatísticas descritivas possibilitam a avaliação do comportamento de cada uma das variáveis utilizadas no estudo, sendo tal comportamento avaliado a partir das seguintes estatísticas: média, desvio padrão, coeficientes de assimetria e curtose. As medidas de assimetria e curtose são utilizadas conforme definido por Pestana e Gageiro (2005). A medida de assimetria é dada pelo quociente entre a assimetria e o desvio de erro, onde este quociente é usado para não rejeitar a simetria, o que acontece se o seu resultado for menor que 1,96 (ou aproximadamente 2) em valor absoluto. Em relação à curtose toma-se como o quociente entre a curtose e o desvio de erro, por exemplo, se tal quociente for menor que 1,96 então, a distribuição é mais achatada do que a normal, isto é, platicúrtica.

Para avaliar as diferenças univariadas entre os setores de atividade, propostos pelas classificações BOVESPA e NAICS, utiliza-se o teste de Kruskal Wallis, o qual indicará se as diferenças entre os setores significam genuínas diferenças entre as populações ou se elas representam meramente o tipo de variação que seria esperada entre amostras aleatórias de uma mesma população. A análise de variância de um fator de Kruskal-Wallis por postos é um teste, conforme Siegel e Castellan (2006), extremamente útil para decidir se K amostras independentes provêm de populações diferentes (hipótese nula - mesma população). O teste de Kruskal-Wallis trabalha com as diferenças entre os postos médios para determinar se elas são tão discrepantes que provavelmente não tenham vindo de amostras que tenham sido extraídos de uma mesma população. O cálculo da estatística desse teste é dado por:

$$KW = \frac{12}{N(N+1)} \sum_{j=1}^k n_j (\bar{R}_j - \bar{R})^2 \quad (14)$$

onde:

K = número de grupos;

n_j = número de casos no j -ésimo grupo;

N = número de casos na amostra combinada;

\bar{R}_j = média dos postos no j -ésimo grupo

Para análise da probabilidade segundo este teste usa-se a aproximação à distribuição do qui-quadrado com $(K-1)$ graus de liberdade. Assim se o valor observado de KW for maior ou igual ao valor tabelado do qui-quadrado com nível de significância $\alpha=0,05$ e para o observado grau de liberdade, então se pode rejeitar H_0 neste nível de significância. As hipóteses a testar são:

H_0 = não há diferença mediana entre os K setores;

H_1 = os k setores diferem com relação à mediana.

3.6 Técnica estatística

A principal técnica estatística utilizada nesta pesquisa é a análise de agrupamento ou análise de *cluster* (originalmente *cluster analysis*), que por sua vez, é um conjunto de técnicas que tem por objetivo identificar padrões ao formar *clusters* (grupos) homogêneos e mutuamente exclusivos, isto é, os mais semelhantes pertencem a um mesmo grupo (HAIR *et al.*, 2005). Da mesma forma, Malhotra (2001) define a análise de *cluster* como uma técnica usada para classificar objetos ou casos, em grupos relativamente homogêneos chamados de *clusters*. Segundo essa técnica, a construção dos grupos é feita de modo que as observações de um mesmo grupo se pareçam mais entre si do que com as observações existentes nos outros grupos formados (BUSSAB *et al.*, 1990).

De acordo com estes autores, os *clusters* obtidos devem apresentar tanto uma homogeneidade interna (dentro de cada *cluster*), sendo semelhantes entre si, como grande heterogeneidade externa (entre *clusters*), se diferenciando dos demais. Portanto, se a segmentação for bem sucedida, os objetos dentro dos *clusters* estarão próximos quando representados graficamente e *clusters* diferentes estarão distantes.

A busca por grupos naturais ou homogêneos nos dados de uma estrutura faz da análise de *cluster* uma importante técnica exploratória uma vez que fornece meios informais para avaliações dimensionais e até mesmo identificação de observações destoantes que podem vir a serem valores atípicos (*outliers*), oportunizando ainda o surgimento de hipóteses interessantes nos relacionamentos ao se produzirem taxonomias (JOHNSON e WICHERN, 2002). Nenhuma condição ou suposição é feita sobre o número de *clusters* ou sobre a estrutura de cada um e não há prévia alocação dos casos nos *clusters*.

De acordo com Bergman e Feser (1998) o termo análise de *cluster* foi introduzido primeiramente por Tryon em 1939, contudo, idéias semelhantes já vinham ocorrendo desde o

final do século XIX. Esta técnica vem sendo utilizada em importantes áreas científicas para a identificação de padrões de comportamento nos dados analisados, auxiliando com isso o processo de descoberta do conhecimento, pois a divisão em grupos ou classes facilita a compreensão das observações e o desenvolvimento subsequente de teorias científicas.

Outro ponto de relevância dessa técnica está em sua utilização ainda na fase exploratória da pesquisa, onde a falta de hipóteses a priori sobre as observações e o desconhecido número de grupos permitem que as mesmas auxiliem na organização dos dados em estruturas significativas de fácil interpretação.

Com o intuito de diferenciar a análise de *clusters* de outros métodos cabe ressaltar que, apesar da existência de referências a esse termo, não estamos trabalhando com um método de classificação, a qual, segundo Johnson e Wichern (2002), pressupõem que se conheçam o número de grupos finais e suas características, sendo seu objetivo operacional associar novas observações a um desses grupos. Assim, comparada à classificação a análise de *cluster* apresenta-se como uma técnica mais primitiva onde não existe assunção sobre a existência e característica dos grupos, tampouco sobre a quantidade dos mesmos, pois os agrupamentos são feitos com base nas medidas de distância (coeficientes de similaridade ou dissimilaridade).

Como pré-requisito para aplicar as técnicas de grupamento, além da existência dos dados no conjunto inicial, é necessário definir qual o critério a ser utilizado na definição de proximidade entre os objetos. Por outro lado, na classificação, os objetos são associados a grupos ou classes pré-definidas, e nos agrupamentos, os grupos (quantidade dos mesmos e que indivíduos pertencem a cada um) surgem durante o processo. Portanto, a análise de *cluster* também pode ser utilizada para auxiliar a uma classificação que não tenha sido aplicada ainda devido à falta dos grupos e características dos mesmos. A título de ilustração citam-se como exemplos de classificações, a classificação das estrelas, introduzida por A. J. Cannon em 1910 e aperfeiçoada por volta de 1920, conhecida por Classificação de Harvard que é de suma importância para a astronomia e a classificação geral das galáxias, elaborada por Hubble, em 1927, que ficou conhecida como Classificação Morfológica de Hubble.

Outro ponto que cabe destacar quando se utiliza a técnica análise de *cluster*, é a impossibilidade da técnica de determinar antecipadamente as variáveis dependentes e independentes, uma vez que essa técnica permite examinar as relações de interdependência entre todo o conjunto de variáveis ou casos. Destaca-se, ainda a interdependência como ponto de similaridade com a análise fatorial e a forma de tratamento dos objetos como forma de

diferenciação das mesmas, uma vez que a análise fatorial trabalha somente com as variáveis, com vistas a reduzir o conjunto das mesmas, através da criação de fatores (HAIR *et al.*, 2005).

Diversos autores destacam ainda que as técnicas utilizadas na análise de *cluster* referem-se à aprendizagem não-supervisionada (KRZANOWSKI e MARRIOT, 1995) e dentre estes existem aqueles que ressaltam o fato de não existirem testes de significância estatística envolvidos na sua utilização, mesmo em casos onde existem graus de liberdade (EVERITT, 1974). Entretanto, esses mesmos autores consideram que as técnicas de análise de *cluster* encontram a solução mais “significativa possível”, ao minimizar a variabilidade dentro de cada grupo e maximizar a variabilidade entre os grupos, os quais são objetivos semelhantes existentes nas análises Discriminantes e Fatoriais.

A análise de *cluster* tem aplicabilidade em diversas áreas do conhecimento, com destaque para marketing e, mais recentemente, nas áreas de investimentos, economia e finanças. Além da fácil interpretação, outra razão pela opção da utilização da análise de *cluster* é a não exigência de pressupostos iniciais quanto à distribuição de probabilidade dos dados.

Algumas técnicas estatísticas de análise multivariada são pertinentes ao processo de segmentação de mercado, por permitirem identificar relações simultâneas entre mais de duas variáveis. A escolha do método análise de *cluster* nessa pesquisa deve-se ao fato de ser entre os métodos estatísticos de segmentação o que melhor responde ao escopo do estudo. Outros métodos como, por exemplo, a análise dos componentes principais, não mostra uma diferenciação suficiente dentro do universo de investimento, conseqüentemente torna difícil segmentar o mercado de uma maneira relevante e útil (CLIFFORD, 2002; BOILLAT *et al.*, 2002).

Nesse sentido, Hair *et al.* (2005) argumentam que as opções de estratégias baseadas na identificação de grupos dentro de uma dada população, tal como segmentação do mercado, não seria possível sem uma metodologia objetiva. E ainda, segundo esses autores, a análise de *cluster* é a única técnica multivariada a não estimar a variável estatística empiricamente, ou seja, usa a variável estatística especificada pelo pesquisador. Logo, possibilita comparar os objetos com base na variável estatística e não na estimação da variável estatística em si, assim, a definição da variável estatística é ponto de análise crítica.

Outro ponto de relevância da técnica é sua flexibilidade Lapointe e Legendre (1994) e Toms *et al.*(2001) argumentam que este método permite agrupar objetos de acordo com o grau de similaridade, podendo ser aplicado a uma larga espécie de objetos. A flexibilidade do

método e dos algoritmos que estão associados a ele explica a diversidade de sua aplicação, podendo variar das ciências físicas às ciências sociais.

Mais precisamente, em consonância com o escopo dessa pesquisa, a análise de *cluster* será utilizada para segmentar as ações listadas conforme mencionado no subitem 3.2 do corrente capítulo para posterior análise dessas ações quando classificadas segundo as classificações BOVESPA e NAICS, ou seja, setores convencionais. Assim esta técnica proporciona a esta pesquisa uma forma de correspondência ou não das ações em setores convencionais, ou em setores diferentes dos convencionais. A seguir descreve-se com mais detalhes a técnica análise de *cluster*, justificando vários pontos de relevância para a consolidação do método.

3.6.1 Estágios de aplicação da técnica análise de *cluster*

Nesta técnica multivariada a variável estatística se apresenta de uma forma bastante diferente de outras técnicas multivariadas. A variável estatística do *cluster* é “o conjunto de variáveis que representam as características usadas para comparar objetos na análise de *cluster*” (HAIR *et al.*, 2005, p.384). Como ela incluiu apenas as variáveis usadas para comparar objetos, ela determina o caráter dos objetos.

Malhotra (2001) estabelece etapas para aplicação da análise de *cluster*:

Etapa 1 - a formulação do problema, onde se define com quais variáveis a aglomeração será baseada;

Etapa 2 - a escolha de uma medida de distância, que determina o quão semelhantes ou diferentes serão os grupos formados; a escolha de um processo de aglomeração;

Etapa 3 - a decisão quanto ao número de *clusters*, que exige julgamento por parte do pesquisador;

Etapa 4 - a interpretação e perfilação dos *clusters*;

Etapa 5 - a avaliação quanto à validade do processo.

Hair *et al.* (2005) apresentam seis estágios para o processo de construção de um modelo para o uso da análise de *cluster*. Define-se a seguir esses estágios.

Estágio 1 - esse estágio é de definição dos objetivos da análise de *cluster*, tal como a descrição taxonômica, ou seja, a formação de uma classificação de objetos baseada na experiência; simplificação de dados, pois a análise de *cluster* executa a mesma tarefa para as

observações; a identificação de relação que é um meio para revelar as relações entre as observações, o que talvez não fosse possível com as observações individuais.

Nesse estágio destacam ainda a seleção de variáveis a serem incluídas na variável estatística de agrupamento, a qual deve ser feita em relação às considerações teóricas e conceituais, bem como práticas. Se o argumento a respeito da seleção das variáveis for baseado em uma teoria explícita, pesquisa anterior, ou suposição deve-se ter em mente a importância de incluir apenas variáveis que caracterizam objetos agregados ou se relacionem especificamente aos objetivos da análise de *cluster*. Esta técnica não tem meios de diferenciar variáveis relevantes das irrelevantes.

Estágio 2 - esse estágio trata do projeto de pesquisa em análise de *cluster*. Antes de iniciar o processo de partição, três questões devem ser respondidas. Que são:

i) as observações atípicas (independentes dos *clusters* existentes) podem ser detectadas e se for caso podem ser eliminadas?

ii) como deve ser medida a similaridade?

iii) os dados devem ser padronizados?

Nesse estágio aborda-se como se faz a detecção de observações atípicas. A análise de *cluster* é muito sensível à inclusão de variáveis irrelevantes e também igualmente sensíveis a observações atípicas, que se entende mais precisamente por objetos muito diferentes de todos os outros. Assim uma projeção preliminar de observações atípicas é sempre necessária.

Quanto às medidas de similaridade entre objetos ou medida de correspondência ou semelhança entre objetos a serem agrupados, sugerem primeiro especificar as características que definem a similaridade e após definidas combinar as características em uma medida de similaridade calculada para todos os pares de objetos. Então, o procedimento prossegue agregando os objetos semelhantes em *clusters*.

Destacam três métodos para a medida de similaridade nas aplicações da análise de *cluster*:

i) medida correlacional - essa medida requer dados métricos e representam a similaridade pela correspondência de padrões ao longo das características (X variáveis), esta medida não olha a magnitude, mas sim os padrões dos valores.

ii) medida de distância - essa medida requer dados métricos e representam a similaridade como a proximidade entre observações ao longo das variáveis na variável

estatística do *cluster*. Mais precisamente as medidas de distância são, na verdade, uma medida de dissimilaridade, no qual os valores maiores denotam menor similaridade. A distância é convertida em uma medida de similaridade pelo uso de uma relação inversa.

iii) medida de associação – é usada para comparar objetos cujas características são medidas apenas em termos não métricos (medida nominal e ordinal).

Após a seleção da medida de similaridade os autores remetem-se a padronização das variáveis explicitando que a forma mais comum de padronização é a conversão de cada variável em escores padrão (z-escore) que é dado pelo escore subtraído da média após divisão pelo desvio padrão para cada variável.

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (15)$$

onde:

X_i = uma dada variável

\bar{X} = média do conjunto das observações

S = desvio padrão amostral

De acordo com Hair *et al.* (2005) um dos benefícios da padronização das variáveis é a facilidade para comparar as variáveis quando elas estão na mesma escala; a magnitude representa o número de desvios-padrão que o valor original está distante da média. Outro benefício é que não há diferença nos valores padronizados quando muda apenas a escala.

Por outro lado, se existe alguma relação “natural” refletida na escala das variáveis, a padronização pode não ser apropriada.

Estágio 3 - suposições em análise de *cluster*, essa técnica não é uma técnica de inferência estatística em que os parâmetros de uma amostra são avaliados como possíveis representantes de uma população. Assim, esse estágio deve ser concentrado em duas questões críticas, representatividade da amostra e a multicolinearidade.

Quanto à representatividade da amostra sugerem que todos os esforços devam ser feitos para garantir que a amostra é representativa e que os resultados são generalizáveis para a população de interesse. Quanto ao impacto da multicolinearidade argumentam que o efeito

na análise de *cluster* é diferente porque as variáveis multicolineares são implicitamente ponderadas com maior peso. A multicolinearidade atua como um processo de ponderação não visível para o observador, mas que afeta a análise. Então se devem examinar as variáveis usadas em busca de multicolinearidade substancial e se encontrá-las reduzir a números iguais em cada *cluster*.

Estágio 4 - determinação de *clusters* e avaliação do ajuste geral, nesse estágio a pergunta importante a responder é qual procedimento deve ser usado para colocar objetos similares em *clusters*? Ou qual algoritmo é o mais adequado?

Os algoritmos mais comumente usados classificam-se em duas categorias gerais, hierárquica e não hierárquicos. Os procedimentos hierárquicos envolvem a construção de uma estrutura do tipo árvore. Podem ser de dois tipos aglomerativos e divisivos.

Nos aglomerativos cada objeto de observação começa como seu próprio *cluster*. Nos próximos passos os dois *clusters* mais próximos são combinados em um novo *cluster*, reduzindo assim o número de *clusters* em uma unidade em cada passo. Uma característica importante desse procedimento hierárquico é que os resultados de um estágio anterior são sempre aninhados com os resultados de um estágio posterior. A representação gráfica desse processo chama-se dendograma. Cita-se a seguir 5 algoritmos aglomerativos:

- i) ligação individual ou vizinho mais próximo - esse procedimento é baseado na distância mínima. Ele encontra os dois objetos separados pela menor distância e os coloca num primeiro *cluster*. Em seguida, a próxima distância mais curta é determinada, e um terceiro objeto se junta aos dois primeiros para formar um *cluster* ou um *cluster* de dois membros é formado. O processo continua até que todos os objetos formem um só cluster;
- ii) ligação completa - é semelhante ao da ligação individual, exceto em que o critério de agrupamento se baseia na distância máxima. A distância máxima entre indivíduos em cada *cluster* representa a menor esfera que pode incluir todos os objetos em ambos os *clusters*. É chamado de ligação completa porque todos os objetos em um *cluster* são conectados um com o outro a alguma distância máxima ou similaridade mínima;
- iii) ligação média - começa da mesma forma que a ligação individual ou completa, mas o critério de agrupamento é a distância média de todos os indivíduos em um *cluster* aos demais em outro. Esta abordagem tende a combinar *clusters* com pequena variação interna e tendem a produzir *clusters* com aproximadamente a mesma variância;

iv) método *ward* - a distância entre dois *clusters* é a soma dos quadrados entre os dois *clusters* feitos sobre todas as variáveis. Em cada etapa do procedimento, a soma interna de quadrados é minimizada sobre todas as partições que podem ser obtidas pela combinação de dois *clusters* do estágio anterior. Esse procedimento tende a combinar *clusters* com um pequeno número de observações;

v) método centróide - a distância entre dois *clusters* é a distância entre seus centróides. Centróides são os valores médios das observações sobre as variáveis na variável estatística de agrupamento. Toda vez que os indivíduos são reunidos um novo centróide é computado. A vantagem desse método é que ele é menos afetado por observações atípicas do que outros métodos hierárquicos.

No método divisivo, começa-se com um grande *cluster* que contém todas as observações (objetos). Em passos sucessivos, as observações que mais diferem entre si são separadas e transformadas em *clusters* menores. Esse processo continua até que cada objeto seja um agrupamento por si mesmo.

Procedimentos não hierárquicos ou *k*-médias: não envolvem o processo de construção de árvore, em vez disso designam objetos aos *clusters* assim que o número de *cluster* estiver sido especificado. O processo funciona seguindo as etapas, sendo a primeira a seleção de uma semente de agrupamento como o centro inicial de um *cluster* e todos os objetos dentro de uma distância de referência pré-especificada é incluído no *cluster* resultante. Em seguida outra semente de agrupamento é escolhida e a designação continua até que todos os objetos tenham sido designados. Os objetos podem então serem redesignados se estiverem mais próximos de outro *cluster* do que daquele ao qual foram originalmente associados.

Estágio 5 - interpretação dos *clusters*, este estágio envolve o exame de cada *cluster* em termos da variável estatística de agrupamento para nomear ou designar um rótulo que descreva precisamente a natureza do *cluster*.

Uma medida frequentemente usada quando se inicia o processo de interpretação é o centróide. Se o procedimento de agrupamento for realizado sobre os dados iniciais, a interpretação será uma descrição lógica. Se os dados forem padronizados daí deve-se voltar aos escores iniciais, isto é, para as variáveis originais e computar os perfis médios usando esses dados.

O perfil e a interpretação dos *clusters* fornecem um meio de avaliar a correspondência dos *clusters* obtidos com aqueles propostos por alguma teoria ou por experiência prática. Se for usada pra confirmação, os perfis da análise de *cluster* fornecem um meio direto para

avaliar a correspondência, pois fornecem uma rota para se fazer avaliações de significância prática.

Estágio 6 - validação e perfil do *cluster*, apesar de não haver qualquer método pra garantir a validade e significância prática, destacam-se duas abordagens para fornecer alguma base para avaliação.

Validação da solução, a validação inclui tentativas para garantir que a solução seja representativa da população geral, e assim generalizável para outros objetos e estável com o passar do tempo. Um método comum é particionar a amostra em dois grupos. Cada um é analisado separadamente e os resultados são então comparados.

Outra técnica inclui uma forma modificada de partição de amostra onde centros de grupos obtidos a partir de uma solução são empregados para definir *clusters* de outras observações e os resultados são comparados, é uma forma direta de validação cruzada.

Perfil da solução por *cluster*, o estágio de perfil envolve a descrição das características de cada *cluster* para explicar como eles podem diferir em dimensões relevantes. O procedimento inicia depois que os *clusters* são identificados no procedimento de agrupamento que caracterizam cada *cluster*.

Em relação aos procedimentos desse estudo, detalham-se alguns pontos fundamentais do método análise de *cluster*, isto é, pontuam-se a seguir alguns pressupostos básicos em relação à definição do método, à função distância, padronização das variáveis, homogeneidade dos *clusters* e estabilidade, todos voltados para o escopo desse estudo usando um tratamento vetorial multidimensional, que o mesmo requer.

3.7 Definição da análise de *cluster* via tratamento vetorial

A visão apresentada a seguir serve para elucidar a análise de *cluster* no presente estudo. A partir da definição da análise de *cluster*, a qual consiste em agrupar objetos similares, usa-se a noção vetorial para a representação de tais objetos.

Formalmente, denota-se um dado objeto pelo vetor x , usando a seguinte notação:

$$x = (x_1, x_2, \dots, x_n) \tag{16}$$

onde:

x_i = componentes do vetor com $i = 1, \dots, n$

Em particular, considerando-se o retorno mensal de uma dada empresa num período com n meses, têm-se a seguinte notação para o vetor retorno \mathbf{x}^k de uma k empresa:

$$\mathbf{x}^k = (x_1, x_2, \dots, x_n) \quad \text{com} \quad k = 1, \dots, p \quad (17)$$

onde:

x_i = componentes dos retornos mensais de uma dada empresa

Da mesma forma, outros vetores representarão o retorno de outras empresas, sendo tais vetores os objetos almejados para os *clusters*. E para as demais variáveis utilizadas no estudo apropria-se da mesma idéia e notação.

Formalmente, denota-se um dado *cluster* X , com m objetos, sendo cada objeto de dimensão n , por:

$$X = \{x^1, x^2, x^3, \dots, x^m\} \text{ tal que, } x^i = (x_1^i, x_2^i, \dots, x_n^i) \text{ e } i = 1, \dots, m \quad (18)$$

onde:

x^i = vetor

x_j^i = componente do vetor com $i = 1, \dots, m$ e $j = 1, \dots, n$

Em particular, considerando-se a variável retorno mensal num período com n meses, para uma k empresa, com $k = 1, \dots, p$, então um dado *cluster* X de dimensão m será denotado como segue:

$$X = \{x_1^1, x_2^1, \dots, x_n^1\} \text{ tal que, } x_j^i = (x_1^i, x_2^i, \dots, x_n^i) \text{ e } i = 1, \dots, p \text{ e } j = 1, \dots, m \quad (19)$$

onde:

x_j^i = vetor retorno de uma dada empresa com $i = 1, \dots, p$ e $j = 1, \dots, m$

3.7.1 Noções de distância

De acordo com Lima (1981), define-se distância como uma função $d: R^n \times R^n \rightarrow R$, tal que satisfaz as seguintes condições:

i) $d(x, x) = 0$

ii) $d(x, y) \geq 0$

iii) $d(x, y) \leq d(x, z) + d(z, y)$

onde:

x , y e z são vetores arbitrários de dimensão n , da forma $x = (x_1, \dots, x_n)$; $y = (y_1, \dots, y_n)$ e $z = (z_1, \dots, z_n)$.

Em particular, seguindo a mesma notação anterior, mostram-se a definição de distância para dois vetores representantes do retorno mensal de duas dadas empresas, num período com n meses. Então:

$$d: R^n \times R^n \rightarrow R \quad (20)$$

$$(x^1, x^2) \rightarrow a(\text{const})$$

Tomando-se dois vetores x^1 e x^2 , algumas medidas de distância, de acordo com Boillat *et al.* (2002) e Clifford (2002) são expressas como:

- **distância euclidiana**

$$d(x^1, x^2) = \left[\sum_{i=1}^n (x_i^1 - x_i^2)^2 \right]^{\frac{1}{2}} \quad (21)$$

Esta distância é a recomendada para os métodos de agrupamento centróide e *ward*.

- **distância city block**

$$d(x^1, x^2) = \sum_{i=1}^n |x_i^1 - x_i^2| \quad (22)$$

Esta distância para serem usadas as variáveis em questão não pode ser correlacionada umas com as outras.

- **distância Mahalanobis**

$$d(x^1, x^2) = (x^1 - x^2)' V^{-1} (x^1 - x^2) \quad (23)$$

onde:

V = matriz de covariância de x^1 e x^2 ; $(x^1 - x^2)'$ indica a matriz transposta.

- **distância de Pearson**

$$d_p(x^1, x^2) = |\delta_{12} - 1| \quad (24)$$

onde:

d_p = distância de *Pearson* entre os entes

$\delta_{1,2}$ = correlação de *Pearson* entre x^1 e x^2

A seguir mostra-se a medida de distância euclidiana e de *Pearson* para um par de objetos (vetores) através da notação desenvolvida anteriormente. Para tanto, toma-se a variável retorno mensal num período com n meses de duas empresas x^1 e x^2 .

$$d(x^1, x^2) = \left[\sum_{i=1}^n (x_i^1 - x_i^2)^2 \right]^{\frac{1}{2}} \quad \text{e,} \quad d(x^1, x^2) = |\delta_{12} - 1| \quad (25)$$

Clifford (2002) usa a distância de *Pearson* em seu estudo. Na presente pesquisa o coeficiente escolhido para ser utilizado na avaliação da proximidade entre os setores foi à distância euclidiana um coeficiente de dissimilaridade usado na maioria dos estudos pesquisados (ARNOTT, 1980; BOILLAT *et al.*, 2002; SILVA *et al.*, 2007).

3.7.2 Algoritmo de ligação

O algoritmo de ligação trata-se de um procedimento de agregação dos objetos dentro de um *cluster*, ou seja, procedimento de colocação de objetos similares em um *cluster*. A seguir sob a ótica da análise vetorial, citam-se tais algoritmos:

Seja $D(X, Y)$ uma medida de dissimilaridade entre dois *clusters* (X e Y).

Assim, de acordo com o método de ligação individual o qual é baseado na distância mínima, tem-se:

$$D(X, Y) = \min \{d(x^i, y^j) \mid i \in (1, \dots, m); j \in (1, \dots, k)\} \quad (26)$$

No método de ligação completa, que se baseia na distância máxima, tem-se:

$$D(X, Y) = \max \{d(x^i, y^j) \mid i \in (1, \dots, m); j \in (1, \dots, k)\} \quad (27)$$

No método de ligação média, o critério de agrupamento é a distância média de todos os indivíduos em um *cluster* aos demais em um outro. Assim:

$$D(X, Y) = \frac{1}{m \cdot k} \sum_{i,j} d(x^i, y^j) \text{ com } i \in (1, \dots, m) \text{ e } j \in (1, \dots, k) \quad (28)$$

No método centróide, a distância entre dois *clusters* é a distância entre seus centróides, então:

$$D(X, Y) = d(x_{\text{médio}}, y_{\text{médio}}) \quad (29)$$

onde:

$$x_{\text{médio}} = \frac{1}{m} \sum_i d(0, x^i)$$

$$y_{\text{médio}} = \frac{1}{k} \sum_j d(y^j, 0)$$

No método de *ward* a distância entre dois *clusters* é a soma dos quadrados entre os dois *clusters* feitos sobre todas as variáveis, então:

$$D(X, Y) = \frac{m.k}{m+k} (d(x_{\text{médio}}, y_{\text{médio}}))^2 \quad (30)$$

Apesar de o método individual ser o mais comumente usado, ele é aplicado somente para objetos que são relativamente distintos uns dos outros (BOILLAT *et al.*, 2002). No estudo em questão não é adequado, pois os dados das empresas que formam a base do estudo podem apresentar pequena variação relativa, assim o método não é ideal. Por outro lado, Boillat *et al.*, (2002) apresentam duas propriedades do método *ward* as quais são fundamentais para a proposta do estudo em questão.

i) seja a medida de dissimilaridade dada pelo método *ward*

$$D(X, Y) = \frac{m.k}{m+k} (d(x_{\text{médio}}, y_{\text{médio}}))^2 \text{ donde o fator } \frac{m.k}{m+k}, \text{ que é parte da fórmula,}$$

cria uma tendência que pressupõe a formação de *clusters* internamente homogêneos contendo números similares de objetos. Por exemplo, suponha-se que a distância entre os centros de massas dos *clusters* *X* e *Y* seja 1, se cada um tiver 5 objetos, a distância *ward* será ponderada por 2,5. Se por outro lado se *X* contém 1 objeto e *Y* contém 9 objetos, o fator de ponderação será de 0,9. Logo com a ponderação proporcionada por tal fator previnem a formação de um número pequeno de *clusters* grandes, com alguns objetos independentes e isolados flutuantes entre eles;

ii) se os dados que representam os objetos a ser classificados forem normalizados então a distância entre os *clusters* sob o método *ward* é diretamente proporcional ao coeficiente de correlação entre tais objetos.

Logo esses resultados mostram a existência de uma relação direta entre o método usado para formar os *clusters* e o tradicional método utilizado no campo das finanças para identificar relacionamentos entre variáveis (multicolinearidade), no que se refere à correlação linear entre elas. Assim, nesta pesquisa utiliza-se o método de ligação *ward*.

Demonstra-se a seguir a proposição (ii), conforme apresentado por Boillat *et al.* (2002). De fato, seja a normalização efetuada como segue:

$$x' = \frac{x_i - \langle x \rangle}{(\langle x^2 \rangle - \langle x \rangle^2)^{0,5}} \quad (31)$$

onde:

x_i = cada componente do vetor

$\langle x \rangle$ = média das componentes.

A partir de (29) e da medida da dissimilaridade do método *ward* (28), com as devidas substituições e após alguns passos algébricos, chega-se a proporcionalidade da distância *ward* ao coeficiente de correlação.

$$d_{ward} \cong \text{constante } t - 2 \cdot \left[\frac{\sum_i x_i y_i - \langle x \rangle \cdot \langle y \rangle}{((\langle x^2 \rangle - \langle x \rangle^2) \cdot (\langle y^2 \rangle - \langle y \rangle^2))^{0,5}} \right] \quad (32)$$

3.7.3 Dendograma

O dendograma é a forma gráfica usada para representar o resultado final dos diversos agrupamentos. Entretanto, por ser este gráfico uma simplificação em duas dimensões de uma relação n -dimensional, é inevitável que algumas distorções quanto à similaridade apareçam.

Na construção de um dendograma, sob o viés vetorial, os objetos que serão aglomerados ficam no eixo vertical e as distâncias que separam os *clusters* ficam no eixo horizontal. Conseqüentemente torna-se fácil ver a posição do objeto em relação aos seus vizinhos. Cada linha vertical representa um *cluster*, a composição inicial do *cluster* é formada por dois objetos e gradualmente o tamanho da composição vai aumentando. O dendograma no presente estudo permitiu obter-se uma imagem singular da totalidade do processo de aglomeração resultante do método, também foram utilizados, principalmente, para se observar os saltos que ocorrem na formação dos grupos. Assim, através do dendograma busca-se detectar a formação de grupos heterogêneos, ou seja, essas estruturas permitiram também a identificação de quanto seria necessário consentir ou “relaxar” na definição de grupos homogêneos, corroborando para a definição do número de *clusters* quando da utilização da regra de parada.

3.7.4 Critério de seleção do número de *clusters*

Quanto ao número de *clusters* que devem ser formados segundo vários autores (CORRAR, 2007; PESTANA E GAGEIRO, 2005; HAIR et al., 2005; JOHNSON e WICHERN, 2002) não existe um procedimento objetivo padrão de seleção. Alguns autores entre eles Johnson e Wichern (2002) sugerem um tipo de regra de parada (*stopping rule*) relativamente simples que é o exame da medida de distância entre *clusters* em cada passo sucessivo, sendo a solução quanto ao número de agrupamento definido quando a medida das diferenças de distâncias entre duas etapas adjacentes exceder a um valor especificado ou quando existir um salto repentino no tamanho das diferenças pode-se considerar como o número adequado de *clusters*. Esta solução baseia-se na lógica de que a última combinação causou um substancial decréscimo na similaridade. Além disso, deve-se também considerar a análise do dendograma como suporte complementar ao julgamento estritamente empírico realizado. Salienta-se que este procedimento refere-se aos processos aglomerativos no caso do método hierárquico.

3.7.5 Homogeneidade dos *clusters*

O método, análise de *cluster*, mostra claramente *clusters* homogêneos do ponto de vista matemático, o que é obviamente claro, dado a estrutura do método. Entretanto, espera-se que estes *clusters* sejam homogêneos também sob o ponto de vista da análise setorial, pois a homogeneidade está entre os pressupostos da formação dos grupos. Nesse sentido, o princípio básico adotado neste estudo é o de que um *cluster* será tão mais homogêneo quanto uma maior proporção de empresas formadoras dos *clusters* pertencerem ao mesmo setor.

Mais precisamente, a partir dos resultados da análise de *cluster*, cada um dos *clusters* terá sua composição avaliada. Cada empresa pertencente a determinado *cluster* será classificada em um dos setores, conforme as classificações adotadas nesse estudo. Dada à classificação das empresas dentro de um *cluster*, faz-se a distribuição de frequências dos setores para o referido *cluster*. Assim, quanto maior a frequência em um único setor mais homogêneo será o *cluster*.

3.7.6 Estabilidade dos *clusters*

É razoável argumentar que o período sobre o qual os dados são observados tem substancial influência sobre os resultados e, portanto sobre a formação dos *clusters*. A questão natural é saber se estas estruturas são estáveis suficientemente para serem usadas. Assim, para a avaliação da estabilidade dos *clusters* ao longo do período do estudo utilizou-se o Teste de Wilcoxon, o qual dá mais peso a um par que mostra uma diferença grande entre as duas condições do que a um par que mostra uma diferença pequena e aplica-se para se analisar as diferenças entre duas condições (SIEGEL e CASTELLAN, 2006). Assim, na mecânica do teste, leva-se em conta valor absoluto das diferenças, sendo a diferença expressa por $|D_i| = |Y_i - X_i|$, onde os dados que ocorrem aos pares são analisados numa sucessão única de valores ordenados. As diferenças nulas, se existirem são suprimidas e no caso de diferenças iguais em valor absoluto (empates) atribui-se a cada uma delas a média das ordens que lhes caberiam se os valores fossem distintos. As hipóteses são:

$$H_0 = E(X) = E(Y) \quad (33)$$

$$H_1 = E(X) \neq E(Y) \text{ ou } E(X) > E(Y) \text{ ou } E(X) < E(Y)$$

A estatística quando existe empate é dado por:

$$T = \frac{W^+ - W^-}{\sqrt{\sum_{i=1}^n R_i^2}} \quad (34)$$

onde:

W^+ = soma das ordens que correspondem a diferenças

W^- = soma das ordens que correspondem a diferenças negativas

R_i = ordem atribuída às diferenças

No entanto para amostras grandes ($n \geq 20$) como é o caso do presente estudo, a distribuição T tende para a normal, e segundo Reis (1997) a aproximação à normal é sempre válida caso haja empates ou o número de pares em análise seja grande.

A estatística do teste quando não existem empates é dada por:

$$T = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{n(n+1)(2n+1)/24}} \quad (35)$$

onde:

W^+ = soma das ordens que correspondem a diferenças

n = número de elementos da amostra

Como nesse estudo a análise de *cluster* será aplicada para cada ano do período proposto se utiliza o Teste de Wilcoxon para comparar as composições medianas dos *clusters* geradas para cada ano. Portanto, se a análise de *cluster* é estável ao longo dos anos, a composição setorial mediana de cada *cluster*, deverá ser mantida a mesma, ou seja, a participação dos setores dentro do *cluster* não deverá se alterar significativamente. Portanto, constata-se a estabilidade do *cluster* se a hipótese nula, igualdade da composição mediana dos *clusters*, for aceita.

4. ANÁLISE DOS RESULTADOS

Este capítulo aborda quatro momentos distintos: o primeiro momento traz a definição da amostra, procedimentos adotados e tratamento dos dados. O segundo momento apresenta uma análise inicial dos dados e engloba a apresentação das estatísticas descritivas das variáveis médias, incluindo uma descrição da amostra e uma análise univariada. O terceiro momento é dedicado à aplicação da análise de *cluster*. E por fim, o quarto momento apresenta os testes para avaliação da eficiência da análise de *cluster*.

4.1 Procedimento de amostragem

Utilizando a base de dados do Economática realizou-se a coleta de dados referentes às variáveis do estudo, para cada um dos sete anos, isto é, de 2000 a 2007, e de todas as ações de empresas brasileiras listadas no programa. Obteve-se uma listagem inicial de 403 ações e, como dentre essas, 5 ações são de empresas que foram registradas em 2008, optou-se por retirá-las passando então a amostra a contar com um total de 398 ações. Ressalta-se ainda que para cada um dos anos do período de estudo a quantidade de ações reporta-se as ações ativas na BOVESPA, com o devido registro válido durante todo o ano. No Quadro 04 apresenta-se a quantidade de ações para cada ano do período do estudado.

ANO	NÚMERO DE AÇÕES
2000	287
2001	290
2002	294
2003	296
2004	299
2005	306
2006	335
2007	391
2000 a 2007	398

Quadro 04 – Quantidade de ações em negociação nos respectivos anos, no período de 2000 - 2007

Observa-se que ao longo dos anos estudados houve um aumento no número de ações ativas, indicando que o número de ações sendo negociadas foi superior ao número de ações que deixaram de ser negociadas. A amostra durante o período de estudo apresenta 398 ações diferentes com um crescimento percentual no período de 36,24.

4.2 Tratamento dos dados

Os dados da pesquisa receberam dois tratamentos distintos. Primeiramente apresenta-se o tratamento para os dados omissos (item 4.2.1). Na seqüência apresentam-se os procedimentos adotados para a verificação dos *outliers* (item 4.2.2).

4.2.1 Dados omissos

A Tabela 01 apresenta a quantidade média de dados omissos para cada variável em relação aos anos de 2000 a 2007. A média foi obtida a partir do número de dados omissos em cada ano conforme apresentado no Apêndice A.

Tabela 01 – Quantidade de percentual médio (M), mediana (MD), desvio padrão (DP) de dados omissos para cada variável em relação ao período de 2000 - 2007

VARIÁVEIS	MÉDIA	%	MEDIANA	DESVIO PADRÃO
liquidez corrente	17,75	5,68	16,50	5,99
liquidez seca	38,38	12,28	36,00	9,65
margem operacional	49,88	15,97	46,00	10,29
margem ebitida	69,75	22,33	66,50	16,68
rentabilidade do ativo	13,63	4,36	13,50	4,34
rentabilidade do patrimônio líquido	50,13	16,05	52,00	9,26
exigível total/ativo total	11,63	3,72	11,00	4,96
endividamento geral	11,63	3,72	11,00	4,96
endividamento financeiro	62,38	19,97	56,50	14,21
dividendo <i>yeld</i>	175,75	56,26	174,50	10,01
<i>payout</i>	139,88	44,78	137,00	15,14
valor patrimonial da ação	24,75	7,92	10,50	41,75
liquidez em bolsa	60,75	19,45	61,00	6,58
retorno	96,25	30,81	101,00	13,35
ln (valor de mercado)	66,25	21,21	76,50	31,95
ln (receita líquida)	46,50	14,89	46,00	3,16
ln (ativo total)	21,75	6,96	11,50	28,34

Da Tabela 01 pode-se observar que em média as variáveis “dividendo *yeld*” e “*payout*” apresentam o maior percentual de dados omissos no período considerado com 56,26% e 44,78 %, respectivamente. Salienta-se que os percentuais correspondentes as variáveis apresentados na tabela 01 relacionam-se a um total médio de 312,37 de números de casos. Pela mediana constata-se certa proximidade com a média, exceto para as variáveis “valor patrimonial da ação”, “ln (valor de mercado)” e “ln (ativo total)” que apresentam uma diferença entre elas maior do que 20 %, como acusado pelo desvio padrão.

4.2.2 Dados *outliers*

A identificação de *outliers* é feita com base nos *clusters* índice e mercado, em cada um dos respectivos anos. Ressalta-se que o *cluster* índice envolve todas as variáveis menos as de mercado, ao passo que o *cluster* mercado envolve somente as variáveis de mercado. Os resultados deste procedimento são apresentados na Tabela 02.

Tabela 02 – Quantidade de *outliers* para os *clusters* índices e mercados em relação ao período de 2000 – 2007

ANOS	QUANTIDADE DE <i>OUTLIERS</i> COM RESPECTIVOS PERCENTUAIS											
	SEM VARIÁVEIS DE TAMANHO						COM VARIÁVEIS DE TAMANHO					
	<i>CLUSTER</i> ÍNDICE			<i>CLUSTER</i> MERCADO			<i>CLUSTER</i> ÍNDICE			<i>CLUSTER</i> MERCADO		
	CASOS VÁLIDOS	Nº	%	CASOS VÁLIDOS	Nº	%	CASOS VÁL.	Nº	%	CASOS VÁL.	Nº	%
2000	231,0	5,0	2,2	224,0	12,0	5,4	232,0	9,0	3,9	230,0	4,0	1,7
2001	234,0	9,0	3,8	230,0	10,0	4,5	235,0	13,0	5,5	235,0	5,0	2,1
2002	235,0	6,0	2,6	234,0	6,0	2,7	236,0	11,0	4,7	230,0	6,0	2,6
2003	239,0	5,0	2,1	236,0	4,0	1,8	241,0	6,0	2,5	242,0	5,0	2,1
2004	247,0	13,0	5,3	250,0	11,0	4,9	249,0	13,0	5,2	254,0	8,0	3,1
2005	254,0	19,0	7,5	255,0	8,0	3,6	256,0	20,0	7,8	258,0	2,0	0,8
2006	276,0	16,0	5,8	275,0	15,0	6,7	279,0	13,0	4,7	282,0	5,0	1,8
2007	313,0	17,0	5,4	307,0	11,0	4,9	324,0	18,0	5,6	334,0	2,0	0,6
MÉDIA	253,6	11,2	4,3	251,3	9,6	4,3	256,5	12,8	5,0	258,1	4,6	1,8

Diante do exposto na Tabela 02, pode-se observar que a quantidade média percentual de *outliers* para os *clusters* índice e mercado mostra-se em torno de 4,3% quando não se considera as variáveis de tamanho e com a inclusão de tais variáveis observa-se que fica em torno de 3,4%. Então, seguindo a literatura (CORRAR *et al.*, 2007; JOHNSON e WICHERN, 2002) optou-se por não retirá-los da amostra.

4.3 Classificações setoriais

Apresenta-se a seguir a distribuição das empresas conforme as classificações setoriais propostas pela BOVESPA e pela *North America Industrial Classification* (NAICS nível 1).

A Tabela 03 apresenta as empresas distribuídas conforme a classificação setorial da BOVESPA, esta tabela apresenta o respectivo percentual de distribuição das empresas nos setores para cada um dos anos.

Tabela 03 - Classificação setorial segundo a BOVESPA para cada ano das ações em negociação no período de 2000 – 2007

CLASSIFICAÇÃO BOVESPA										
SETORES	2000 (%)	2001 (%)	2002 (%)	2003 (%)	2004 (%)	2005 (%)	2006 (%)	2007 (%)	M (%)	DP
bens industriais	11,1	11,4	11,2	11,1	11,0	10,8	11,0	9,7	10,9	0,5
construção e transporte	10,5	10,7	10,9	10,8	11,0	11,1	11,3	14,6	11,4	1,2
Consumo cíclico	18,5	18,6	18,4	18,2	18,1	18,3	16,7	15,9	17,8	0,9
Consumo não cíclico	7,3	7,2	7,5	7,4	8,4	8,5	9,6	10,0	8,2	1,0
financeiro e outros	16,4	16,2	16,0	15,9	16,1	16,0	16,1	18,2	16,4	0,7
Materiais básicos	16,4	16,2	16,3	16,2	15,4	14,7	13,7	12,8	15,2	1,3
petróleo, gás e biocombustíveis	0,7	0,7	0,7	0,7	0,7	0,7	1,2	1,0	0,8	0,2
tecnologia da informação	0,7	0,7	0,7	0,7	0,7	1,0	1,8	1,8	1,0	0,5
Telecomunicações	6,3	6,2	6,1	6,1	6,0	5,9	5,1	4,3	5,8	0,6
Utilidade pública	12,2	12,1	12,2	12,8	12,7	13,1	13,4	11,8	12,5	0,5
TOTAL (%)	100									
TOTAL DE EMPRESAS	287	290	294	296	299	306	335	391		

M = média, DP= desvio padrão

Da classificação anual segundo a BOVESPA observa-se um crescimento do número de empresas no decorrer do período nos setores de “construção e transporte”, “consumo não cíclico”, “petróleo gás e bicombustíveis” e “tecnologia da informação”. Nos demais, se verifica uma diminuição do número de empresas. Observa-se também que, em média, o setor que apresenta uma maior concentração de empresas no período é o setor de “consumo cíclico” (17,8 %) e o menor, o setor de “petróleo gás e bicombustíveis” (0,8%). E para uma visão geral da classificação BOVESPA em todo o período apresenta-se a tabela 04.

Tabela 04 - Classificação segundo a BOVESPA considerando todas as ações em negociação no período de 2000 – 2007

CLASSIFICAÇÃO BOVESPA		
SETORES	FREQ	%
bens industriais	38,0	9,5
construção e transporte	57,0	14,3
Consumo cíclico	62,0	15,6
Consumo não cíclico	39,0	9,8
financeiro e outros	71,0	17,8
Materiais básicos	54,0	13,6
petróleo, gás e biocombustíveis	4,0	1,0
tecnologia da informação	7,0	1,8
telecomunicações	19,0	4,8
Utilidade pública	47,0	11,8
TOTAL	398	100
MÉDIA	39,8	10

A partir da visão geral apresentada na Tabela 04 observa-se que, para as ações da amostra, o setor “financeiros e outros” com 17,8% do total é o mais representativo e o setor “petróleo gás e biocombustíveis” é o de menor representatividade na amostra.

A NAICS trata-se de uma classificação setorial internacional estruturada em três níveis. Num primeiro nível as empresa dividem-se em 20 categorias, sendo que cada uma destas categorias subdividem-se num total de 96 categorias para formar o nível 2, e no nível 3, apresenta um total de 313 categorias. Na presente pesquisa optou-se pela classificação no nível 1 devido ao número de empresas pertinentes a amostra. Assim, o estudo estrutura-se nas classificações setoriais de acordo com a classificação da BOVESPA e da NAICS no nível 1, que a partir de agora se denota por NAICS.

Tabela 05 - Classificação setorial segundo a NAICS para cada ano das ações em negociação, no período de 2000 – 2007

CLASSIFICAÇÃO NAICS										
SETORES	2000	2001	2002	2003	2004	2005	2006	2007	M	DP
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)		
Adm. empresas e empreend.	7,7	7,6	7,8	7,8	7,7	7,5	7,2	6,1	7,4	0,5
agric., pec., silvi., pesca e caça	0,3	0,3	0,3	0,3	0,7	0,7	0,9	1,0	0,6	0,3
artes, entretenimento e recreação	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,0
assistência médica e social	0,0	0,0	0,0	0,0	0,3	0,3	1,2	1,3	0,4	0,5
Comércio atacadista	0,0	0,0	0,0	0,0	0,3	0,3	0,6	0,5	0,2	0,2

CLASSIFICAÇÃO NAICS										
SETORES	2000	2001	2002	2003	2004	2005	2006	2007	M	DP
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)		
Comércio varejista	3,8	3,8	3,7	3,7	3,7	3,6	3,3	3,3	3,6	0,2
construção	4,2	4,5	4,4	4,4	4,3	4,2	5,7	9,2	5,1	1,6
educação	1,0	1,4	1,4	1,4	1,3	1,3	1,2	1,3	1,3	0,1
empresa de eletríc., gás e água	12,2	12,1	12,2	12,5	12,4	12,7	13,4	12	12,4	0,4
hotel e restaurante	0,7	0,7	0,7	0,7	0,7	0,7	0,6	0,5	0,7	0,1
imob. e locadora de outros bens	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,8	0,4	0,2
Indústria manufatureira	46,3	45,9	45,6	45,3	44,1	43,5	41,5	38,6	43,9	2,5
informação	7,3	7,2	7,1	7,1	7,0	7,2	6,9	5,9	7,0	0,4
mineração	1,4	1,4	1,4	1,4	1,3	1,3	2,1	1,8	1,5	0,3
outros serv. (exceto adm. púb.)	0,7	0,7	0,7	0,7	0,7	0,7	0,6	0,5	0,7	0,1
Serviços de apoio e ger. res. e rem.	0,3	0,7	0,7	0,7	0,7	0,7	0,9	1,3	0,8	0,3
Serviços financeiros e seguros	9,4	9,3	9,2	9,1	9,4	9,5	8,7	10,5	9,4	0,5
Serviços profiss., cient. e técnicos	0,7	0,7	1,0	1,0	1,0	1,0	0,9	0,8	0,9	0,1
transporte e armazenamento	3,1	3,1	3,1	3,4	3,7	4,2	3,9	4,3	3,6	0,5
TOTAL (%)	100									
TOTAL DE EMPRESAS	287	290	294	296	299	306	335	391		

Da classificação anual segundo a NAICS observa-se que em nove dos seus vinte segmentos apresentam crescimento em relação ao número de empresas. Setores como de “agricultura-pecuária-silvicultura-pesca-caça”, “assistência médico social”, “construção”, “imobiliária-locadora de outros bens”, “mineração”, “serviços de apoio as empresas e gerenciamento de resíduos e remediação”, “serviços profissionais científicos e técnicos” e por fim o setor de “transporte e armazenamento” apresenta tal crescimento o que não se verifica nos demais. Observa-se também que em média o setor que apresenta uma maior concentração de empresas é o setor de indústria manufatureira (43,9%) e em média o de menor concentração é o setor de comércio atacadista (0,2%).

E para uma visão geral da classificação NAICS em todo o período apresenta-se a Tabela 06.

Tabela 06 - Classificação segundo a NAICS considerando todas as ações em negociação no período de 2000 - 2007

CLASSIFICAÇÃO NAICS		
SETORES	FREQ	%
administração de empresas e empreendimentos	25,0	6,3
agricultura, pecuária, silvicultura, pesca e caça	4,0	1,0
artes, entretenimento e recreação	1,0	0,3
assistência médica e social	5,0	1,3
Comércio atacadista	2,0	0,5
Comércio varejista	13,0	3,3

CLASSIFICAÇÃO NAICS		
SETORES	FREQ	%
Construção	36,0	9,0
Educação	5,0	1,3
empresa de eletricidade, gás e água	47,0	11,8
hotel e restaurante	2,0	0,5
imobiliária e locadora de outros bens	3,0	0,8
Indústria manufatureira	155,0	38,9
Informação	25,0	6,3
Mineração	7,0	1,8
outros serviços (exceto administração pública)	2,0	0,5
Serviços de apoio a empresa e ger. res. e rem.	5,0	1,3
Serviços financeiros e seguros	41,0	10,3
Serviços profis., científicos e técnicos	3,0	0,8
transporte e armazenamento	17,0	4,3
TOTAL	398	100
MÉDIA	20,9	5,3

Pela Tabela 06 observa-se uma forte concentração na “indústria manufatureira” que apresenta 38,94% do total, além disso, observa-se que somente 3 setores apresentam uma concentração em torno dos 10%, que são os setores de “serviços financeiros e seguros” (10,3 %), “empresa de eletricidade gás e água” (11,8 %) e “construção” (9,05 %). Sendo os que apresentam menor concentração em torno de 0,5 % os setores de “artes, entretenimento e recreação” (0,3 %), “comércio atacadista” (0,5 %), “hotel e restaurante” (0,5 %) e “outros serviços” (0,5 %).

4.4 Estatísticas descritivas

As estatísticas descritivas possibilitam a avaliação do comportamento de cada uma das variáveis utilizadas no estudo sob o ponto de vista univariado. Tal comportamento foi avaliado a partir das seguintes estatísticas: número de casos válidos (N), valor mínimo da distribuição (MIN), valor máximo da distribuição (MAX), média (M), mediana (MED), desvio padrão (DP) e coeficientes de assimetria (ASSIM) e curtose (CURT). A análise refere-se às estatísticas descritivas das variáveis médias. Conforme já descrito no capítulo de método, as variáveis médias indicam a média dos valores anuais para as respectivas variáveis. Salienta-se que as estatísticas se referem aos valores antes da padronização das variáveis, uma vez que, após a padronização, as variáveis apresentam média igual a zero e desvio padrão igual a um.

Tabela 07 – Número de casos válidos, valores mínimo e máximo, média, mediana, desvio-padrão, assimetria e curtose das variáveis médias

VARIÁVEIS MÉDIAS	N	MÍN.	MÁX.	M	MED	DP	ASSIM.	CURT.
liquidez corrente	394	0,0	190,5	2,5	1,3	10,2	16,5	296,6
liquidez seca	364	0,0	190,5	2,3	1,1	10,6	16,0	276,7
margem operacional	351	-429090,6	16802,2	-1413,6	4,5	23436,2	-17,6	320,0
margem ebitida	342	-61215,2	5274,9	-200,0	13,8	3378,0	-17,5	315,0
rentabilidade do ativo	398	-91766,7	66,9	-329,4	2,0	4745,3	-18,3	350,4
rentab. Patrim. líquido	381	-12409,7	3666,9	-37,3	7,5	676,0	-15,8	299,5
exigível total/ativo total	398	1,8	20507,6	261,8	65,2	1494,6	10,3	119,2
Endividamento geral	398	-34613,1	14958,4	339,4	145,8	2156,8	-9,4	181,1
Endivid. financeiro	338	0,3	100,0	50,8	46,6	27,2	0,4	-0,9
Dividendo <i>yeld</i>	249	0,0	53,7	5,8	4,2	7,0	4,0	21,9
<i>Payout</i>	289	-2440,9	2736,0	55,4	38,8	235,5	1,3	102,3
valor patrim. ação	398	-20008,0	1013,4	-36,2	4,3	1014,8	-19,3	380,7
liquidez em bolsa	373	0,0	8,1	0,2	0,0	0,8	6,8	58,1
Retorno	289	-0,5	8,0	0,5	0,3	0,9	4,9	31,2
ln (valor de mercado)	370	5,8	18,5	12,8	13,3	2,5	-0,5	-0,4
ln (receita líquida)	351	4,5	18,4	12,6	12,7	2,2	-0,8	1,3
ln (ativototal)	398	1,5	19,2	13,2	13,4	2,4	-0,9	3,0

M = média; MED = mediana; DP = desvio padrão

Da Tabela 07 pode-se observar que todas as variáveis apresentam valores expressivos de assimetria e curtose. Quanto à assimetria, as duas variáveis relacionadas à margem (margem operacional e margem ebitida) e as duas relacionadas à rentabilidade (rentabilidade do ativo e rentabilidade do patrimônio líquido) e o endividamento geral apresentam assimetria negativa indicando que as caudas são alongadas à esquerda. As demais variáveis apresentam assimetria positiva, o que de um modo geral indicam que, em média, a mensuração de lucratividade dada pelas margens sinaliza empresas com capacidades menores do que o ponto de máximo se a distribuição fosse normal, menor capacidade de gerar lucro. Como também, menores capacidade da empresa garantir os capitais de terceiros.

No que se referem à curtose, todas as variáveis médias assumiram valores positivos indicando que as curvas são mais alongadas do que a curva normal. Em mediana as ações apresentam liquidez corrente e seca em torno de 1,34 e 1,11 respectivamente. Em relação à margem apresenta mediana em torno de 4,5 para a operacional e 13,8 para a ebitida; quanto à rentabilidade do ativo apresentam em mediana em torno de 2,0 e a rentabilidade do patrimônio líquido em torno de 7,5. Ainda, apresentam endividamento geral e financeiro, em mediana, em torno de 145,8 e 46,6 respectivamente; dividendo *yeld*, em mediana, em torno de 4,2 e *payout* com mediana em torno de 38,8.

Com relação à participação no mercado suas ações têm valor patrimonial mediano em torno de 4,3, liquidez em bolsa com mediana em torno de 0,01 e retorno anual mediano igual a 0,3. Com relação ao tamanho, como as variáveis foram calculadas utilizando-se o logaritmo natural pode-se deduzir que o valor de mercado em mediana apresenta-se em torno de 603.197,53 ($e^{13,31} = 603.197,53$) ao passo que em mediana o valor apresentado para a receita líquida está em torno de 334.368,85 ($e^{12,72} = 334.368,85$) e por fim em mediana o ativo total apresenta-se em torno de 673.336,17 ($e^{13,42} = 673.336,17$). As estatísticas descritivas para cada ano do estudo podem ser visualizadas no Apêndice B.

Para análise da diferença univariada entre os setores de atividade, propostos pelas classificações BOVESPA e NAICS, optou-se pela utilização do Teste de Kruskal Wallis. As variáveis utilizadas para tal avaliação referem-se às variáveis médias do período, por exemplo, a variável liquidez corrente é a média dos valores da liquidez corrente nos anos de 2000 a 2007, conforme foi definido no método.

A Tabela 08 apresenta o valor do teste e sua respectiva significância.

Tabela 08 - Teste Kruskal Wallis para as variáveis médias segundo as classificações setoriais da NAICS e BOVESPA

TESTE KRUSKAL-WALLIS PARA CLASSIFICAÇÃO SETORIAL				
VARIÁVEIS MÉDIAS	NAICS		BOVESPA	
	Qui-quadrado	p -valor	Qui-quadrado	p -valor
liquidez corrente	73,97	0,00	39,77	0,00
liquidez seca	62,50	0,00	33,34	0,00
margem operacional	44,05	0,00	47,33	0,00
margem ebitida	61,85	0,00	77,25	0,00
rentabilidade do ativo	19,26	0,38	25,84	0,00
rentabilidade patrimônio líquido	37,32	0,01	44,60	0,00
exigível total/ativo total	51,55	0,00	15,05	0,09
endividamento geral	58,39	0,00	21,82	0,01
endividamento financeiro	73,39	0,00	59,47	0,00
dividendo <i>yeld</i>	50,87	0,00	37,56	0,00
<i>Payout</i>	22,43	0,10	15,76	0,07
valor patrimonial da ação	35,06	0,01	18,24	0,03
liquidez em bolsa	40,37	0,00	31,79	0,00
ln (valor de mercado)	74,50	0,00	54,49	0,00
ln (receita líquida)	83,74	0,00	76,82	0,00
ln (ativo total)	91,47	0,00	86,48	0,00
retorno	46,16	0,00	33,39	0,00

Para a grande maioria das variáveis o valor de significância é inferior a 0,05 para as duas classificações setoriais indicando que existe diferença significativa entre os setores de atividade para as referidas variáveis. Apenas as variáveis “rentabilidade do ativo” e “*payout*” não apresentaram diferença significativa na análise segundo a classificação NAICS, e segundo a classificação da BOVESPA, as variáveis médias exigível total/ativo total e *payout* não foram significativas. De maneira geral, os resultados do Teste Kruskal Wallis mostram que, do ponto de vista univariado, o setor de atividade é um fator de classificação preponderante uma vez que as medianas de cada setor diferem estatisticamente.

Seguindo o procedimento adotado na análise anterior, o Teste Kruskal Wallis também foi aplicado anualmente e os resultados são apresentados no Apêndice C. Observa-se que estes resultados são muito semelhantes aos apresentados na Tabela 08 indicando que as diferenças setoriais não são fortemente afetadas pelo ano de estudo, já que as diferenças se conservam ao longo do tempo.

4.5 Análise dos *clusters*

Conforme definido anteriormente, a análise dos *clusters* será realizada para dois conjuntos de dados. O primeiro, denominado *cluster* índice e o segundo denominado *cluster* mercado, ambos envolvem a estimação dos *clusters* a partir das variáveis apresentadas no Quadro 05 abaixo. Para tais *clusters* será feita uma avaliação da influência da utilização de variáveis para o controle de tamanho. Assim, primeiramente, os *clusters* serão estimados sem as variáveis de tamanho e, posteriormente, tais variáveis serão inseridas.

CLUSTER ÍNDICE VARIÁVEIS	CLUSTER MERCADO VARIÁVEIS	CONTROLE VARIÁVEIS
liquidez corrente	valor patrimonial da ação	ln (valor de mercado)
liquidez seca	liquidez em bolsa	ln (receita líquida)
margem operacional	retorno	ln (ativo total)
margem ebitida		
rentabilidade do ativo		
rentabilidade patrimônio líquido		
<i>Payout</i>		
dividendo <i>yeld</i>		
exigível total/ativo total		
endividamento geral		
endividamento financeiro		

Quadro 05 – Variáveis formadoras dos *clusters* índice e mercado, e variáveis de controle

Para a definição do número de *clusters* serão utilizados dois métodos. O primeiro, denominado método de parada envolve o procedimento descrito no item 3.7.4. O segundo tem o número de *clusters* previamente definido conforme o número de setores existentes em cada classificação, ou seja, 10 para a classificação BOVESPA e 20 para a classificação NAICS. Assim, no item 4.5.1 tem-se uma descrição dos métodos utilizados na definição do número de *clusters*. A construção dos *clusters* é apresentada no item 4.6 e finalmente têm-se os resultados de controle para o tamanho no item 4.7.

4.5.1 Definição do número de *clusters*

Como forma de ilustração do procedimento descrito no item 3.7.4 do capítulo do método apresenta-se o número de *clusters* a partir da matriz de aglomeração (*agglomeration schedule*) complementada pela exibição da diferença de coeficientes e com as respectivas mudanças de percentuais considerando-se o *cluster* mercado referente ao ano 2000, com a inclusão das variáveis de tamanho.

Etapas	COMBINADOS		COEF	DIF. ENTRE COEF.	MUDANÇA % PRÓXIMO NÍVEL	PRIMEIRA ETAPA DE AGLOMERAÇÃO		Nova etapa	Nº de Clusters
	Cluster 1	Cluster 2				Cluster 1	Cluster 2		
1	317	324	0,00000			0	0	2	
2	66	317	0,00000			0	1	4	
3	264	315	0,00000			0	0	4	
4	66	264	0,00000			2	3	6	
15	94	293	0,00000	0,00000		0	0	42	
18	147	269	0,00000	0,00000		0	0	26	
...
273	61	360	14,07900	0,93800	0,07138	265	0	281	14
274	22	62	15,02900	0,95000	0,06748	271	255	280	13
275	27	45	16,54200	1,51300	0,10067	267	269 275	277 282	12 10
279	5	124	28,06600	5,09800	0,22196	278	272	282	8
280	1	22	33,91600	5,85000	0,20844	266	274	284	7
281	61	265	41,51100	7,59500	0,22394	273	251	285	6
282	5	15	50,41800	8,90700	0,21457	279	277	284	5
283	113	261	60,84900	10,43100	0,20689	0	0	285	4
284	1	5	72,98100	12,13200	0,19938	280	282	286	3
285	61	113	88,08100	15,10000	0,20690	281	283	286	2
286	1	61	118,02400	29,94300	0,33995	284	285	0	1

Quadro 06 – Dados da matriz de aglomeração para o *cluster* mercado e sem as variáveis de tamanho, para o ano de 2000

Assim, por meio da análise do Quadro 06, que exhibe os resultados da avaliação da mudança de percentual ocorrida quando se altera o número de clusters, observa-se que o coeficiente de aglomeração demonstra maiores ganhos ao passar de 4 para 3 *clusters*. Nessa situação opta-se por tomar como 3 o número adequado de *clusters*. Adotando-se o mesmo procedimento, para os demais anos chega-se ao seguinte quadro resumo.

ANO	NÚMERO DE <i>CLUSTERS</i>			
	SEM AS VARIÁVEIS DE TAMANHO		COM AS VARIÁVEIS DE TAMANHO	
	<i>Clusters índices</i>	<i>Clusters mercados</i>	<i>Clusters índices</i>	<i>Clusters mercados</i>
2000	10	3	4	6
2001	4	5	5	6
2002	5	2	5	9
2003	4	9	6	11
2004	4	4	3	2
2005	6	2	3	2
2006	8	7	7	6
2007	10	4	11	10

Quadro 07 – Definição do número de *clusters* considerando ou não as variáveis de tamanho na construção dos *clusters* no período de 2000 – 2007

Para os *clusters* índices (CI) o número de *clusters* após a inclusão das variáveis de tamanho observa-se um aumento do número de *clusters* em três anos do período (2001, 2003, 2007), permanecendo igual em um ano (2002) e menor nos demais. Para as variáveis de mercado (CM) a inclusão das variáveis de tamanho provoca um aumento do número de *clusters* em mais de 60% dos anos do período. Mais especificadamente, para os *clusters* índices e mercados, sem a inclusão das variáveis de tamanho, o número predominante de *clusters* são respectivamente 4 e 2.

No segundo método, o número de *clusters* leva em conta o número de subdivisões das classificações adotadas no estudo, ou seja, 10 (subdivisão da BOVESPA) e 20 (subdivisão da NAICS nível 1). Entretanto, quando o número de *clusters* foi estipulado por 10 nem todos os *clusters* apresentaram mais de 1 elemento, como pode ser visualizado no quadro 8, sendo que o número de *clusters* exibidos representa os *clusters* com pelo menos 2 elementos.

CLASSIFICAÇÃO BOVESPA – COM 10 <i>CLUSTERS</i>				
ANO	SEM VARIÁVEIS DE TAMANHO		COM VARIÁVEIS DE TAMANHO	
	<i>Cluster índice</i>	<i>Cluster mercado</i>	<i>Cluster índice</i>	<i>Cluster mercado</i>
2000	8	8	8	8
2001	9	9	9	9
2002	7	9	7	9
2003	7	9	7	9
2004	7	9	8	9
2005	9	8	8	9
2006	6	8	6	10
2007	10	9	10	10

Quadro 08 – Definição do número de clusters segundo o mesmo número de subdivisões da classificação da BOVESPA, ou seja, 10, no período de 2000 – 2007

Optando-se pela formação do número de *clusters* igual a 10 nota-se que a inclusão das variáveis de tamanho nos *clusters* índice e mercado praticamente não altera o número de *clusters*. Observa-se também que o número máximo de *clusters* é 10 e ocorre somente no último período. Quando tomamos o número de *clusters* igual a 20, os *clusters* índice e mercado não apresentam grandes diferenças em relação ao número de *clusters* quando da inclusão das variáveis de tamanho, o que pode ser verificado no Apêndice D.

4.6 Construção dos *clusters*

Conforme estabelecido anteriormente nesta etapa da análise os *clusters* serão estimados para os dois conjuntos de dados sem as variáveis de tamanho. O primeiro denominado *cluster* índice envolve todas as variáveis menos as de mercado. O segundo denominado *cluster* mercado envolve só as variáveis de mercado. Para cada um destes *clusters* os resultados são avaliados considerando as classificações segundo a BOVESPA e NAICS, levando-se em conta o número de *cluster* sugerido pelo método da parada, o número de *clusters* igual a 10 (classificação BOVESPA) e o número de *cluster* igual a 20 (classificação NAICS). Assim, para cada um dos conjuntos apresenta-se um total de 6 estimações por ano para análise. Ressalta-se ainda que, apenas as tabelas referentes aos dois grupos de *clusters* índice e mercado com a classificação BOVESPA são apresentadas ao longo do texto.

4.6.1 Análise dos *clusters* segundo a regra de parada

Esta análise é feita considerando-se as duas classificações BOVESPA e NAICS, para tanto se apresenta as tabelas de 09 a 17 dispostas segundo a classificação da BOVESPA e as tabelas de 18 a 25 dispostas segundo a classificação NAICS, ambas com o número de *clusters* sugeridos pelo critério de parada conforme exemplificado na seção 4.5.1.

Inicialmente apresenta-se a Tabela 09 com a classificação BOVESPA para os *clusters* índice e mercado considerando o ano 2000.

Tabela 09 – Frequência percentual e total de casos, segundo a classificação BOVESPA, no ano 2000

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2000													
SETORES	Cluster índice (%)										Cluster mercado (%)		
	1	2	3	4	5	6	7	8	9	10	1	2	3
bens industriais	8,0	23,5	11,1	0,0	0,0	8,3	0,0	0,0	0,0	0,0	11,6	0,0	0,0
construção e transporte	8,0	13,7	11,1	33,3	0,0	10,2	0,0	0,0	0,0	0,0	10,9	0,0	0,0
Consumo cíclico	20,0	27,5	15,9	0,0	33,3	14,8	0,0	33,3	0,0	0,0	18,5	11,1	50,0
Consumo não cíclico	6,0	11,8	4,8	0,0	0,0	8,3	0,0	0,0	0,0	0,0	7,6	0,0	0,0
financeiro e outros	20,0	3,9	3,2	0,0	50,0	25,9	100,0	0,0	100,0	0,0	17,0	0,0	0,0
Materiais básicos	22,0	17,6	9,5	0,0	0,0	18,5	0,0	33,3	0,0	0,0	16,7	11,1	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	0,0	1,9	0,0	0,0	0,0	0,0	0,4	11,1	0,0
tecnologia da informação	2,0	2,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,7	0,0	0,0
telecomunicações	6,0	0,0	12,7	0,0	16,7	4,6	0,0	0,0	0,0	100,0	4,7	55,6	0,0
utilidade pública	8,0	0,0	31,7	66,7	0,0	7,4	0,0	33,3	0,0	0,0	12,0	11,1	50,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	50	51	63	3	6	108	1	3	1	1	276	9	2

Segundo o critério da regra de parada para a decisão do número de *clusters* dos 10 *clusters* sugeridos para o *cluster* índice observa-se pela Tabela 09 que o *cluster* 6 (37,6%) e os *clusters* 1 (17,4%), 2 (17,8%) e 3 (22%) concentram a maioria das ações da amostra para o ano 2000. Observa-se que nestes *clusters* as ações estão distribuídas em diversos setores, entretanto, com concentrações pequenas em torno de 25% nos setores mais concentrados o que sugere a não homogeneidade setorial. Apenas para o *cluster* 5 com apenas 6 casos a concentração é um pouco mais expressiva com 50% dos casos no setor “financeiro e outros”. Com relação ao *cluster* mercado, o *cluster* 1 apresenta uma concentração de 96,2% dos casos sendo que os setores da classificação BOVESPA com concentração maior do que 10% referem-se aos setores de “consumo cíclico” (18,5%), “financeiros e outros” (17%),

“materiais básicos” (16,7%), “utilidade pública” (12%), “bens industriais” (11,6%) e “construção e transporte” (10,9%).

A Tabela 10 apresenta a classificação BOVESPA para os *clusters* índices e mercado para o ano 2001.

Tabela 10 – Classificação BOVESPA, com 4 *clusters* para *cluster* índice e 5 *clusters* para *cluster* mercado, no ano 2001

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2001									
SETORES	Cluster índice				Cluster mercado				
	1	2	3	4	1	2	3	4	5
bens industriais	9,4	16,9	20,0	0,0	7,7	11,1	34,3	0,0	0,0
construção e transporte	9,9	14,1	0,0	0,0	11,0	12,7	8,6	0,0	0,0
consumo cíclico	15,6	28,2	20,0	0,0	19,9	22,2	8,6	10,0	0,0
consumo não cíclico	5,7	12,7	0,0	0,0	7,7	4,8	11,4	0,0	0,0
financeiro e outros	18,4	5,6	40,0	100,0	20,4	7,9	11,4	10,0	0,0
materiais básicos	17	15,5	0,0	0,0	14,4	25,4	11,4	10,0	0,0
petróleo, gás e biocomb.	0,9	0	0,0	0,0	0,6	0,0	0,0	10,0	0,0
tecnologia da informação	0,5	1,4	0,0	0,0	0,0	3,2	0,0	0,0	0,0
telecomunicações	7,5	1,4	20,0	0,0	6,1	3,2	0,0	50,0	0,0
utilidade pública	15,1	4,2	0,0	0,0	12,2	9,5	14,3	10,0	100,0
TOTAL (%)	100	100	100	100	100	100	100	100	100
TOTAL CASOS	212	71	5	2	181	63	35	10	1

A Tabela 10 apresenta para o *cluster* índice uma concentração no *cluster* 1 de 73,10% seguido de 24,5% no *cluster* 2, ou seja, estes dois *clusters* correspondentes à metade do número de *clusters* para esse conjunto guardam um total de 97,6% dos casos. Observa-se que nestes *clusters* as ações estão distribuídas em diversos setores, com pequenas concentrações, sendo o de maior concentração apresentado no *cluster* 2 (28,2%). Para o *cluster* mercado o *cluster* 1 apresenta uma concentração de 62,4%, seguido dos *clusters* 2 (21,7%) e 3 (12,10%). Nesses *clusters* as ações encontram-se distribuídas em diversos setores sendo as maiores concentrações em torno de 25% como, por exemplo, o setor de “materiais básicos” (25,4%) do *cluster* 2.

A Tabela 11 apresenta a classificação BOVESPA para os *clusters* índice e mercado para o ano 2002.

Tabela 11 – Classificação BOVESPA, com 5 clusters para o cluster índice e 2 clusters para o cluster mercado, no ano 2002

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2002								
SETORES	Cluster índice					Cluster mercado		
	1	2	3	4	5	1	2	
Bens industriais	9,8	18,0	7,5	0,0	25,0	11,6	9,6	
construção e transporte	11,7	16,0	6,0	0,0	8,3	11,2	9,6	
consumo cíclico	17,8	26,0	10,4	50,0	33,3	19,8	11,5	
consumo não cíclico	3,7	14,0	10,4	0,0	16,7	7,4	7,7	
financeiro e outros	22,1	4,0	9,0	50,0	16,7	17,8	7,7	
materiais básicos	16,0	16,0	20,9	0,0	0,0	13,6	28,8	
petróleo, gás e biocomb.	0,0	2,0	1,5	0,0	0,0	0,4	1,9	
tecnologia da informação	1,2	0,0	0,0	0,0	0,0	0,8	0,0	
telecomunicações	5,5	0,0	13,4	0,0	0,0	4,1	15,4	
utilidade pública	12,3	4,0	20,9	0,0	0,0	13,2	7,7	
TOTAL (%)	100	100	100	100	100	100	100	
TOTAL CASOS	163	50	67	2	12	242	52	

Observa-se pela Tabela 11 que no cluster índice o cluster que apresenta maior número de casos é o cluster 1 perfazendo um total de 55,4%, os clusters 2 e 3 apresentam percentuais de 17% e 22,8% respectivamente, sendo que estes três clusters juntos representam um total de 95,2% dos casos. Os clusters 1, 2 e 3 apresentam uma classificação setorial dispersa sendo o setor de “consumo cíclico” do cluster 2 o que apresentou maior concentração (26%). Para o cluster mercado o cluster 1 apresenta uma concentração de 82,3% do total de casos, ou seja, apresenta uma considerável concentração. Quanto à classificação setorial desse cluster observa-se que o setor que apresenta maior concentração é o setor de “consumo cíclico” (19,8%). A Tabela 12 apresenta a classificação BOVESPA para os clusters índice e mercado para o ano 2003.

Tabela 12 – Classificação BOVESPA, com 4 clusters para o cluster índice e 9 clusters para cluster mercado no ano 2003

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2003													
SETORES	Cluster índice				Cluster mercado								
	1	2	3	4	1	2	3	4	5	6	7	8	9
bens industriais	7,3	22,7	7,1	22,2	9,4	30,0	16,9	0,0	9,1	7,7	16,7	0,0	0,0
construção e transporte	10,9	9,1	10,7	22,2	12,9	20,0	8,5	0,0	12,1	12,3	8,3	0,0	0,0
consumo cíclico	16,1	28,8	7,1	22,2	15,3	30,0	11,9	0,0	15,2	21,5	45,8	6,3	0,0
consumo não cíclico	6,7	9,1	10,7	0,0	10,6	10,0	6,8	0,0	12,1	4,6	0,0	6,3	0,0
financeiro e outros	20,2	3,0	14,3	22,2	22,4	0,0	5,1	0,0	21,2	21,5	4,2	18,8	0,0
materiais básicos	15,0	19,7	17,9	11,1	8,2	0,0	39,0	0,0	6,1	15,4	8,3	25,0	0,0
petróleo, gás e biocomb.	0,5	1,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,5	0,0	0,0	33,3
tecnologia da informação	1,0	0,0	0,0	0,0	0,0	0,0	1,7	0,0	3,0	0,0	0,0	0,0	0,0
Telecomunicações	7,8	1,5	7,1	0,0	7,1	0,0	0,0	0,0	3,0	3,1	12,5	25,0	66,7
utilidade pública	14,5	4,5	25,0	0,0	14,1	10,0	1,0	100,0	18,2	12,3	4,2	18,8	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	193	66	28	9	85	10	59	1	33	65	24	16	3

Novamente nos *clusters* mais representativos as ações encontram-se distribuídas nos diversos setores sendo que os com maiores concentrações de casos apresentam um percentual na faixa dos 20%. Para o *cluster* mercado, os *clusters* que apresentam um percentual de concentração de casos acima de 10% são os *clusters* 1 (28,7%), 3 (19,9%), 5 (11%) e o 6 (22%) que juntos representam 81,6% dos casos.

A Tabela 13 apresenta a classificação BOVESPA para os *clusters* índice e mercado para o ano 2004.

Tabela 13 – Classificação BOVESPA, num total de 4 *clusters* para o *cluster* índice e 4 *clusters* para o *cluster* mercado, no ano 2004

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2004								
SETORES	Cluster índice				Cluster mercado			
	1	2	3	4	1	2	3	4
Bens industriais	9,3	18,3	0,0	16,7	8,7	45,5	0,0	0,0
construção e transporte	11,5	11,7	0,0	0,0	11,8	9,1	0,0	0,0
consumo cíclico	14,1	30,0	33,3	33,3	19,8	9,1	0,0	0,0
consumo não cíclico	7,5	11,7	0,0	16,7	8,4	9,1	0,0	7,7
financeiro e outros	17,2	6,7	66,7	16,7	17,5	0,0	0,0	15,4
materiais básicos	16,3	13,3	0,0	16,7	14,8	13,6	0,0	30,8
petróleo, gás e biocomb.	0,4	1,7	0,0	0,0	0,4	0,0	0,0	7,7
tecnologia da informação	0,4	1,7	0,0	0,0	0,8	0,0	0,0	0,0
telecomunicações	7,0	3,3	0,0	0,0	5,3	0,0	0,0	30,8
utilidade pública	16,3	1,7	0,0	0,0	12,5	13,6	100,0	7,7
TOTAL (%)	100	100	100	100	100	100	100	100
TOTAL CASOS	227	60	6	6	263	22	1	13

Observa-se pela Tabela 13 que no *cluster* índice o *cluster* 1 com 65,2% é o mais expressivo. As ações encontram-se distribuídas nos diversos setores, sendo os que apresentam maiores concentrações apresentam um percentual na faixa dos 15%, a exceção do setor “consumo cíclico” do *cluster* 2 que apresenta uma concentração de 30%. Do mesmo modo para o *cluster* mercado o *cluster* 1 predomina com 88% dos casos. A distribuição nos setores desses *clusters* apresenta pequenas concentrações nos diversos setores, sendo as maiores em torno dos 15%, a exceção do setor “bens industriais” no *cluster* 2 que apresenta uma concentração bem expressiva (45,5%). Observa-se ainda que para a maioria dos setores, as empresas se dividem entre dois ou mais *clusters*, mostrando que não há aglomeração das empresas de um mesmo setor em um único *cluster*.

A Tabela 14 apresenta a classificação BOVESPA para os *clusters* índice e mercado para o ano 2005.

Tabela 14 – Classificação BOVESPA, com 6 clusters para o cluster índice e 2 clusters para o cluster mercado, no ano 2005

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2005									
SETORES	Cluster índice						Cluster mercado		
	1	2	3	4	5	6	1	2	
bens industriais	7,1	21,2	5,0	9,2	0,0	33,3	11	3,8	
construção e transporte	10,6	16,7	10,0	9,2	0,0	0,0	11	11,5	
consumo cíclico	13,3	27,3	20,0	16,3	50,0	0,0	19	7,7	
consumo não cíclico	8,0	9,1	5,0	10,2	0,0	0,0	8,6	7,7	
financeiro e outros	4,4	7,6	20,0	32,7	16,7	66,7	16	15,4	
materiais básicos	12,4	10,6	30,0	17,3	16,7	0,0	14	23,1	
petróleo, gás e biocomb.	0,9	1,5	0,0	0,0	0,0	0,0	0,4	3,8	
tecnologia da informação	0,0	1,5	5,0	1,0	0,0	0,0	1,1	0,0	
telecomunicações	12,4	0,0	5,0	2,0	16,7	0,0	5,4	11,5	
utilidade pública	31,0	4,5	0,0	2,0	0,0	0,0	13	15,4	
TOTAL (%)	100	100	100	100	100	100	100	100	
TOTAL CASOS	113	66	20	98	6	3	272	26	

No *cluster* 1 predomina o setor de “utilidade pública” (31%) e no *cluster* 4 o “setor financeiro e outros” (32,7). Já para o *cluster* mercado o *cluster* 1 acumula a grande maioria dos casos (88%) e os mesmos se distribuem em diversos setores, o que mostra uma grande heterogeneidade setorial. A Tabela 15 apresenta a classificação BOVESPA para os *clusters* índice e mercado para o ano 2006.

Tabela 15 – Classificação BOVESPA, com 8 clusters para o cluster índice e 7 clusters para o cluster mercado, no ano 2006

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2006																
SETORES	Cluster índice								Cluster mercado							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	
bens industriais	5,0	11,5	22,6	4,0	9,7	50,0	0,0	0,0	16,7	14,2	25,0	7,3	11,1	0,0	0,0	
construção e transporte	9,0	14,2	16,1	8,0	3,2	0,0	0,0	0,0	0,0	12,3	25,0	13,4	3,7	14,3	0,0	
Consumo cíclico	18,0	14,2	24,2	24,0	3,2	0,0	0,0	0,0	33,3	16,1	0,0	12,2	33,3	3,6	0,0	
consumo não cíclico	10,0	12,4	6,5	8,0	3,2	0,0	0,0	100,0	0,0	7,7	25,0	12,2	7,4	14,3	0,0	
financeiro e outros	7,0	27,4	3,2	24,0	22,6	0,0	100,0	0,0	0,0	14,8	12,5	22,0	11,1	21,4	0,0	
Materiais básicos	18,0	11,5	12,9	8,0	16,1	0,0	0,0	0,0	0,0	12,3	12,5	7,3	25,9	17,9	0,5	
petróleo, gás e biocomb.	2,0	0,0	1,6	0,0	0,0	50,0	0,0	0,0	0,0	1,3	0,0	0,0	1,9	0,0	0,5	
tecnologia da informação	0,0	4,4	1,6	0,0	0,0	0,0	0,0	0,0	0,0	2,6	0,0	1,2	1,9	0,0	0,0	
telecomunicações	4,0	1,8	6,5	16,0	9,7	0,0	0,0	0,0	16,7	1,9	0,0	8,5	1,9	17,9	0,0	
Utilidade pública	27,0	2,7	4,8	8,0	32,3	0,0	0,0	0,0	33,3	16,8	0,0	15,9	1,9	10,7	0,0	
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
TOTAL CASOS	100	113	62	25	31	2	1	1	6	155	8	82	54	28	2	

Da Tabela 15 observa-se que para o *cluster* índice novamente predominam dois *clusters* (1 e 2) sendo estes bastantes heterogêneos. Por outro lado, os *clusters* 7 e 8 apresentam apenas 1 caso, portanto não são representativos. A Tabela 16 apresenta a classificação BOVESPA para os *clusters* índice e mercado para o ano 2007.

Tabela 16 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice e 4 *clusters* para o *cluster* mercado, no ano 2007

CLASSIFICAÇÃO BOVESPA – MÉTODO DE PARADA, ANO 2007														
SETORES	Cluster índice										Cluster mercado			
	1	2	3	4	5	6	7	8	9	10	1	2	3	4
bens industriais	15,0	17,4	4,8	0,0	0,0	0,0	0,0	20,0	25,0	0,0	10,5	14,3	2,4	11,1
construção e transporte	11,3	15,9	15,7	12,9	0,0	13,3	39,3	0,0	25,0	0,0	14,7	0,0	14,6	22,2
consumo cíclico	15,8	29,0	8,4	6,5	9,5	40,0	7,1	20,0	0,0	50,0	16,2	28,6	7,3	33,3
consumo não cíclico	12,8	11,6	9,6	3,2	0,0	13,3	10,7	0,0	0,0	0,0	9,9	0,0	14,6	0,0
financeiro e outros	22,6	2,9	8,4	58,1	4,8	20,0	25,0	40,0	25,0	0,0	18,6	0,0	14,6	33,3
materiais básicos	12,0	11,6	24,1	6,5	0,0	13,3	3,6	0,0	0,0	50,0	12,0	28,6	19,5	0,0
Petróleo, gás e biocomb.	1,5	1,4	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,9	14,3	0,0	0,0
tecnologia da informação	0,8	4,3	0,0	0,0	0,0	0,0	10,7	0,0	0,0	0,0	2,1	0,0	0,0	0,0
Telecomunicações	5,3	0,0	6,0	6,5	4,8	0,0	3,6	20,0	0,0	0,0	3,6	0,0	12,2	0,0
Utilidade pública	3,0	5,8	21,7	6,5	81,0	0,0	0,0	0,0	25,0	0,0	11,7	14,3	14,6	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	133	69	83	31	21	15	28	5	4	2	334	7	41	9

Da Tabela 16 observa-se que para o *cluster* índice apenas os *clusters* 1, 2 e 3 com os respectivos percentuais de 34%, 17,6% e 21,20% representam juntos 72,8% dos casos, esses *clusters* apresentam uma distribuição setorial com concentrações pequenas nos setores, ficando as maiores em torno de 20%, e o setor que apresenta maior concentração é o setor “financeiro e outros” (18,6%) e a menor concentração é o setor “petróleo, gás e bicomustíveis” (0,9%). As Tabelas 17 e 18 apresentam um resumo para os *clusters* índice e mercado respectivamente, para o critério regra de parada, segundo a classificação BOVESPA.

Tabela 17 – Resumo da classificação setorial BOVESPA, para o *cluster* índice no período 2000 - 2007

CLUSTER ÍNDICE – CRITÉRIO REGRA DE PARADA						
ANO	Nº CLUSTERS	QUANTIDADE DE SETORES			FREQUÊNCIA SETORIAL	
		MAX	MIN	MÉDIA	MÁXIMA	SETOR
2000	10	9	1	4,4	25,9	financeiro e outros
2001	4	10	1	6,0	18,4	financeiro e outros
2002	5	9	2	6,6	22,1	financeiro e outros
2003	4	10	5	8,0	20,2	financeiro e outros
2004	4	10	2	6,8	17,2	financeiro e outros
2005	6	9	2	6,3	31,0	utilidade pública
2006	8	10	1	6,0	24,2	consumo cíclico
2007	10	10	2	6,1	22,6	financeiro e outros

Dos resultados exibidos na Tabela 17 observa-se, pelo valor mínimo que, para a maioria dos anos, foi obtido pelo menos um *cluster* com empresas de um único setor. Tais *clusters*, geralmente, são aqueles formados por apenas uma ou duas empresas. O número máximo de setores por *cluster* foi 9 para os anos 2000, 20002 e 2005, e 10 para os anos de 2001, 2003, 2004, 2006 e 2007. Em média há 6,3 setores por *clusters* e a frequência máxima em um único setor é em torno de 22,7%.

Tabela 18 – Resumo da classificação setorial BOVESPA, para o *cluster* mercado no período 2000 - 2007

CLUSTER MERCADO - CRITÉRIO REGRA DE PARADA						
ANO	Nº CLUSTERS	QUANTIDADE DE SETORES			FREQUÊNCIA SETORIAL MÁXIMA	SETOR
		MAX	MIN	MÉDIA		
2000	3	10	2	7,1	18,5	consumo cíclico
2001	5	9	1	6,4	25,4	materiais básicos
2002	2	10	9	9,5	19,8	consumo cíclico
2003	9	9	1	6,1	21,5	financeiros/outros, consumo cíclico
2004	4	10	1	6,8	19,8	consumo cíclico
2005	2	10	9	9,5	19,3	consumo cíclico
2006	7	10	2	6,7	33,3	consumo cíclico
2007	4	10	4	6,8	18,6	financeiros e outros

Dos resultados exibidos na Tabela 18 observa-se pelo valor mínimo que para os anos, de 2001, 2003 e 2004 foi obtido pelo menos um *cluster* com empresas de um único setor, em geral, tais *clusters* são aqueles formados por apenas uma ou duas empresas. O número máximo de setores por *cluster* foi 10 para a grande parte dos anos. Em média existem 7,4 setores por *clusters* e a frequência máxima em um único setor em torno de 17,7%.

Portanto, como a classificação BOVESPA possui 10 setores e como, por este critério, o número máximo de *clusters* estimado foi 9 (2003), caso os setores fossem heterogêneos para esse ano, esperar-se-ia que cada cluster apresentasse empresas de um setor. Como os resultados demonstram uma alta heterogeneidade setorial em cada cluster então tais resultados apontam para a ineficiência da análise setorial, nesse caso. Os *clusters* índice e os *clusters* mercado estimados segundo a regra de parada também foram classificados segundo o critério NAICS. O Apêndice E, apresenta os resultados para cada um dos anos do período em estudo. Os resultados obtidos não se diferenciam expressivamente dos apresentados para a classificação BOVESPA. No que se refere à concentração amostral para o *cluster* índice na maioria dos anos dois *clusters* concentram um percentual expressivo da amostra. Por

exemplo, o ano com menor concentração de *clusters* é o ano de 2005 (68,9%) da amostra e o ano com maior concentração é o ano de 2001 (97,5%). Já para o *cluster* mercado no que se refere à concentração amostral na maioria dos anos também dois *clusters* concentram um percentual predominante, sendo o ano de 2005 (100%) com a maior concentração e 2003 com a menor concentração (50,67%). Quanto à distribuição setorial, novamente os *clusters* são bastante heterogêneos. Cabe destacar apenas que, nessa classificação, o setor “indústria manufatureira” ganha muita representatividade dentro de cada *cluster*. Por exemplo, no ano 2002 o setor “indústria manufatureira” apresenta no *cluster* índices percentuais de 42,3% no *cluster* 1, apresenta 35,8% no *cluster* 3, e por fim apresenta uma concentração de 66% para o *cluster* 2, sendo que os *clusters* 1,3 e 2 na ordem são os de maior representatividade de casos para esse ano.

4.6.2 Análise dos *clusters* índices e mercados formados com o mesmo número de setores das classificações BOVESPA (10) e NAICS (20)

Esta análise é feita considerando-se o número de *clusters* igual a 10, que corresponde ao número de setores da classificação BOVESPA. Considera-se nessa análise somente a classificação BOVESPA, e para melhor visualização dos resultados os *clusters* índices e mercados são mostrados em tabelas separadas. Salienta-se que a classificação NAICS também será avaliada tomando-se para tal o número de *clusters* igual a 20 que corresponde ao mesmo número de setores dessa classificação. As Tabelas 19 e 20 apresentam os resultados das composições dos 10 *clusters*, critério BOVESPA, para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2000.

Tabela 19 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2000

CLASSIFICAÇÃO BOVESPA – ANO 2000										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	8,0	23,5	11,1	0,0	0,0	8,3	0,0	0,0	0,0	0,0
construção e transporte	8,0	13,7	11,1	33,3	0,0	10,2	0,0	0,0	0,0	0,0
consumo cíclico	20,0	27,5	15,9	0,0	33,3	14,8	0,0	33,3	0,0	0,0
consumo não cíclico	6,0	11,8	4,8	0,0	0,0	8,3	0,0	0,0	0,0	0,0
financeiro e outros	20,0	3,9	3,2	0,0	50,0	25,9	100,0	0,0	100,0	0,0
materiais básicos	22,0	17,6	9,5	0,0	0,0	18,5	0,0	33,3	0,0	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	0,0	1,9	0,0	0,0	0,0	0,0
tecnologia da informação	2,0	2,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	6,0	0,0	12,7	0,0	16,7	4,6	0,0	0,0	0,0	100,0
utilidade pública	8,0	0,0	31,7	66,7	0,0	7,4	0,0	33,3	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	50	51	63	3	6	108	1	3	1	1

Tabela 20 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2000

CLASSIFICAÇÃO BOVESPA – ANO 2000										
SETORES	Cluster Mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	7,9	13,1	13,6	5,9	14,7	0,0	0,0	14,3	0,0	0,0
construção e transporte	18,4	8,2	7,4	0,0	14,7	0,0	0,0	0,0	0,0	0,0
consumo cíclico	10,5	21,3	19,8	0,0	29,4	0,0	0,0	57,1	100,0	33,3
consumo não cíclico	9,2	3,3	8,6	5,9	11,8	0,0	0,0	0,0	0,0	0,0
financeiro e outros	22,4	14,8	17,3	11,8	11,8	0,0	0,0	14,3	0,0	0,0
materiais básicos	10,5	19,7	19,8	29,4	11,8	16,7	0,0	14,3	0,0	0,0
petróleo, gás e biocomb.	0,0	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	33,3
tecnologia da informação	0,0	1,6	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	5,3	1,6	3,7	23,5	2,9	66,7	0,0	0,0	0,0	33,3
utilidade pública	15,8	14,8	8,6	23,5	2,9	16,7	100,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	76	61	81	17	34	6	1	7	1	3

Segundo este critério de decisão do número de *clusters*, dos 10 *clusters* sugeridos observa-se pelas tabelas 19 e 20 que tanto para o *cluster* índice como para o *cluster* mercado as ações expressivamente concentram-se em poucos *clusters*, 4 e 3 *clusters* respectivamente. Mais precisamente, considerando o *cluster* índice observa-se que nos *clusters* com as maiores concentrações (6, 3, 2 e 1) as ações estão distribuídas em diversos setores e com pequenas concentrações, ficando em torno de 20% nos setores de maior concentração. Para o *cluster* mercado observa-se que para os *clusters* de maiores concentrações (3, 1 e 2), as ações estão distribuídas em diversos setores e com pequenas concentrações, tal como no *cluster* índice. Dessa forma, do ponto de vista da composição do *cluster*, estes resultados *sugerem* uma não homogeneidade dos *clusters*.

Por outro lado, analisando-se as distribuições dos setores nos *clusters*, observa-se que para o *cluster* índice a maioria dos setores apresenta-se distribuídos em torno de 5 *clusters*, onde a distribuição mais heterogênea nos *clusters* ocorre para o setor “financeiro e outros”. Para o *cluster* mercado observa-se também que a maioria dos setores encontra-se distribuídos em torno de 5 *clusters*, ocorrendo a decomposição mais heterogênea no setor de “consumo cíclico”. Assim, esses resultados *sugerem* uma não homogeneidade setorial tanto para o *cluster* índice como para o *cluster* mercado.

As Tabelas 21 e 22 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2001.

Tabela 21 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2001

CLASSIFICAÇÃO BOVESPA – ANO 2001										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	7,4	16,9	11,1	15,0	0,0	0,0	0,0	0,0	50,0	0,0
construção e transporte	10,6	14,1	8,9	10,0	0,0	16,7	0,0	0,0	0,0	0,0
consumo cíclico	13,8	28,2	16,7	15,0	0,0	16,7	0,0	0,0	50,0	50,0
consumo não cíclico	8,5	12,7	3,3	5,0	0,0	0,0	0,0	0,0	0,0	0,0
financeiro e outros	27,7	5,6	10,0	20,0	100,0	0,0	0,0	100,0	0,0	0,0
materiais básicos	20,2	15,5	12,2	20,0	0,0	33,3	0,0	0,0	0,0	0,0
petróleo, gás e biocomb.	0,0	0,0	2,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	1,1	1,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	1,1	1,4	13,3	10,0	0,0	0,0	100,0	0,0	0,0	50,0
utilidade pública	9,6	4,2	22,2	5,0	0,0	33,3	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	94	71	90	20	2	6	1	2	2	2

Tabela 22 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado ano 2001

CLASSIFICAÇÃO BOVESPA – ANO 2001										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	7,1	8,3	10,3	36,4	6,7	12,5	0,0	0,0	0,0	0,0
construção e transporte	15,7	9,4	15,4	9,1	0,0	8,3	0,0	0,0	0,0	0,0
consumo cíclico	15,7	26,0	10,3	6,1	0,0	41,7	12,5	0,0	50,0	0,0
consumo não cíclico	11,4	5,2	7,7	12,1	6,7	0,0	0,0	0,0	0,0	0,0
financeiro e outros	20,0	21,9	12,8	12,1	13,3	0,0	12,5	0,0	0,0	0,0
materiais básicos	11,4	14,6	30,8	9,1	26,7	16,7	12,5	0,0	50,0	0,0
petróleo, gás e biocomb.	0,0	1,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	50,0
tecnologia da informação	0,0	0,0	0,0	0,0	0,0	8,3	0,0	0,0	0,0	0,0
Telecomunicações	4,3	4,2	5,1	0,0	26,7	0,0	50,0	0,0	0,0	50,0
utilidade pública	14,3	9,4	7,7	15,2	20,0	12,5	12,5	100,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	70	96	39	33	15	24	8	1	2	2

Observa-se pelas Tabelas 21 e 22, respectivamente, *cluster* índice e *cluster* mercado apresentam poucos *clusters* com expressivas concentrações. O *cluster* índice concentra-se praticamente nos *clusters* 1, 3 e 2, ao passo que o *cluster* mercado apresenta-se concentrado basicamente nos *clusters* 2 e 1. Para o *cluster* mercado, o *cluster* 2 é o de maior concentração com 33,10% do total de casos, a composição setorial desse *cluster* apresenta os setores “consumo cíclico” (26%), “financeiros e outros” (21,9%) e “materiais básicos” (14,6%) como os setores de maiores concentrações de um modo geral, para o *cluster* mercado a maioria dos *clusters* apresentam pequenas concentrações em vários setores. Para o *cluster* índice, os *clusters* apresentam pequenas concentrações de ações em vários setores, o que sugere

novamente uma não homogeneidade dos *clusters*. Do ponto de vista da decomposição dos setores nos *clusters*, ressalta-se que a maioria dos setores apresenta decomposição em torno de 6 *clusters* tanto para o *cluster* índice como para o *cluster* mercado. Ressalta-se que os setores com menores decomposições apresentam um pequeno número de ações como, por exemplo, o setor de “petróleo, gás e bicomcombustíveis” que apresenta um percentual de distribuição no *cluster* 2 de 2,2% equivalendo a quase duas ações do total de ações nesse *cluster*.

As Tabelas 23 e 24 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2002.

Tabela 23 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2002

CLASSIFICAÇÃO BOVESPA – ANO 2002										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	9,1	18,0	10,3	7,5	12,5	0,0	11,1	0,0	100,0	50,0
construção e transporte	10,0	16,0	17,2	6,0	12,5	0,0	0,0	0,0	0,0	50,0
consumo cíclico	15,5	26,0	24,1	10,4	20,8	100,0	44,4	0,0	0,0	0,0
consumo não cíclico	5,5	14,0	0,0	10,4	0,0	0,0	22,2	0,0	0,0	0,0
financeiro e outros	26,4	4,0	20,7	9,0	4,2	0,0	22,2	100,0	0,0	0,0
materiais básicos	15,5	16,0	17,2	20,9	16,7	0,0	0,0	0,0	0,0	0,0
petróleo, gás e biocomb.	0,0	2,0	0,0	1,5	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	1,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	7,3	0,0	3,4	13,4	0,0	0,0	0,0	0,0	0,0	0,0
utilidade pública	9,1	4,0	6,9	20,9	33,3	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
TOTAL CASOS	110	50	29	67	24	1	9	1	1	2

Tabela 24 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2002

CLASSIFICAÇÃO BOVESPA – ANO 2002										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	8,9	12,5	8,8	0,0	17,1	0,0	16,0	33,3	0,0	0,0
construção e transporte	13,9	13,6	2,9	0,0	7,3	0,0	20,0	0,0	0,0	0,0
consumo cíclico	13,9	22,7	20,6	0,0	24,4	5,9	20,0	0,0	0,0	0,0
consumo não cíclico	11,4	8,0	2,9	0,0	2,4	11,8	4,0	33,3	0,0	0,0
financeiro e outros	20,3	15,9	26,5	0,0	9,8	5,9	4,0	33,3	25,0	0,0
materiais básicos	10,1	15,9	11,8	0,0	17,1	35,3	36,0	0,0	0,0	0,0
petróleo, gás e biocomb.	0,0	1,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	50,0
tecnologia da informação	0,0	1,1	0,0	0,0	2,4	0,0	0,0	0,0	0,0	0,0
Telecomunicações	5,1	2,3	2,9	0,0	7,3	29,4	0,0	0,0	50,0	50,0
utilidade pública	16,5	6,8	23,5	100,0	12,2	11,8	0,0	0,0	25,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	79	88	34	1	41	17	25	3	4	2

Os resultados exibidos nas Tabelas 23 e 24 mostram que os *clusters* índice e mercado apresentam 2 *clusters* como os de maiores concentrações, sendo os *clusters* 1 e 4 para o *cluster* índice e os *clusters* 2 e 1 para o *cluster* mercado. Para o *cluster* índice a composição setorial dos *clusters* mostra pequenas concentrações em vários *clusters*, o mesmo ocorrendo para o *cluster* mercado, o que sugere uma não homogeneidade dos *clusters*. Em relação à decomposição setorial nos *clusters*, tanto o *cluster* índice como o *cluster* mercado as ações dos setores encontram-se distribuídas em vários *clusters* o que sugere uma não homogeneidade setorial. As Tabelas 25 e 26 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2003.

Tabela 25 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2003

CLASSIFICAÇÃO BOVESPA – ANO 2003										
SETORES	Cluster Índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	7,0	22,7	7,4	8,7	0,0	0,0	0,0	100,0	25,0	0,0
construção e transporte	9,9	9,1	11,5	8,7	20,0	0,0	0,0	0,0	50,0	0,0
consumo cíclico	8,5	28,8	20,5	8,7	0,0	100,0	0,0	0,0	25,0	0,0
consumo não cíclico	5,6	9,1	7,4	13,0	0,0	0,0	0,0	0,0	0,0	0,0
financeiro e outros	7,0	3,0	27,9	13,0	20,0	0,0	100,0	0,0	0,0	50,0
materiais básicos	26,8	19,7	8,2	21,7	0,0	0,0	0,0	0,0	0,0	50,0
petróleo, gás e biocomb.	1,4	1,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,0	0,0	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	7,0	1,5	8,2	4,3	20,0	0,0	0,0	0,0	0,0	0,0
utilidade pública	26,8	4,5	7,4	21,7	40,0	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	71	66	122	23	5	1	1	1	4	2

Tabela 26 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2003

CLASSIFICAÇÃO BOVESPA – ANO 2003										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	11,1	30,0	16,9	0,0	9,1	8,6	7,7	16,7	0,0	0,0
Construção e transporte	3,7	20,0	8,5	0,0	12,1	17,2	12,3	8,3	0,0	0,0
consumo cíclico	14,8	30,0	11,9	0,0	15,2	15,5	21,5	45,8	6,3	0,0
Consumo não cíclico	7,4	10,0	6,8	0,0	12,1	12,1	4,6	0,0	6,3	0,0
financeiro e outros	18,5	0,0	5,1	0,0	21,2	24,1	21,5	4,2	18,8	0,0
materiais básicos	11,1	0,0	39,0	0,0	6,1	6,9	15,4	8,3	25,0	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	0,0	0,0	1,5	0,0	0,0	33,3
tecnologia da informação	0,0	0,0	1,7	0,0	3,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	18,5	0,0	0,0	0,0	3,0	1,7	3,1	12,5	25,0	66,7
utilidade pública	14,8	10,0	10,2	100,0	18,2	13,8	12,3	4,2	18,8	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	27	10	59	1	33	58	65	24	16	3

Observa-se que para o *cluster* índice, o *cluster* 3 é aquele que apresenta a concentração mais expressiva, com 41,2 % do total. Quanto à composição setorial desse *cluster*, este apresenta vários setores com pequenas concentrações sendo que “materiais básicos” e “utilidade pública”, ambos com 26,8%, são os que apresentam as maiores concentrações, os demais apresentam percentuais de concentrações abaixo de 10%. Isso também se verifica para os demais *clusters*, ou seja, pequenas concentrações em vários setores. O *cluster* mercado apresenta o *cluster* 7 como o de maior concentração com 22% do total de casos. Os *clusters* de maior expressividade apresentam uma composição setorial com pequenas concentrações em vários setores. Por outro lado, a decomposição dos setores nos *clusters* fica em torno de 5 *clusters* tanto para o *cluster* índice como para o *cluster* mercado.

As Tabelas 27 e 28 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2004.

Tabela 27 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2004

CLASSIFICAÇÃO BOVESPA – ANO 2004										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	8,6	18,3	3,8	14,3	8,3	0,0	0,0	0,0	100,0	0,0
construção e transporte	7,5	11,7	19,2	12,9	0,0	0,0	0,0	0,0	0,0	0,0
consumo cíclico	10,8	30,0	23,1	12,9	8,3	33,3	100,0	33,3	0,0	0,0
consumo não cíclico	6,5	11,7	7,7	8,6	8,3	0,0	0,0	0,0	0,0	100,0
financeiro e outros	9,7	6,7	13,5	27,1	33,3	66,7	0,0	33,3	0,0	0,0
materiais básicos	19,4	13,3	17,3	11,4	16,7	0,0	0,0	33,3	0,0	0,0
petróleo, gás e biocomb.	1,1	1,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,0	1,7	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	9,7	3,3	3,8	5,7	8,3	0,0	0,0	0,0	0,0	0,0
utilidade pública	26,9	1,7	9,6	7,1	16,7	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	93	60	52	70	12	6	1	3	1	1

Tabela 28 - Classificação BOVESPA, com 10 clusters para cluster mercado ano 2004

CLASSIFICAÇÃO BOVESPA – ANO 2004										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	6,5	13,9	47,4	6,8	0,0	7,8	3,6	0,0	33,3	0,0
construção e transporte	3,2	10,1	5,3	12,2	0,0	19,6	10,7	0,0	33,3	0,0
consumo cíclico	9,7	27,8	10,5	21,6	0,0	15,7	10,7	0,0	0,0	0,0
consumo não cíclico	9,7	10,1	5,3	6,8	0,0	9,8	3,6	9,1	33,3	0,0
financeiro e outros	12,9	19,0	0,0	16,2	0,0	21,6	14,3	18,2	0,0	0,0
materiais básicos	16,1	11,4	15,8	12,2	0,0	5,9	46,4	36,4	0,0	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	1,4	0,0	0,0	0,0	0,0	0,0	50,0
tecnologia da informação	0,0	1,3	0,0	0,0	0,0	0,0	3,6	0,0	0,0	0,0
Telecomunicações	22,6	0,0	0,0	8,1	0,0	2,0	0,0	27,3	0,0	50,0
utilidade pública	19,4	6,3	15,8	14,9	100,0	17,6	7,1	9,1	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	31	79	19	74	1	51	28	11	3	2

Para o *cluster* índice, o *cluster* 1 é o que apresenta maior concentração. A composição setorial desse *cluster* e para os demais apresentam pequenas concentrações e vários setores. Quanto à distribuição dos setores nos *clusters* ela fica em torno de 4 *clusters*, entretanto com pequenos percentuais. Para o *cluster* mercado, mesmo os *clusters* de maiores concentrações, apresentam concentrações pouco expressiva (em torno de 20%). A composição setorial desses *clusters* também demonstra uma não homogeneidade dos *clusters*.

As Tabelas 29 e 30 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2005.

Tabela 29 – Classificação BOVESPA, com 10 clusters para o cluster índice no ano 2005

CLASSIFICAÇÃO BOVESPA – ANO 2005										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	7,1	21,2	14,3	11,3	5,6	0,0	0,0	0,0	0,0	100,0
Construção e transporte	10,6	16,7	14,3	6,5	13,9	0,0	7,7	0,0	0,0	0,0
consumo cíclico	13,3	27,3	0,0	12,9	22,2	33,3	30,8	66,7	0,0	0,0
consumo não cíclico	8,0	9,1	0,0	11,3	8,3	0,0	7,7	0,0	0,0	0,0
financeiro e outros	4,4	7,6	14,3	40,3	19,4	33,3	23,1	0,0	100,0	0,0
materiais básicos	12,4	10,6	57,1	11,3	27,8	0,0	15,4	33,3	0,0	0,0
petróleo, gás e biocomb.	0,9	1,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,0	1,5	0,0	1,6	0,0	0,0	7,7	0,0	0,0	0,0
telecomunicações	12,4	0,0	0,0	1,6	2,8	33,3	7,7	0,0	0,0	0,0
utilidade pública	31,0	4,5	0,0	3,2	0,0	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	113	66	7	62	36	3	13	3	2	1

Tabela 30 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2005

CLASSIFICAÇÃO BOVESPA – ANO 2005										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	4,3	8,3	16,7	12,7	0,0	9,5	6,5	0,0	0,0	0,0
construção e transporte	0,0	25,0	9,7	9,8	0,0	4,8	21,0	0,0	0,0	0,0
consumo cíclico	17,4	16,7	15,3	24,5	0,0	19,0	16,1	0,0	0,0	0,0
consumo não cíclico	8,7	8,3	6,9	7,8	0,0	9,5	11,3	0,0	0,0	100,0
financeiro e outros	17,4	16,7	22,2	10,8	0,0	0,0	22,6	22,2	0,0	0,0
materiais básicos	13,0	8,3	18,1	20,6	0,0	4,8	1,6	44,4	33,3	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	1,0	0,0	0,0	0,0	0,0	33,3	0,0
tecnologia da informação	0,0	0,0	2,8	1,0	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	21,7	0,0	2,8	3,9	0,0	4,8	4,8	22,2	33,3	0,0
utilidade pública	17,4	16,7	5,6	7,8	100,0	47,6	16,1	11,1	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	23	12	72	102	1	21	62	9	3	1

Para o *cluster* índice, os *clusters* 1, 2, 4 e 5 são os que apresentam as maiores concentrações para o *cluster* índice, sendo os respectivos percentuais de 36,9 %, 21,6 %, 20,3 % e 11,8 %, e juntos computam o total de 90,6%. Na composição setorial do *cluster* 1, o setor de “utilidade pública” é o que apresenta uma relevante concentração de 31%. A composição setorial desses *clusters* apresenta pequenas concentrações em vários setores, o que sugere uma não homogeneidade desses *clusters*. Para o *cluster* mercado, o *cluster* 4 é o de maior concentração, com 33,3% do total de casos. A composição setorial desse *cluster* apresenta os setores “materiais básicos” (20,6 %), “consumo cíclico” (24,5 %), “bens industriais” (12,7 %) e “financeiros e outros” (10,8 %), como os setores de maiores concentrações. Os demais *clusters* também apresentam composições setoriais variadas e de baixa concentração.

As Tabelas 31 e 32 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2006.

Tabela 31 – Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2006

CLASSIFICAÇÃO BOVESPA – ANO 2006										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	5,0	11,5	22,6	0,0	9,7	6,7	100,0	0,0	0,0	0,0
construção e transporte	9,0	14,2	16,1	10,0	3,2	6,7	0,0	0,0	0,0	0,0
consumo cíclico	18,0	14,2	24,2	20,0	3,2	26,7	0,0	0,0	0,0	0,0
consumo não cíclico	10,0	12,4	6,5	20,0	3,2	0,0	0,0	0,0	100,0	0,0
financeiro e outros	7,0	27,4	3,2	10,0	22,6	33,3	0,0	100,0	0,0	0,0
materiais básicos	18,0	11,5	12,9	0,0	16,1	13,3	0,0	0,0	0,0	0,0
petróleo, gás e biocomb.	2,0	0,0	1,6	0,0	0,0	0,0	0,0	0,0	0,0	100,0
tecnologia da informação	0,0	4,4	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telecomunicações	4,0	1,8	6,5	20,0	9,7	13,3	0,0	0,0	0,0	0,0
utilidade pública	27,0	2,7	4,8	20,0	32,3	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	100	113	62	10	31	15	1	1	1	1

Tabela 32 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2006

CLASSIFICAÇÃO BOVESPA – ANO 2006										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	20,0	14,0	14,3	0,0	7,3	11,1	0,0	15,4	100,0	0,0
construção e transporte	0,0	10,9	28,6	0,0	13,4	3,7	14,3	19,2	0,0	0,0
consumo cíclico	40,0	14,7	0,0	0,0	12,2	33,3	3,6	23,1	0,0	0,0
consumo não cíclico	0,0	9,3	28,6	0,0	12,2	7,4	14,3	0,0	0,0	0,0
financeiro e outros	0,0	13,2	14,3	0,0	22,0	11,1	21,4	23,1	0,0	0,0
materiais básicos	0,0	14,7	14,3	0,0	7,3	25,9	17,9	0,0	0,0	50,0
petróleo, gás e biocomb.	0,0	1,6	0,0	0,0	0,0	1,9	0,0	0,0	0,0	50,0
tecnologia da informação	0,0	3,1	0,0	0,0	1,2	1,9	0,0	0,0	0,0	0,0
Telecomunicações	20,0	2,3	0,0	0,0	8,5	1,9	17,9	0,0	0,0	0,0
utilidade pública	20,0	16,3	0,0	100,0	15,9	1,9	10,7	19,2	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	5	129	7	1	82	54	28	26	1	2

Para os resultados apresentados na Tabela 31, observa-se que os *clusters* 2 e 1 são os *clusters* que apresentam maior concentração de casos, respectivamente com 33,7% e 29,9%. Na composição setorial do *cluster* 1, os setores de “utilidade pública” (27%) e “consumo cíclico” e “materiais básicos” ambos com o mesmo percentual (18%) são os setores de maior relevância, do mesmo modo, os outros *clusters* pertencentes ao *cluster* índice apresentam pequenas concentrações em um variado número de setores. Para o *cluster* mercado, os *clusters* 2 e 5 são os que apresentam as concentrações mais expressivas. O *cluster* 2 é o de

maior concentração com 38,5% do total de casos. A composição setorial desse *cluster*, como os demais, apresentam baixas concentrações em vários setores.

As Tabelas 33 e 34 apresentam os resultados das composições dos 10 *clusters* para o *cluster* índice e o *cluster* mercado, segundo a classificação BOVESPA para o ano de 2007.

Tabela 33 - Classificação BOVESPA, com 10 *clusters* para o *cluster* índice no ano 2007

CLASSIFICAÇÃO BOVESPA – ANO 2007										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	15,0	17,4	4,8	0,0	0,0	0,0	0,0	20,0	25,0	0,0
construção e transporte	11,3	15,9	15,7	12,9	0,0	13,3	39,3	0,0	25,0	0,0
consumo cíclico	15,8	29,0	8,4	6,5	9,5	40,0	7,1	20,0	0,0	50,0
consumo não cíclico	12,8	11,6	9,6	3,2	0,0	13,3	10,7	0,0	0,0	0,0
financeiro e outros	22,6	2,9	8,4	58,1	4,8	20,0	25,0	40,0	25,0	0,0
materiais básicos	12,0	11,6	24,1	6,5	0,0	13,3	3,6	0,0	0,0	50,0
petróleo, gás e biocomb.	1,5	1,4	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,8	4,3	0,0	0,0	0,0	0,0	10,7	0,0	0,0	0,0
telecomunicações	5,3	0,0	6,0	6,5	4,8	0,0	3,6	20,0	0,0	0,0
utilidade pública	3,0	5,8	21,7	6,5	81,0	0,0	0,0	0,0	25,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	133	69	83	31	21	15	28	5	4	2

Tabela 34 - Classificação BOVESPA, com 10 *clusters* para o *cluster* mercado no ano 2007

CLASSIFICAÇÃO BOVESPA – ANO 2007										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	8,3	9,8	0,0	2,9	20,5	14,3	0,0	11,1	25,0	0,0
construção e transporte	18,6	9,8	0,0	17,6	12,8	28,6	0,0	22,2	0,0	0,0
consumo cíclico	14,7	15,9	0,0	8,8	23,1	14,3	0,0	33,3	50,0	0,0
consumo não cíclico	9,6	9,8	0,0	17,6	12,8	0,0	0,0	0,0	0,0	0,0
financeiro e outros	26,3	13,6	0,0	5,9	7,7	0,0	57,1	33,3	0,0	0,0
materiais básicos	10,9	12,1	0,0	14,7	17,9	0,0	42,9	0,0	25,0	50,0
petróleo, gás e biocomb.	0,6	1,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	50,0
tecnologia da informação	1,3	3,0	0,0	0,0	2,6	0,0	0,0	0,0	0,0	0,0
telecomunicações	1,9	5,3	0,0	14,7	2,6	14,3	0,0	0,0	0,0	0,0
utilidade pública	7,7	18,9	100,0	17,6	0,0	28,6	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	156	132	1	34	39	7	7	9	4	2

Segundo os resultados exibidos na Tabela 33, para o *cluster* índice, observa-se que os *clusters* 1 e 2 são os que apresentam as concentrações mais expressivas, juntos computam o total de 73,7%. O *cluster* 1 é o de maior concentração com 39,9% do total de casos. A composição setorial desse *cluster* apresenta o setor “financeiro e outros” (26,3%) como o de maior relevância, as demais ações apresentam-se distribuídas em vários setores com pequenas concentrações. O *cluster* 2 com percentual de concentração de 33,8% apresenta na sua composição os setores “utilidade pública” (18,9%), “consumo cíclico” (15,9%), “financeiros e outros” (13,6%), “materiais básicos” (12,1%) como os setores de maiores concentrações. A distribuição setorial no *cluster* índice aponta também para uma não homogeneidade dos *clusters*.

Para os resultados do *cluster* mercado, exibidos na Tabela 34, a maior concentração dos casos encontra-se no *cluster* 1. Quanto à composição setorial do *cluster* 1 os setores de maiores concentrações são os setores “financeiros e outros” (22,6%), “consumo cíclico” (15,8%), “consumo não cíclico” (12,8%), “materiais básicos” (12%) e “construção e transporte” (11,3%), os demais *clusters* também apresentam uma distribuição das ações em vários setores com pouca concentração.

Tabela 35 – Resumo da classificação setorial BOVESPA, para o *cluster* índice para o período 2000 - 2007

CLUSTER INDICE – CRITÉRIO CLASSIFICAÇÃO BOVESPA						
ANO	Nº CLUSTERS	QUANTIDADE DE SETORES			FREQUÊNCIA SETORIAL	
		MAX	MIN	MÉDIA	MÁXIMA (%)	SETOR
2000	10	9	1	4,4	25,9	financeiro e outros
2001	10	9	1	4,6	28,2	consumo cíclico
2002	10	9	1	4,8	26,4	financeiro e outros
2003	10	9	1	4,7	28,8	consumo cíclico
2004	10	10	1	5,1	30,0	consumo cíclico
2005	10	10	6	8,8	31,0	utilidade pública
2006	10	10	1	5,2	24,2	consumo cíclico
2007	10	10	2	6,1	22,6	financeiro e outros

Dos resultados exibidos na Tabela 35 observa-se pelo valor mínimo que para todos os anos, foi obtido pelo menos um *cluster* com empresas de um único setor. Tais *clusters*, geralmente, são aqueles formados por apenas uma ou duas empresas. O número máximo de setores por *cluster* foi 9 para os anos 2000 a 2003 e 10 para os anos de 2004 a 2007. Em

média há 5,5 setores por *clusters* e a frequência máxima em um único setor é em torno de 27,1%.

Tabela 36 – Resumo da classificação setorial BOVESPA, para o *cluster* mercado para o período 2000 – 2007

CLUSTER MERCADO - CRITÉRIO CLASSIFICAÇÃO BOVESPA						
ANO	Nº CLUSTERS	QUANTIDADE DE SETORES			FREQ SETORIAL MÁXIMA (%)	SETOR
		MAX	MIN	MÉDIA		
2000	10	10	1	5,3	21,3	consumo cíclico
2001	10	9	1	5,4	26,0	consumo cíclico
2002	10	10	1	5,6	22,7	consumo cíclico
2003	10	9	1	6,3	21,5	Financeiro, cons. Cíclico
2004	10	9	1	5,8	21,6	consumo cíclico
2005	10	10	1	5,7	24,5	consumo cíclico
2006	10	10	1	5,4	33,3	consumo cíclico
2007	10	10	1	5,3	26,3	financeiros e outros

Dos resultados exibidos na Tabela 36 observa-se pelo valor mínimo que para todos os anos, foi obtido pelo menos um *cluster* com empresas de um único setor. Em geral, tais *clusters* são aqueles formados por apenas uma ou duas empresas. O número máximo de setores por *cluster* foi 9 para os anos 2001, 2003 e 2004, sendo 10 para os anos 2000, 2002, 2005, 2006 e 2007. Em média há 5,6 setores por *clusters* e a frequência máxima em um único setor gira em torno de 24,6%.

Portanto, como a classificação BOVESPA possui 10 setores e como foram estimados 10 *clusters*, caso os setores fossem heterogêneos entre si esperar-se-ia que cada *cluster* apresentasse empresas de um setor. Logo, os resultados demonstram uma alta heterogeneidade setorial em cada *cluster*, o que aponta para a ineficiência da análise setorial nesses casos.

Para a classificação NAICS, foram estimados os *clusters* índices e mercados formados com 20 *clusters*, o mesmo número de setores da classificação NAICS (20). No entanto, observou-se que tanto para o *cluster* índice como o *cluster* mercado diversos *clusters* foi construído com poucos casos. O Quadro 09 apresenta um levantamento do número de *clusters* com 1, 2 ou 3 casos em cada período.

LEVANTAMENTO DO NÚMERO DE <i>CLUSTERS</i>			
CLASSIFICAÇÃO NAICS - CRITÉRIO NAICS (20 <i>CLUSTERS</i>)			
NÚMERO DE <i>CLUSTERS</i>			
	ANO	1, 2 ou 3 CASOS	%
<i>CLUSTER</i> ÍNDICE	2000	11	55
	2001	12	60
	2002	11	55
	2003	9	45
	2004	8	40
	2005	7	35
	2006	9	45
	2007	10	50
<i>CLUSTER</i> MERCADO	2000	6	30
	2001	6	30
	2002	10	50
	2003	4	20
	2004	6	30
	2005	4	20
	2006	9	45
	2007	5	25

Quadro 09 - Levantamento do número de *clusters* quando se impõe o número de clusters iguais a 20 o mesmo número de subdivisões da classificação NAICS, no período de 2000 – 2007

Observa-se que para o *cluster* índice e para o *cluster* mercado em torno de 55% e 35%, respectivamente, são os números de *clusters* formados por apenas uma, duas ou três ações. Portanto, conclui-se que a utilização da análise de cluster com a regra de formação de 20 clusters é inadequada para os objetivos dessa pesquisa, pois para grande parte dos *clusters* não há aglomeração de empresas.

4.7 Análise dos *clusters* índice e mercado com a inclusão das variáveis de tamanho, segundo a classificação BOVESPA

Para avaliação do efeito das variáveis de tamanho na formação dos *clusters*, estimam-se novamente os clusters índice e mercado com a classificação BOVESPA. A tabela 37 apresenta os resultados para o *cluster* índice e a Tabela 38 para o *cluster* mercado, ambas para o ano de 2001.

A Tabela 37 apresenta os resultados para o *cluster* índice e mostra os percentuais de distribuições nos 10 *clusters*, segundo a classificação BOVESPA, para o ano 2001.

Tabela 37 – Classificação BOVESPA, com 10 clusters para o cluster índice, no ano 2001

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2001										
SETORES	Cluster Índice									
	1	2	3	4	5	6	7	8	9	10
Bens Industriais	1,5	16,4	10,8	18,8	28,6	0,0	0,0	0,0	0,0	0,0
Construção e Transporte	4,4	13,7	9,6	18,8	28,6	0,0	0,0	0,0	0,0	0,0
Consumo Cíclico	2,9	28,8	16,9	27,1	28,6	0,0	25,0	0,0	0,0	50,0
Consumo Não Cíclico	7,4	11,0	7,2	4,2	28,6	0,0	0,0	0,0	0,0	0,0
Financeiro e Outros	11,8	6,8	24,1	18,8	28,6	100,0	0,0	0,0	100,0	0,0
Materiais Básicos	20,6	17,8	18,1	8,3	28,6	0,0	25,0	0,0	0,0	0,0
Petróleo, Gás e Biocomb.	1,5	0,0	1,2	0,0	28,6	0,0	0,0	0,0	0,0	0,0
Tecnologia da Informação	0,0	1,4	0,0	2,1	28,6	0,0	0,0	0,0	0,0	0,0
Telecomunicações	19,1	0,0	2,4	2,1	28,6	0,0	0,0	100,0	0,0	50,0
Utilidade Pública	30,9	4,1	9,6	0,0	28,6	0,0	50,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	68	73	83	48	7	2	4	1	2	2

Observa-se que os *clusters* estimados com as variáveis de tamanho apresentam resultados semelhantes aos obtidos sem a inclusão de tais variáveis. Na maioria dos casos os *clusters* apresentam uma grande variedade setorial e cada setor aparece em diversos *clusters*.

A Tabela 38 apresenta os resultados para o *cluster* mercado e mostra os percentuais de distribuições nos 10 *clusters*, segundo a classificação BOVESPA, para o ano 2001.

Tabela 38 – Classificação BOVESPA, com 10 clusters para o cluster mercado, no ano 2001

CLASSIFICAÇÃO BOVESPA – C/VARIAVEIS DE TAMANHO, ANO 2001										
SETORES	Cluster Mercado									
	1	2	3	4	5	6	7	8	9	10
Bens Industriais	5,4	17,2	14,5	5,0	18,2	26,7	0,0	9,1	0,0	0,0
Construção e Transporte	6,5	20,7	14,5	0,0	15,2	13,3	0,0	9,1	0,0	0,0
Consumo Cíclico	12,9	17,2	15,8	0,0	36,4	46,7	10,0	36,4	0,0	50,0
Consumo Não Cíclico	3,2	10,3	13,2	15,0	6,1	0,0	0,0	0,0	0,0	0,0
Financeiro e Outros	12,9	17,2	22,4	25,0	12,1	6,7	10,0	18,2	0,0	0,0
Materiais Básicos	18,3	13,8	17,1	30,0	12,1	6,7	10,0	0,0	0,0	50,0
Petróleo, Gás e Biocomb.	1,1	0,0	0,0	0,0	0,0	0,0	10,0	0,0	0,0	0,0
Tecnologia da Informação	1,1	0,0	0,0	0,0	0,0	0,0	0,0	9,1	0,0	0,0
Telecomunicações	8,6	0,0	2,6	10,0	0,0	0,0	50,0	9,1	0,0	0,0
Utilidade Pública	30,1	3,4	0,0	15,0	0,0	0,0	10,0	9,1	100,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	93	29	76	20	33	15	10	11	1	2

Pelos resultados exibidos na Tabela 38, do mesmo modo que ocorreu para o *cluster* índice, a inclusão das variáveis de tamanho não altera significativamente os resultados. Na maioria dos casos os *clusters* apresentam uma grande variedade setorial e cada setor aparece em diversos *clusters*.

Para os demais anos, os resultados assemelham-se aos apresentados para o ano de 2001. A inclusão das variáveis de tamanho não promove um aumento da homogeneidade setorial de cada *cluster* nem aumenta a aglomeração dos setores em cada *cluster*.

As tabelas dos *clusters* índice e mercado com as variáveis de tamanho para os demais anos encontram-se no Apêndice F.

5. ESTABILIDADE DOS *CLUSTERS*

Para avaliar se os *clusters* são estáveis ao longo do tempo aplicou-se o teste de Wilcoxon. Este teste foi aplicado para análise da estabilidade quando se opta pelo número de *clusters* sugeridos conforme a regra de parada e pela classificação BOVESPA. Como o objetivo é avaliar o comportamento do *cluster* ao longo do tempo, a composição do *cluster* em cada ano (2001 a 2007) foi comparada a composição no primeiro ano da série (2000).

Para os *clusters* estimados, segundo a regra de parada, o número de *clusters* a serem comparados referem-se a menor quantidade de *clusters* gerados pelo método no ano 2000, ou seja, 4 para o *cluster* índice e 2 para o *cluster* mercado.

Para a estimação com o critério BOVESPA, o número de *clusters* estimados é constante e igual a dez.. Portanto, para este critério foram comparados os dez *clusters* gerados em cada ano. A Tabela 39 apresenta os resultados do Teste de Wilcoxon para o *cluster* índice e *cluster* mercado, segundo o critério de regra de parada, para a classificação BOVESPA.

Tabela 39 – Valores de significância para o teste de Wilcoxon, com 4 *clusters* para o *cluster* índice e 2 *clusters* para o *cluster* mercado, com critério de regra de parada, para os anos do período de 2001-2007

TESTE RANK SINAL DE WILCOXON														
CLASSIFICAÇÃO BOVESPA – CRITÉRIO REGRA DE PARADA														
nº <i>clusters</i> /ano	<i>Cluster</i> índice							<i>Cluster</i> mercado						
	2001	2002	2003	2004	2005	2006	2007	2001	2002	2003	2004	2005	2006	2007
1	0,95	0,95	0,92	0,84	0,88	0,57	0,88	0,65	0,58	0,88	0,51	0,51	0,96	0,84
2	0,59	0,59	0,96	0,96	0,86	0,86	0,77	0,28	0,44	0,31	0,67	0,31	0,39	0,51
3	0,73	0,73	0,86	0,77	0,67	0,37	0,68							
4	0,67	0,67	0,67	0,58	0,44	0,92	0,73							

O Teste de Wilcoxon compara a composição setorial do *cluster* no ano 2000 à composição em cada um dos anos seguintes (2001 a 2007). Observa-se que todos os valores de significância são superiores a 0,05, indicando que não há alterações significativas na composição dos *clusters* ao longo do período estudado.

A Tabela 40 apresenta os resultados do Teste de Wilcoxon para o *cluster* índice e *cluster* mercado, segundo o critério da classificação BOVESPA.

Tabela 40 – Significância para o teste de Wilcoxon, com 10 clusters para os clusters índice e mercado, para os anos do período de 2001-2007

TESTE RANK SINAL DE WILCOXON														
CLASSIFICAÇÃO BOVESPA- CRITÉRIO BOVESPA														
nºclusters/ano	Cluster índice							Cluster mercado						
	2001	2002	2003	2004	2005	2006	2007	2001	2002	2003	2004	2005	2006	2007
1	0,95	0,95	0,92	0,84	0,88	0,57	0,88	0,67	0,78	1,00	1,00	1,00	1,00	0,28
2	0,59	0,59	0,96	0,96	0,86	0,86	0,77	0,96	0,65	0,96	0,96	0,96	0,86	0,88
3	0,73	0,73	0,86	0,77	0,67	0,37	0,68	0,77	0,68	0,77	0,37	0,95	0,59	0,11
4	0,67	0,67	0,67	0,58	0,44	0,92	0,73	0,89	0,34	0,34	0,77	0,80	0,34	0,94
5	1,00	1,00	0,89	0,87	0,93	0,78	0,72	0,89	0,95	0,95	0,16	0,16	0,95	0,86
6	0,86	0,86	0,11	0,48	0,68	1,00	0,37	0,46	0,46	0,40	0,40	0,61	0,33	0,92
7	1,00	1,00	1,00	1,00	0,23	1,00	0,24	0,49	0,24	0,11	0,16	0,16	0,24	1,00
8	0,71	0,71	0,71	1,00	0,66	0,71	0,92	0,49	0,68	1,00	0,61	0,75	0,75	0,89
9	1,00	1,00	0,71	1,00	1,00	1,00	0,71	1,00	0,71	0,34	0,71	0,71	1,00	1,00
10	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,66	1,00	0,71	1,00	1,00

Do mesmo modo que na tabela anterior, todos os valores de significância apresentados são superiores a 0,05. Portanto, os *clusters* estimados segundo o critério BOVESPA também apresentam composições semelhantes ao longo dos anos.

Independentemente do critério de estimação utilizado, regra de parada ou critério BOVESPA, conclui-se que a composição setorial de cada *cluster* não se alterou significativamente ao longo do período. Portanto conclui-se que a composição setorial dos *clusters* é estável no período analisado.

6. CONCLUSÕES

Esta dissertação avaliou a eficiência da classificação setorial. A utilização da classificação setorial nos estudos na área de Administração tem por base o pressuposto de que empresas de um mesmo setor são semelhantes. Portanto, se a classificação setorial é eficiente, espera-se que empresas de um mesmo setor formem um *cluster* e que empresas de setores diferentes sejam agrupadas em *clusters* díspares.

Inicialmente optou-se por uma avaliação univariada das diferenças entre os setores. Os resultados do Teste de Kruskal Wallis apresentaram valor de significância inferior a 0,05 para as duas classificações setoriais, indicando que existe diferença significativa entre os setores de atividade para as variáveis estudadas. Tais resultados mostram que, do ponto de vista univariado, o setor de atividade é um fator de classificação preponderante, uma vez que as medianas de cada setor diferem estatisticamente. Observou-se ainda que as diferenças setoriais não sejam fortemente afetadas pelo ano de estudo, já que as diferenças se conservam ao longo do período estudado.

Para a análise dos *clusters* considerou-se o *cluster* índice e o *cluster* mercado. O *cluster* índice foi estimado com variáveis predominantemente econômico-financeiras e o *cluster* mercado com variáveis relacionadas ao mercado de ações. Como critério de escolha do número de *clusters* utilizou-se o critério da regra de parada e o critério de definição do número de *clusters* segundo o número de setores em cada classificação. Todos os *clusters* também foram reestimados utilizando-se variáveis de tamanho.

De modo geral, apesar dos *clusters* índice e mercado serem construídos a partir de variáveis diferentes, não foi observado homogeneidade para nenhum desses conjuntos, quando classificados segundo as classificações BOVESPA e NAICS. Salienta-se também que, objetivamente, quando foram incluídas as variáveis de tamanho essa situação não apresentou mudanças consideráveis.

Quanto ao critério de escolha do número de *clusters* cabe salientar que o critério regra de parada tanto para as classificações BOVESPA e NAICS não foi eficiente, no sentido de um número de *clusters* adequado com vistas a obter uma maior homogeneidade setorial nos *clusters*. O critério BOVESPA, com um número de *clusters* iguais a 10, também apresentou uma alta heterogeneidade setorial nos *clusters*. E o critério NAICS, regra de formação com 20 *clusters*, mostrou-se inadequado para os objetivos dessa pesquisa, pois para a grande maioria dos *clusters* não apresentava aglomeração de empresas.

Após a formação dos *clusters* e a sua classificação em setores, buscou-se avaliar a estabilidade dos mesmos ao longo do período estudado. O teste de Wilcoxon apresentou resultados não significativos para todos os *clusters* e em todos os períodos comparados ao ano inicial de análise, independente da utilização do critério para a obtenção do número de *clusters*, regra de parada ou critério BOVESPA. Portanto, conclui-se que a composição setorial de cada *cluster* não se alterou significativamente ao longo do período, o que implica numa composição setorial dos *clusters* estável durante o período analisado.

Os resultados de um modo geral demonstraram uma alta heterogeneidade setorial nos *clusters*, seja pelo número de *clusters* sugerido pela regra de parada ou critério BOVESPA, e esta heterogeneidade é estável, uma vez que se manteve ao longo do período estudado. Portanto, os resultados apontam para a ineficiência da análise setorial.

O estudo do agrupamento de empresas (ações) a partir de dados econômicos e financeiros serviu para identificar a não existência de homogeneidade dos setores do ponto de vista dos indicadores quando sujeitos a uma classificação baseada unicamente no tipo de bem ou serviço produzido. Questões tradicionalmente analisadas sob o ponto de vista de setor constituído por empresas oriundas do mesmo tipo de bem ou serviço produzido, podem não ser sequer adequadas ao estudo do ambiente de tarefas da empresa. Nesse sentido, a análise de um setor definido nessas condições esclarece muito pouco sobre as ameaças e oportunidades existentes para uma dada empresa, subsidiando um arcabouço de natureza estratégica inadequado, seja para os investidores seja para as empresas.

Para tomadas de decisões mais precisas, os setores podem ser reestruturados de modo a incorporar, em cada agrupamento, empresas que possam ser razoavelmente semelhantes. A análise da indústria, onde o setor é previamente definido em função do ramo de atividade em que a empresa está inserida, deve ser utilizada com muita cautela. A visão da empresa baseada em recursos (RBV) ao considerar que toda empresa possui um *portfolio* de recursos físicos, financeiros, intangíveis (marca, imagem), organizacionais (cultura organizacional, sistemas administrativos) e recursos humanos que deve ser avaliado para que a empresa possa criar vantagens competitivas parece mais adequada, uma vez que se foca no *portfolio* de recursos e não necessariamente no setor de atividade para a tomada de decisão.

Do ponto de vista financeiro, os resultados desta pesquisa apontam para a necessidade de se rever o método tradicional de construção de índices-padrão a partir das empresas do mesmo setor. Na definição de estratégias de diversificação, a grande heterogeneidade setorial obtida vai ao encontro dos estudos que defendem que a diversificação por país é mais relevante do que a diversificação por setor. Do mesmo modo, é preciso rever ou aprimorar as

técnicas de formação do grupo de controle nos estudos em que são formadas amostras de controle, principalmente tendo como variável chave o setor de atividade.

Dados setoriais publicados e organizados por tipo de bem ou serviço devem ser utilizados com a devida prudência. A análise da indústria fica prejudicada num contexto onde o grupo de comparação, empresa do mesmo setor, pode ter poucas características em comum.

É importante salientar que, no cenário atual de economia globalizada, de reconfiguração das empresas globais na busca da eficiência coletiva, o papel do setor ainda é tradicionalmente importante. Entretanto, é preciso compreender esse processo de reestruturação do tecido industrial de um país, como o Brasil, sob o ponto de vista dinâmico. O processo de agrupamento por atividade é estático e pode não responder às expectativas de crescimento tanto das empresas como do país. É preciso que as empresas encontrem formas estratégicas mais adequadas à natureza das suas vantagens competitivas, à sua dimensão e à sua cultura organizacional.

Pesquisas futuras podem realizar estudos de agrupamentos setoriais utilizando-se de variáveis mais relacionadas à estratégia e performance empresarial. Além disso, devem considerar a possibilidade de utilização de classificações setoriais que não sejam baseadas no produto final, bem ou serviços prestados.

Como limitações do estudo destacam-se a classificação setorial e o período de análise. Quanto às classificações foram testadas somente duas, a BOVESPA e NAICS nível 1. Poderia ter sido incluída a classificação Económica, porém, devido ao número de ações da amostra, e considerando o ocorrido como a classificação NAICS é esperado que o mesmo se repita com a classificação Económica, uma vez que também é composta por 20 setores. A pesquisa tomou os balanços referentes ao exercício fiscal de 2000 à 2007, assim, os resultados da análise por meio de índices e, por extensão, de *cluster*, refletem o desempenho e a situação patrimonial das empresas brasileiras, na conjuntura econômica vigente no ano de exercício e nos precedentes.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AMIT, R.; SCHOEMAKER, P. J. H. Strategic assets and organizational rent. **Strategic Management Journal**, n. 14, p.33-46, 1993.

ANDRADE, M. M. **Introdução a Metodologia do Trabalho Científico**. 2ª ed. São Paulo: Atlas, 1997.

ANGELO, C. F.; SANVICENTE, A. Z. Agrupamento de empresas por semelhança: uma crítica à análise setorial convencional. **Revista de administração**, São Paulo, v. 25, p. 20-27, 1990.

ANSOFF, H. I. **Corporate strategy**. Harmondsworth: Penguin, 1965.

ARNOTT, R. D. Cluster Analysis and Stock Price Comovement. **Financial Analysts Journal**, v. 37, n. 6 (November/December):56-62, 1980.

ASSAF NETO, Alexandre. **Estrutura e análise de balanços**: um enfoque econômico - financeiro. Comércio e Serviços Indústrias Bancos Comerciais e Múltiplos. 7ª ed. São Paulo: Atlas, 2002.

_____. **Contribuição ao estudo da avaliação de empresas no Brasil**: uma aplicação prática. Ribeirão Preto. Tese (Livre Docência). Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2004.

BACA, S.P.; GARBE, B.L.; WEISS, R.A. The Rise of Sector Effects in Major Equity Markets. **Financial Analysts Journal**, v.56, p.34-40, 2000.

BARNEY, J. B. Firm resources and sustained competitive advantage. **Journal of Management**, n.17, p. 99-120, 1991.

BARNETT, V.; LEWIS, T. **Outliers in statistical data**. Ed. John Wiley&Sons, New York, 1994.

BEAVER, W.H. Financial Ratios and Predictions of Failure, Empirical Research in Accounting Selected Studies. **Suplemento do Journal of Accounting Research**, p.77-111, 1966.

BOILLAT, P.Y.; SKOWRONSKI, N.; TUCHSCHMID, N. Cluster Analysis: application to sector indices and empirical validation. **Financial Markets and Portfolio Management**, v.16, n.4, p. 467-486, 2002.

BOVESPA - Bolsa de Valores de São Paulo - **Folheto informativo: Classificação setorial das empresas listadas**. Disponível em: <<http://www.bovespa.com.br/pdf/ClassifSetorial.pdf>> Acesso em maio de 2008.

BERGMAN, E. M.; FESER, E. J. *Industrial and Regional Clusters: Concepts and Comparative Applications*. Virginia: West Virginia University - Regional Research Institute, 1998. Disponível em: <http://www.rri.wvu.edu/WebBook/Bergman-Feser/contents.htm>. Acesso em março 2007.

BUSSAB, W. de O.; MIAZAKI, E.S.; ANDRADE, D.F. Introdução à Análise de Agrupamentos. In: 9º Simpósio Nacional de Probabilidade e Estatística, São Paulo. Associação Brasileira de Estatística, p.105, 1990.

CAVAGLIA, S.; BRIGHTMAN C.; AKED M. The Increasing Importance of Industry Effects. **Financial Analysts Journal**, v. 56, n.5, p.41-54, 2000.

CAVES, R. E. **American Industry, Structure, Conduct and Performance**. Englewood Cliffs: Prentice Hall, 306 p., 1982.

CHANDLER, Alfred D. *Strategy and structure: chapters in the history of the industrial enterprise*. M.I.T. PRESS, 1962.

CLIFFORD, W. Another Method of Analyzing the Importance of a Company's Country versus Sector Characteristics. Available at SSRN: <http://ssrn.com/abstract=302850> or DOI: [10.2139/ssrn.302850](https://doi.org/10.2139/ssrn.302850), 2002. Acesso em junho de 2007.

CORRAR, L. J.; PAULO E.; FILHO, J. M. D. **Análise Multivariada**. Para os Cursos de Administração, Ciências Contábeis e Economia. 1ª ed. São Paulo: Atlas, 2007.

DALHSTEDT, Roy; SALMI, Timo; LUOMA, Martti; LAAKKONEN, Arto. (1994). On the Usefulness of Standard Industrial Classifications in Comparative Financial Statement Analysis. **Journal of Operational Research**, n. 79, v. 2, p.230-238. Europa, 1994.

DAMODARAN, Aswath. **A Face Oculta da Avaliação: Avaliação de Empresas da Velha Tecnologia, da Nova Tecnologia e da Nova Economia**. 1ª ed. São Paulo: Makron Books, 2002.

_____. **Avaliação de Investimentos: ferramentas e técnicas para a determinação do valor de qualquer ativo**. Rio de Janeiro: Qualitymark, 2001.

DE MOOR, Lieven; SERCU, Piet M.F.A. Country and Sector Effects in International Stock Returns Revisited. IFSAM 2006 Berlin Meetings Available at SSRN: <http://ssrn.com/abstract=676394>. Acesso em junho de 2007.

DOSI, G.; PAVITT, K.; SOETE, L. Technology and trade: an overview of the literature. In: *The economics of technical change and international trade*. Great British: **Harvester Wheatsheaf**, p.15-39, 1990.

ELTON, E. J.; GRUBER, M. J.; BROWN, S. J.; GOETZMANN, W. N. **Moderna Teoria de Carteiras e Análise de Investimentos**. 1ª ed. São Paulo: Atlas, 2004.

EVERITT, B. S. **Cluster Analysis: A Survey**, Springer-Verlog, Berlin, 1974.

FAMA, E.; Mac. BETH, J. Risk, return and equilibrium: empirical tests. **Journal of Political Economy**, v.3, n.80, p. 607-636, 1973.

FAMÁ, Rubens; VENTURA, Alessandra Montini; PEREIRA, Leonel Molero. Índices Financeiros como Previsores de Falência. VI SEMEAD – Seminários em Administração 2003, São Paulo. **Anais...** FEA-USP, São Paulo: 2003.

FARRELL, J. L. Analyzing Covariation of Returns to determine Homogeneous Stock Groupings. **Journal of Business**, v.48, n.2 (April), p.186-207, 1974.

_____. Homogeneous Stock Groupings: Implications for Portfolio Management. **Financial Analysts Journal**, v.31, n.3 (May/June), p.50-62, 1975.

FERREIRA, M. Ângelo. The Importance of Industry and Country Effects in the EMU Equity Markets. **European Financial Management**, v.12, p.341-373, 2006.

FLAVIN, T. J. The Effect of the Euro on Country versus Industry Portfolio Diversification. **Journal of International Money and Finance**. Elsevier, v. 23, n.7-8, p.1137-1158, 2004.

GRIFFIN, J.M.; KAROLYI, G.A.. Another Look at the Role of the Industrial Structure of Markets for International Diversification Strategies. **Journal of Financial Economics**, n.50, p.351-373, 1998.

GRUBEL, H.G. Internationally Diversified Portfolios: Welfare Gains and Capital loss. **American Economic Review**, v. 58, n. 5, p.1299-1314, 1968.

HAIR, J.; ANDERSON, R.E.; TATHAN, R.L.;BLACK, W.C. **Análise Multivariada de Dados**. 5ª ed.; Porto Alegre: Bookman; 2005.

HAMELINK, Foort.; HARASTY, Hélène; HILLION, Pierre. Country, Sector or Style: What matters most when constructing Global Equity Portfolios? An empirical investigation from 1990-2001. **FAME Research Paper**, n.35, 2001.

HESTON, S.L.; ROUWENHORST, K.G. Industry and country effects in international stock returns. **Journal of Portfolio Management**, spring, n.21, p.53–58, 1995.

HORNA, J.**The Study of Leisure**. Oxford University Press, 1994.

IBGE - Fundação Instituto Brasileiro de Geografia e Estatística. **Folheto informativo: Classificação Nacional de Atividades Econômicas – CNAE, 2007**. IBGE. Pesquisa Industrial Anual. Disponível em <<http://www.ibge.gov.br>>. Acesso em: 14 mar. de 2008.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 5ª ed.; New Jersey: Prentice Hall; 2002.

KING, Benjamin F. Market and Industry Factors in Stock Price Behavior. **Journal of Business**, v.39, n.1, p.139-199, 1966.

KLING, R. B. **Principles and practice of structural equation modeling**. New York: The Guilford Press, 1998, 354p.

KROGH, G.; ROOS, J. A perspective on knowledge, competence and strategy. **Personnel Review**, v.24, n.3, p. 56-76, 1995.

KRZANOWSKI, W. J.; MARRIOTT, F. H. C. **Multivariate Analysis Part 2**. John Wiley 1ª Edição, 1995.

LAPORTE, F.J; LEGENDRE, P. A Classification of Pure Malt Scotch Whiskies, Université de Montréal, **Applied Statistics** **43**, n. 1, p. 237–257, 1994.

LESSARD, D.R.. World, national and industry factors in equity returns. **Journal of Finance**, v.29, p.379–391, 1974.

LEVY, H. **Introduction to Investments**. 2ª edição. South-Western College Publishing, 1999.

LIMA, Elon L. Curso de Análise, v.2. 3ªed. Rio de Janeiro. Instituto de Matemática Pura e Aplicada, 1981.

LIVINGSTON, Miles. Industry Movements of Common Stocks. **Journal of Finance**, v.32, n.3 (June), p.861-874, 1977.

LUDÍCIBUS, Sérgio. **Análise de Balanços**. 7º ed. - São Paulo: Atlas, 1998.

MALHOTRA, N.K. **Pesquisa de Marketing: uma orientação aplicada**. Trad. Nivaldo M. Jr. e Alfredo A. de Farias. 3ª. ed. Porto Alegre: Bookman, 2001.

MATARAZZO, Dante C. **Análise financeira de balanços: abordagem básica e gerencial**. 6ª ed. São Paulo: Atlas, 2003.

MARKOWITZ, Harry. Portfolio Selection. **The Journal of Finance**, v.7, n.1, p.77-91, 1952.

MEYERS, Stephen L. A Re-Examination of Market and Industry Factors in Stock Price Behavior. **Journal of Finance**, v.28, n.2, p.695-705, 1973.

PADOVEZE, C. L.; BENEDICTO, G.C. **Análise das Demonstrações Financeiras**. 2ª ed. São Paulo: Thomson Learning, 2007.

PAVITT, K. Sectors Patterns of Technical Change: toward and taxonomy and theory. **Research Policy**, v.13, n.6, p.343-373, 1984.

PHYLAKTIS, Kate; XIA, Lichuan. Sources of Firms' Industry and Country Effects in Emerging Markets. **Journal of International Money and Finance**, v.25, n.3, p.459-475, 2006.

PESTANA, M.; GAGEIRO, J. **Análise de Dados para Ciências Sociais - A Complementaridade do SPSS**. Lisboa: Edições Silabo, 2005.

PETERAF, M. The cornerstones of competitive advantage: a resourcebased view. **Strategic Management Journal**, v.12, n.3, p.95-117, 1993.

PORCARO, R. M. **Produção de informação estatística oficial na (des)ordem social da modernidade**. Tese (Doutorado em Ciência da Informação) - Escola de Comunicação da UFRJ, Rio de Janeiro, 2000.

PORTER, M. E. The structure within industries and companies' performance. **The Review of Economics and Statistics**, v. 61, n. 2, p. 214-227, 1979.

_____. **Estratégia competitiva: técnicas para análise de indústrias e da concorrência**. 7. ed. Rio de Janeiro: Campus, 1986.

_____. What is strategy? **Harvard Business Review**. Nov.-Dec., p.61-78, 1996.

PRATES, Daniela. **Crisis Financeiras dos Países “Emergentes”**: uma explicação heterodoxa. Tese (Doutorado em Economia) - Instituto de Economia, Campinas, SP, 2006.

PRAHALAD, C. K.; HAMEL, G. The core competence of the corporation. **Harvard Business Review**, v.68, n.3, p. 79-92, 1990.

PROCIANOY, JL; HEINEBERG, R. Aspectos Determinantes do Pagamento de Proventos em Dinheiro das Empresas com Ações Negociadas na BOVESPA. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DA PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO (ENANPAD), Atibaia, 2003. **Anais...** São Paulo: 2003.

REIS, E. Estatística Multivariada Aplicada. Lisboa: Ed. Silabo, 1997.

RICHARDSON, R. et al. **Pesquisa social: métodos e técnicas**. São Paulo: Atlas, 1985.

ROLL, R. Industrial structure and the comparative behavior of international stock market indices. **Journal of Finance**, v.47, n.1, p.3-42, 1992.

ROSS, S. A.; WESTERFIELD, R.W.; JAFFE, J.F. **Administração Financeira: corporate finance**, 2ª ed. São Paulo: Atlas, 2007.

ROUWENHORST, K. G. European Equity Markets and the EMU. **Financial Analysts Journal**, v.55, n.3 (May/June), p.57-64, 1999.

SANTOS, E. S; PONTE,V. Modelo de Decisão em Gestão Econômica. **Caderno de Estudos**. São Paulo, FIPECAFI, v.10, n.19, p.43-56 (set/dez), 1998.

SAMPIERI, H. R.; COLLADO, C. F.; LÚCIO, P. B. **Metodologia de Pesquisa**. 3ª ed. São Paulo: McGraw-Hill, 2006.

SELTIZ, C.; JAHODA, M.; DEUTSCH, M.; COOK, S.W. **Research Methods in Social Relations**, 2ª ed., London: Methuen, 1959.

SIEGEL, S.; CASTELLAN, N. Jr. **Estatística não paramétrica para ciências do comportamento**. 2ª ed., Porto Alegre: Artmed, 2006.

SILVA, W.V.; GOMES, L.; MARQUES S.; LINS G. Formação de Carteiras Hipotéticas de Ativos Financeiros Usando a Técnica Análise de Cluster. In: XXXI ENCONTRO DA ASSOCIAÇÃO NACIONAL DA PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO (ENANPAD), 2007, Rio de Janeiro. **Anais da XXXI Reunião Anual da ANPAD**. Rio de Janeiro:2007.

SOLNIK, Bruno. The international pricing of risk: an empirical investigation of the world capital market structure. **Journal of Finance**, v.29, n.2, 1974.

SORENSEN, Christoffer Kok; GUTIÉRREZ, Josep Maria Puigvert. Euro area banking sector integration using hierarchical cluster analysis techniques. Working paper series nº 627, Social Science Research Network electronic library. Alemanha, 2006. Disponível em: http://ssrn.com/abstract_id=900399 . Acesso em janeiro de 2008.

STELIAROS, M.; THOMAS, D. C. The Cross-Sectional Variability of Stock-Price Returns:country and sector effects revisited. Cass Business School Research Paper Available at SSRN, 2006. Disponível em <http://ssrn.com/abstract=888721>, Acesso em janeiro de 2008.

TABACHNICK, B.; FIDELL, L. **Using Multivariate Statistics**. New York: Harper, 2000.

TOMS, M.; CUMMINGS-HILL, M.; CURRY, D.; CONE, S. Using Cluster Analysis for Deriving Menu of Structures for Automotive Mobile Multimedia Applications. International Congress and Exposition - Society of Automotive Engineers, 2001. **Anais...**Detroit, Michigan,2001.

VAN DALEN, D.B.; MEYER, Y. **Manual de técnica de la investigación educacional**. Paidós Educador. DF, México, 1994.

VERMULM, Roberto. **O setor de bens de capital**. Departamento de Economia da FEA/USP. Disponível em: <<http://www.schwartzman.org.br/simon/scipol/vermulm.pdf>>. Acesso em: fevereiro de 2008.

ANEXOS

ANEXO A - Propriedades do operador valor esperado (E)

Um método alternativo para indicar um valor esperado consiste em colocar o símbolo E , ou seja, o operador E , tal que $E(R_{ij})$ representa o valor esperado de R_{ij} , que também pode se denotado por \bar{R}_i (ELTON *et al.*, 2004).

Propriedades do operador E :

$$i) E(R_{1j} + R_{2j}) = \bar{R}_1 + \bar{R}_2$$

$$ii) E(CR_{1j}) = C\bar{R}_1$$

APÊNDICE A - Quantidade de dados omissos com percentuais para cada variável nos respectivos anos do período 2000-2007

VARIÁVEIS	2000		2001		2002		2003		2004		2005		2006		2007	
	Nº	(%)														
liquidez corrente	16	5,57	17	5,86	18	6,12	17	5,74	13	4,35	14	4,58	15	4,48	32	8,18
liquidez seca	35	12,20	36	12,41	36	12,24	36	12,16	32	10,70	34	11,11	36	10,75	62	15,86
margem operacional	43	14,98	44	15,17	46	15,65	46	15,54	46	15,38	46	15,03	54	16,12	74	18,93
margem ebitida	75	26,13	73	25,17	67	22,79	64	21,62	56	18,73	51	16,67	66	19,70	106	27,11
rentabilidade do ativo	13	4,53	14	4,83	14	4,76	15	5,07	10	3,34	9	2,94	11	3,28	23	5,88
rentabilidade do patrim. líquido exgível	35	12,20	41	14,14	47	15,99	52	17,57	52	17,39	52	16,99	57	17,01	65	16,62
total/ativo total	11	3,83	11	3,79	11	3,74	13	4,39	8	2,68	8	2,61	8	2,39	23	5,88
endividamento geral	11	3,83	11	3,79	11	3,74	13	4,39	8	2,68	8	2,61	8	2,39	23	5,88
endividamento financeiro	56	19,51	54	18,62	59	20,07	57	19,26	54	18,06	54	17,65	70	20,90	95	24,30
dividendo <i>yield</i>	176	61,32	171	58,97	180	61,22	173	58,45	166	55,52	162	52,94	186	55,52	192	49,10
<i>payout</i>	137	47,74	138	47,59	158	53,74	137	46,28	124	41,47	125	40,85	133	39,70	167	42,71
valor patrimonial da ação	10	3,48	11	3,79	11	3,74	13	4,39	9	3,01	8	2,61	8	2,39	128	32,74
liquidez em bolsa	70	24,39	66	22,76	66	22,45	64	21,62	54	18,06	56	18,30	58	17,31	52	13,30
retorno	108	37,63	101	34,83	102	34,69	101	34,12	82	27,42	89	29,08	113	33,73	74	18,93
ln (valor de mercado)	97	33,80	12	4,14	95	32,31	82	27,70	69	23,08	78	25,49	75	22,39	22	5,63
ln (receita líquida)	44	15,33	44	15,17	46	15,65	46	15,54	46	15,38	46	15,03	54	16,12	46	11,76
Ln (ativo total)	11	3,83	91	31,38	12	4,08	13	4,39	9	3,01	8	2,61	8	2,39	22	5,63
TOTAL DE CASOS	287		290		294		296		299		306		335		392	

APÊNDICE B – Estatística descritivas das variáveis para cada ano do período de 2000 - 2007

Tabela B1 - Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2000

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D. P	ASSIM.	CURTOSE
Liquidez corrente 00	271	0,0	21,0	1,6	2,1	6,4	50,9
Liquidez seca 00	252	0,0	21,0	1,3	2,1	7,0	59,1
Margem operacional 00	244	-12.404,4	3.879,8	-75,2	883,3	-11,2	159,8
Margem ebitida 00	212	-7.541,6	238,0	-42,2	591,2	-11,3	133,7
rentabilidade do ativo 00	274	-58.961,1	209,8	-322,5	3.953,7	-13,4	188,0
rentabilidade do patr. líquido 00	252	-12.409,7	10.900,0	-11,5	1.048,5	-2,1	124,6
Exigível total/ativo total 00	276	0,9	86.224,1	418,1	5.201,0	16,5	272,3
endividamento geral 00	276	-8.286,2	46.979,0	564,4	3.306,0	11,0	146,2
dividendo <i>yeld</i> 00	111	0,0	70,3	8,0	10,4	3,9	18,7
<i>payout</i> 00	150	-471,2	5.210,6	69,5	429,5	11,6	140,4
valor patrimonial ação 00	277	-58.497,3	960,5	-186,0	3.517,6	-16,6	276,6
Liquidez em bolsa 00	217	0,0	6,4	0,2	0,9	5,5	33,1
Endividamento financeiro 00	231	0,6	100,0	52,4	30,8	0,2	-1,2
retorno 00	179	-0,7	24,3	0,3	1,9	11,4	143,2
ln(valor mercado)00	190	3,9	17,7	11,8	2,5	-0,2	0,0
ln(receita liquida)00	243	5,5	17,7	12,3	2,1	-0,7	0,7
ln(ativo total)00	276	4,1	18,8	13,0	2,2	-0,4	1,9

Tabela B2 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2001

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
liquidez corrente 01	273	0,0	28,8	1,5	2,5	8,3	80,6
liquidez seca 01	254	0,0	28,8	1,2	2,5	9,1	92,2
margem operacional 01	246	-171.580,0	333,0	-735,7	10.939,9	-15,7	245,7
margem ebitida 01	217	-118.366,7	280,3	-532,1	8.036,4	-14,7	217,0
rentabilidade do ativo 01	276	-54.963,2	29,1	-345,8	3.967,1	-12,3	156,2
rentabilidade do patr. líquido 01	249	-1.874,4	150,5	-18,5	165,4	-9,3	92,9
exigível total/ativo total 01	279	0,7	11.642,6	141,5	744,5	13,9	208,7
endividamento geral 01	279	-4.742,3	20.522,4	404,1	1.747,7	7,9	80,9
dividendo <i>yeld</i> 01	119	0,2	119,3	10,7	16,9	4,3	21,0
<i>payout</i> 01	152	-403,6	1.994,2	62,3	217,1	6,4	51,9
valor patrimonial ação 01	279	-101.620,7	3.427,2	-330,2	6.089,9	-16,7	278,3
liquidez em bolsa 01	224	0,0	10,6	0,3	1,1	6,6	52,2
endividamento financeiro 01	236	0,1	100,0	51,5	30,8	0,3	-1,2
retorno 01	189	-0,7	3,7	0,0	0,5	3,4	20,9
ln(valor mercado)01	199	5,0	17,8	11,7	2,5	-0,1	-0,3
ln(receita liquida)01	246	2,7	17,9	12,4	2,2	-0,9	1,6
ln(ativo total)01	278	2,9	18,9	13,1	2,3	-0,5	1,8

Tabela B3 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2002

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
liquidez corrente 02	276	0,0	104,1	1,9	6,4	14,8	233,4
liquidez seca 02	258	0,0	104,1	1,6	6,6	14,7	226,6
margem operacional 02	248	-26.312,5	640,1	-138,8	1.689,7	-15,2	235,8
margem ebitida 02	227	-1.377,5	180,2	4,6	130,5	-9,7	98,4
rentabilidade do ativo 02	280	-47.701,9	36,7	-182,8	2.851,0	-16,7	279,6
rentabilidade do patr. líquido 02	247	-1.760,2	70,6	-28,5	157,1	-7,7	71,0
exigível total/ativo total 02	283	1,0	59.344,4	325,5	3.538,8	16,6	277,2
endividamento geral 02	283	-3.696,8	10.023,6	382,5	1.049,3	4,4	34,9
dividendo <i>yeld</i> 02	114	0,8	35,1	9,2	7,3	1,4	1,8
<i>payout</i> 02	136	-9.043,1	2.042,3	-27,3	848,9	-9,0	96,6
valor patrimonial ação 02	283	-7.926,3	834,7	-9,7	478,2	-16,2	269,1
liquidez em bolsa 02	228	0,0	10,9	0,3	1,1	6,7	55,7
endividamento financeiro 02	235	0,4	100,0	54,5	30,5	0,2	-1,3
retorno 02	192	-1,0	3,0	0,0	0,6	2,0	6,0
ln(valor mercado)02	199	5,3	17,7	11,6	2,6	-0,1	-0,4
ln(receita líquida)02	248	5,5	18,1	12,5	2,2	-0,7	0,7
ln(ativo total)02	282	3,4	19,1	13,1	2,3	-0,4	1,3

Tabela B4 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2003

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
liquidez corrente 03	279	0,0	1.135,7	5,7	68,0	16,6	276,8
liquidez seca 03	260	0,0	1.135,7	5,7	70,5	16,0	258,0
margem operacional 03	250	-189.861,3	857,8	-771,2	12.008,8	-15,8	249,9
margem ebitida 03	232	-2.209,9	1.081,7	10,2	188,0	-6,8	94,3
rentabilidade do ativo 03	281	-2.200,0	27,7	-17,5	156,2	-11,6	148,4
rentabilidade do patr. líquido 03	244	-600,7	414,6	2,2	69,3	-2,6	35,9
exigível total/ativo total 03	283	0,1	5.500,0	152,2	580,6	8,4	72,2
endividamento geral 03	283	-3.744,1	29.954,5	356,2	1.941,9	12,7	193,1
dividendo <i>yeld</i> 03	123	0,3	22,6	5,7	4,6	1,4	1,7
<i>payout</i> 03	159	-973,2	870,7	50,0	130,1	-1,4	34,8
valor patrimonial ação 03	283	-5.781,0	842,1	-2,6	353,5	-15,5	255,7
liquidez em bolsa 03	232	0,0	11,0	0,3	1,0	6,8	58,3
endividamento financeiro 03	239	0,3	100,0	56,2	30,9	0,1	-1,3
retorno 03	195	-1,3	0,9	0,3	0,4	-1,4	2,6
ln(valor mercado)03	214	4,8	18,2	12,1	2,6	-0,1	-0,3
ln(receita líquida)03	250	4,7	18,4	12,6	2,3	-0,8	1,2
ln(ativo total)03	283	1,1	19,3	13,2	2,4	-0,7	2,7

Tabela B5 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2004

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
liquidez corrente 04	282	0,0	14,8	1,6	1,8	4,7	27,8
liquidez seca 04	264	0,0	14,8	1,2	1,6	5,5	37,6
margem operacional 04	250	-272.118,7	6.351,6	-1.098,5	17.219,5	-15,8	249,4
margem ebitida 04	240	-11.666,7	176,6	-41,6	765,7	-14,8	225,1
rentabilidade do ativo 04	285	-13.605,3	45,6	-65,7	818,8	-16,1	265,9
rentabilidade do patr. líquido 04	244	-1.667,4	178,8	0,1	117,2	-12,3	170,9
exigível total/ativo total 04	287	1,3	6.972,1	190,2	759,5	7,5	57,6
endividamento geral 04	287	-2.156,2	64.493,6	490,4	3.856,9	16,1	267,8
dividendo yield 04	133	0,2	24,8	5,6	4,7	2,0	4,5
payout 04	172	-11.125,8	2.007,9	2,1	873,3	-12,1	156,8
valor patrimonial ação 04	286	-6.811,8	1.050,0	-6,3	413,8	-15,6	259,4
liquidez em bolsa 04	242	0,0	8,8	0,3	0,9	5,7	41,0
endividamento financeiro 04	242	0,3	100,0	53,0	31,7	0,2	-1,4
retorno 04	217	-0,9	12,8	0,6	1,4	6,0	45,8
ln(valor mercado)04	227	5,3	18,5	12,3	2,6	-0,1	-0,4
ln(receita liquida)04	250	3,4	18,5	12,8	2,3	-0,9	1,8
ln(ativo total)04	286	0,7	19,3	13,2	2,5	-0,9	3,6

Tabela B6 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2005

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
Liquidez corrente 05	292	0,0	12,5	1,6	1,5	3,4	16,9
Liquidez seca 05	272	0,0	12,5	1,3	1,4	4,2	24,0
Margem operacional 05	260	-985.004,6	1.322,0	-3.804,9	61.087,5	-16,1	260,0
Margem ebitida 05	255	-9.595,5	1.262,2	-21,1	612,2	-15,2	238,2
rentabilidade do ativo 05	297	-4.300,0	81,0	-26,6	268,7	-14,2	219,5
rentabilidade do patr. líquido 05	254	-1.526,8	459,2	-2,6	128,3	-8,3	89,6
Exigível total/ativo total 05	298	3,1	29.212,4	267,7	1.813,4	14,2	221,4
endividamento geral 05	298	-1.644,0	15.265,7	365,2	1.280,4	8,4	83,3
dividendo yield 05	144	0,1	39,3	6,4	6,4	2,3	6,8
payout 05	181	-3.635,6	1.120,4	50,4	305,5	-9,5	119,9
valor patrimonial ação 05	298	-6.895,6	1.133,4	-5,0	411,8	-15,8	266,7
Liquidez em bolsa 05	250	0,0	7,7	0,3	0,9	5,3	32,0
endividamento financeiro 05	252	0,3	100,0	50,9	33,0	0,3	-1,4
retorno 05	217	-0,8	7,8	0,2	0,9	4,3	28,3
ln(valor mercado)05	228	5,4	18,9	12,7	2,6	-0,2	-0,4
ln(receita liquida)05	260	3,1	18,7	12,9	2,3	-0,8	1,4
ln(ativo total)05	298	0,0	19,4	13,3	2,6	-1,2	4,6

Tabela B7 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2006

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
Liquidez corrente 06	320	0,0	120,7	2,1	6,9	16,0	272,3
Liquidez seca 06	299	0,0	120,7	1,9	7,1	15,7	260,3
		-					
Margem operacional 06	281	1.209.790,0	18.289,5	-4.196,8	72.189,0	-16,8	280,8
Margem ebitada 06	269	-1.884,1	1.864,2	13,5	175,6	-0,8	98,3
rentabilidade do ativo 06	324	-183.533,3	63,0	-612,6	10.203,8	-17,9	322,7
rentabilidade do patr. líquido 06	278	-7.809,0	640,2	-29,7	505,5	-13,9	207,8
Exigível total/ativo total 06	327	0,8	31.600,0	312,7	2.044,2	12,2	173,4
endividamento geral 06	327	-2.560,3	31.682,8	406,5	2.253,2	10,8	131,7
dividendo <i>yeld</i> 06	149	0,1	25,5	4,5	4,1	2,2	6,5
<i>payout</i> 06	202	-185,6	461,3	57,3	63,5	2,2	12,9
valor patrimonial ação 06	327	-511,8	2.162,5	25,9	152,8	10,1	128,3
liquidez em bolsa 06	277	0,0	11,1	0,3	0,9	7,5	72,9
Endividamento financeiro 06	265	0,3	100,0	47,9	33,0	0,4	-1,3
retorno 06	222	-0,5	22,7	0,7	1,9	7,9	82,4
ln(valor mercado)06	260	5,6	19,2	13,1	2,6	-0,4	-0,2
ln(receita líquida)06	281	3,0	18,9	12,8	2,4	-0,9	1,4
ln(ativo total)06	327	0,0	19,5	13,3	2,8	-1,4	4,9

Tabela B8 – Desvio padrão, média, assimetria e curtose para os dados das variáveis em relação ao ano 2007

VARIÁVEIS	N	MÍNIMO	MÁXIMO	MÉDIA	D.P	ASSIM	CURTOSE
liquidez corrente 07	359	0,0	24,8	2,0	2,2	4,7	35,4
liquidez seca 07	329	0,0	24,8	1,8	2,1	5,4	44,4
margem operacional 07	317	-743.273,8	134.348,3	-1.842,9	42.469,3	-16,8	296,9
margem ebitada 07	285	-3.182,4	42.055,8	159,7	2.502,1	16,6	279,7
rentabilidade do ativo 07	368	-3.900,0	426,7	-21,5	248,2	-12,4	174,4
rentabilidade do patri. líquido 07	326	-242,4	1.551,7	16,8	90,9	14,8	252,2
exigível total/ativo total 07	368	0,0	23.903,8	211,9	1.493,6	13,2	190,3
endividamento geral 07	368	-9.959,9	11.050,1	182,6	972,5	-0,1	80,6
dividendo <i>yeld</i> 07	199	0,0	32,3	3,4	3,8	3,3	17,2
<i>payout</i> 07	224	-77,8	1.877,2	63,3	134,8	11,2	148,4
valor patrimonial ação 07	369	-424,8	2.166,3	27,3	151,4	9,7	119,2
liquidez em bolsa 07	345	0,0	10,4	0,2	0,8	9,1	98,8
endividamento financeiro 07	296	0,1	100,0	45,3	32,5	0,4	-1,2
retorno 07	263	-1,0	22,6	1,0	2,4	5,1	33,6
ln(valor mercado)07	339	8,2	19,8	13,7	2,1	-0,3	0,1
ln(receita líquida)07	317	3,7	19,0	13,0	2,3	-0,9	1,7
ln(ativo total)07	369	0,0	19,7	13,7	2,4	-1,2	5,7

APÊNDICE C – Teste Kruskal Wallis aplicado para as classificações setoriais NAICS, e BOVESPA, segundo cada ano do período de 2000-2007

Tabela C1- Resultados do Teste Kruskal Wallis, classificação NAICS, 200-2007

VARIÁVEIS	CLASSIFICAÇÃO NAICS							
	TESTE KRUSKAL WALLIS – SIGNIFICÂNCIA							
	2000	2001	2002	2003	2004	2005	2006	2007
liquidez corrente	0,01	0,00	0,00	0,00	0,39	0,00	0,12	0,00
liquidez seca	0,02	0,00	0,00	0,07	0,23	0,00	0,34	0,00
margem operacional	0,01	0,01	0,02	0,05	0,01	0,07	0,00	0,00
margem ebitida	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00
rentabilidade do ativo	0,06	0,02	0,00	0,32	0,10	0,08	0,11	0,00
rentabilidade do patrim. líquido	0,00	0,05	0,00	0,03	0,00	0,00	0,03	0,00
Exigível total/ativo total	0,03	0,05	0,10	0,05	0,12	0,08	0,00	0,00
endividamento geral	0,10	0,02	0,05	0,01	0,00	0,00	0,00	0,00
endividamento financeiro	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
dividendo <i>yeld</i>	0,03	0,01	0,04	0,10	0,22	0,07	0,00	0,00
<i>Payout</i>	0,54	0,26	0,11	0,59	0,52	0,15	0,01	0,00
valor patrimonial ação	0,00	0,00	0,00	0,01	0,02	0,01	0,03	0,04
liquidez em bolsa	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,06
ln (valor de mercado)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ln (receita líquida)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ln (ativo total)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
retorno	0,96	0,10	0,00	0,63	0,00	0,02	0,05	0,00

Tabela C2- Resultados do Teste Kruskal Wallis, classificação BOVESPA, 200-2007

VARIÁVEIS	CLASSIFICAÇÃO BOVESPA							
	TESTE KRUSKAL WALLIS – SIGNIFICÂNCIA							
	2000	2001	2002	2003	2004	2005	2006	2007
liquidez corrente	0,19	0,01	0,01	0,05	0,19	0,04	0,01	0,00
liquidez seca	0,10	0,43	0,02	0,26	0,35	0,36	0,10	0,01
margem operacional	0,00	0,03	0,06	0,00	0,00	0,00	0,00	0,00
margem ebitida	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
rentabilidade do ativo	0,12	0,03	0,04	0,00	0,00	0,01	0,00	0,00
rentabilidade do patrim. líquido	0,03	0,02	0,02	0,00	0,00	0,05	0,00	0,00
exigível total/ativo total	0,02	0,02	0,41	0,11	0,06	0,04	0,01	0,01
endividamento geral	0,04	0,02	0,22	0,03	0,02	0,01	0,00	0,00
endividamento financeiro	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00
dividendo <i>yeld</i>	0,11	0,02	0,04	0,15	0,69	0,02	0,00	0,00
<i>Payout</i>	0,76	0,02	0,13	0,51	0,55	0,42	0,00	0,01
valor patrimonial ação	0,07	0,02	0,10	0,06	0,10	0,05	0,03	0,02
liquidez em bolsa	0,00	0,02	0,00	0,00	0,00	0,01	0,01	0,03
ln (valor de mercado)	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00
ln (receita líquida)	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00
ln (ativo total)	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00
retorno	0,37	0,02	0,00	0,00	0,00	0,03	0,03	0,26

APÊNDICE D – Número de *clusters* com mais de um elemento, considerando o mesmo número de subdivisões da classificação NAICS nível 1, no período de 2000 – 2007

CLASSIFICAÇÃO NAICS - COM 20 CLUSTERS				
ANO	SEM VARIÁVEIS DE TAMANHO		COM VARIÁVEIS DE TAMANHO	
	<i>Cluster índice</i>	<i>Cluster mercado</i>	<i>Cluster índice</i>	<i>Cluster mercado</i>
2000	13	17	13	18
2001	15	16	15	18
2002	15	16	15	19
2003	16	18	16	19
2004	16	17	16	19
2005	15	17	15	18
2006	14	13	16	17
2007	12	17	15	19

APÊNDICE E – Classificação NAICS com frequência percentual nos setores, com número de clusters definidos conforme a regra de parada, no período de 2000 – 2007

Tabela E1 – Classificação NAICS, com 10 clusters para o cluster índice e 3 clusters para o cluster mercado, no ano 2000

CLASSIFICAÇÃO NAICS – ANO 2000 - MÉTODO DE PARADA													
SETORES	Cluster índice										Cluster mercado		
	1	2	3	4	5	6	7	8	9	10	1	2	3
adm. empresas e empreendimentos	16,0	3,9	4,8	0,0	16,7	5,6	100,0	0,0	100,0	0,0	8,0	0,0	0,0
agr., pec., silvic., pesca e caça	0,0	0,0	0,0	0,0	0,0	0,9	0,0	0,0	0,0	0,0	0,4	0,0	0,0
artes, entretenimento e recreação	0,0	0,0	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,0
comércio varejista	2,0	7,8	4,8	0,0	0,0	1,9	0,0	33,3	0,0	0,0	4,0	0,0	0,0
Construção	2,0	9,8	1,6	0,0	0,0	4,6	0,0	0,0	0,0	0,0	4,3	0,0	0,0
Educação	0,0	0,0	1,6	0,0	0,0	1,9	0,0	0,0	0,0	0,0	1,1	0,0	0,0
empresa de eletricidade, gás e água	6,0	2,0	31,7	66,7	0,0	7,4	0,0	33,3	0,0	0,0	12,0	11,1	50,0
hotel e restaurante	0,0	2,0	0,0	0,0	0,0	0,9	0,0	0,0	0,0	1,0	0,7	0,0	0,0
imobiliária e locadora de outros bens	0,0	0,0	0,0	0,0	16,7	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,0
indústria manufatureira	54,0	70,6	27,0	33,3	33,3	45,4	0,0	33,3	0,0	0,0	47,8	0,0	50,0
Informação	4,0	0,0	17,5	0,0	16,7	5,6	0,0	0,0	0,0	0,0	5,4	66,7	0,0
Mineração	0,0	0,0	1,6	0,0	0,0	2,8	0,0	0,0	0,0	0,0	0,7	22,2	0,0
outros serviços	2,0	2,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,7	0,0	0,0
serviços de apoio e ger. de res. remediação	0,0	0,0	0,0	0,0	0,0	0,9	0,0	0,0	0,0	0,0	0,4	0,0	0,0
serviços financeiros e seguros	8,0	0,0	0,0	0,0	16,7	20,4	0,0	0,0	0,0	0,0	9,8	0,0	0,0
serviços profis., científicos e técnicos	2,0	0,0	0,0	0,0	0,0	0,9	0,0	0,0	0,0	0,0	0,7	0,0	0,0
Transporte e armazenamento	4,0	2,0	7,9	0,0	0,0	0,9	0,0	0,0	0,0	0,0	3,3	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	50	51	63	3	6	108	1	3	1	1	276	9	2

Tabela E2 – Classificação NAICS, com 4 clusters para o cluster índice e 5 clusters para o cluster mercado, no ano 2001

CLASSIFICAÇÃO NAICS - ANO 2001 - MÉTODO DE PARADA										
SETORES	Cluster índice				Cluster mercado					
	1	2	3	4	1	2	3	4	5	
adm. empresas e empreendimentos	7,1	5,6	40,0	50,0	9,4	6,3	2,9	0,0	0,0	
agric., pecuária, silv., pesca e caça	0,0	1,4	0,0	0,0	0,6	0,0	0,0	0,0	0,0	
artes, entretenimento e recreação	0,5	0,0	0,0	0,0	0,6	0,0	0,0	0,0	0,0	
comércio varejista	3,3	5,6	0,0	0,0	5,0	3,2	0,0	0,0	0,0	
Construção	2,8	9,9	0,0	0,0	3,9	4,8	8,6	0,0	0,0	
Educação	1,9	0,0	0,0	0,0	2,2	0,0	0,0	0,0	0,0	
empresa de eletricidade, gás e água	15,1	4,2	0,0	0,0	11,6	9,5	17,1	10,0	100,0	
hotel e restaurante	0,0	1,4	20,0	0,0	1,1	0,0	0,0	0,0	0,0	

comércio varejista	2,6	7,6	3,6	0,0	5,9	0,0	5,1	0,0	3,0	3,1	0,0	0,0	0,0
Construção	3,1	7,6	3,6	11,1	3,5	10,0	1,7	0,0	3,0	9,2	4,2	0,0	0,0
Educação	1,6	0,0	0,0	11,1	4,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
empresa de eletricidade, gás e água	14,0	4,5	25,0	0,0	12,9	10,0	10,2	100,0	21,2	10,8	4,2	18,8	0,0
hotel e restaurante	0,0	3,0	0,0	0,0	1,2	0,0	0,0	0,0	0,0	1,5	0,0	0,0	0,0
imobiliária e locadora de outros bens	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,0	0,0	0,0	0,0	0,0
indústria manufatureira	37,3	68,2	46,4	44,4	23,5	70,0	72,9	0,0	39,4	47,7	66,7	25,0	0,0
Informação	8,8	3,0	7,1	0,0	9,4	0,0	0,0	0,0	3,0	3,1	12,5	31,3	66,7
Mineração	2,1	0,0	0,0	0,0	2,4	0,0	0,0	0,0	0,0	0,0	0,0	6,3	33,3
outros serviços	0,5	1,5	0,0	0,0	1,2	0,0	1,7	0,0	0,0	0,0	0,0	0,0	0,0
serviços de apoio e ger. de res. remediação	1,0	0,0	0,0	0,0	1,2	10,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
serviços financeiros e seguros	13,0	0,0	7,1	0,0	12,9	0,0	1,7	0,0	15,2	12,3	0,0	12,5	0,0
serviços profis., científicos e técnicos	1,6	0,0	0,0	0,0	1,2	0,0	0,0	0,0	3,0	1,5	0,0	0,0	0,0
transporte e armazenamento	3,6	1,5	3,6	11,1	5,9	0,0	1,7	0,0	3,0	3,1	4,2	0,0	0,0
TOTAL (%)	100												
TOTAL CASOS	193	66	28	9	85	10	59	1	33	65	24	16	3

Tabela E5 – Classificação NAICS, com 4 clusters para o cluster índice e 4 clusters para o cluster mercado, no ano 2004

CLASSIFICAÇÃO NAICS – ANO 2004 - MÉTODO DE PARADA								
SETORES	Cluster índice				Cluster mercado			
	1	2	3	4	1	2	3	4
adm. empresas e empreendimentos	7,9	3,3	33,3	16,7	8,7	0,0	0,0	0,0
agric., pecuária, silv., pesca e caça	0,9	0,0	0,0	0,0	0,8	0,0	0,0	0,0
artes, entretenimento e recreação	0,4	0,0	0,0	0,0	0,4	0,0	0,0	0,0
assistência médica e social	0,4	0,0	0,0	0,0	0,4	0,0	0,0	0,0
comércio atacadista	0,4	0,0	0,0	0,0	0,4	0,0	0,0	0,0
comércio varejista	3,1	6,7	0,0	0,0	4,2	0,0	0,0	0,0
Construção	4,0	6,7	0,0	0,0	4,9	0,0	0,0	0,0
Educação	0,4	1,7	16,7	16,7	1,5	0,0	0,0	0,0
empresa de eletricidade, gás e água	15,9	1,7	0,0	0,0	11,8	18,2	100,0	7,7
hotel e restaurante	0,0	3,3	0,0	0,0	0,4	4,5	0,0	0,0
imobiliária e locadora de outros bens	0,0	0,0	16,7	0,0	0,4	0,0	0,0	0,0
indústria manufatureira	39,2	65,0	16,7	50,0	42,6	72,7	0,0	30,8
Informação	7,0	6,7	0,0	16,7	6,5	0,0	0,0	30,8
Mineração	1,8	0,0	0,0	0,0	0,8	0,0	0,0	15,4
outros serviços	0,4	1,7	0,0	0,0	0,8	0,0	0,0	0,0
serviços de apoio e ger. de res. remediação	0,9	0,0	0,0	0,0	0,8	0,0	0,0	0,0
serviços financeiros e seguros	11,9	0,0	16,7	0,0	9,9	0,0	0,0	15,4
serviços profis., científicos e técnicos	1,3	0,0	0,0	0,0	1,1	0,0	0,0	0,0
transporte e armazenamento	4,0	3,3	0,0	0,0	3,8	4,5	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100
TOTAL CASOS	227	60	6	6	263	22	1	13

Tabela E6 – Classificação NAICS, com 6 clusters para o cluster índice e 2 clusters para o cluster mercado, no ano 2005

CLASSIFICAÇÃO NAICS - ANO 2005 - MÉTODO DE PARADA									
SETORES	Cluster índice						Cluster mercado		
	1	2	3	4	5	6	1	2	
adm. empresas e empreendimentos	8,8	6,1	15,0	4,1	0,0	66,7	7,9	3,8	
agric., pecuária, silv., pesca e caça	0,0	0,0	0,0	2,0	0,0	0,0	0,7	0,0	
artes, entretenimento e recreação	0,9	0,0	0,0	0,0	0,0	0,0	0,4	0,0	
assistência médica e social	0,0	0,0	0,0	1,0	0,0	0,0	0,4	0,0	
comércio atacadista	0,9	0,0	0,0	0,0	0,0	0,0	0,4	0,0	
comércio varejista	2,7	7,6	0,0	3,1	0,0	0,0	3,9	0,0	
Construção	2,7	7,6	5,0	4,1	0,0	0,0	3,9	7,7	
Educação	0,9	1,5	5,0	1,0	0,0	0,0	1,4	0,0	
empresa de eletricidade, gás e água	29,2	4,5	5,0	2,0	0,0	0,0	12,5	15,4	
hotel e restaurante	0,0	1,5	0,0	0,0	16,7	0,0	0,7	0,0	
imobiliária e locadora de outros bens	0,0	0,0	5,0	0,0	0,0	0,0	0,4	0,0	
indústria manufatureira	26,5	63,6	55,0	46,9	50,0	33,3	44,3	34,6	
Informação	15,0	1,5	5,0	2,0	16,7	0,0	6,8	11,5	
Mineração	1,8	0,0	0,0	2,0	0,0	0,0	0,7	7,7	
outros serviços	0,0	1,5	0,0	1,0	0,0	0,0	0,7	0,0	
serviços de apoio e ger. de res. remediação	0,9	0,0	0,0	1,0	0,0	0,0	0,7	0,0	
serviços financeiros e seguros	1,8	0,0	5,0	25,5	16,7	0,0	9,3	11,5	
serviços profis., científicos e técnicos	0,9	1,5	0,0	1,0	0,0	0,0	0,7	3,8	
transporte e armazenamento	7,1	3,0	0,0	3,1	0,0	0,0	4,3	3,8	
TOTAL (%)	100	100	100	100	100	100	100	100	
TOTAL CASOS	113	66	20	98	6	3	280	26	

Tabela E7 – Classificação NAICS, com 8 clusters para o cluster índice e 7 clusters para o cluster mercado, no ano 2006

CLASSIFICAÇÃO NAICS - ANO 2006 - MÉTODO DE PARADA															
SETORES	Cluster índice								Cluster mercado						
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
adm. empresas e empreendimentos	11,0	8,0	1,6	12,0	0,0	0,0	0,0	0,0	16,7	5,8	0,0	7,3	11,1	7,1	0,0
agric., pec., silv., pesca e caça	0,0	0,9	1,6	4,0	0,0	0,0	0,0	0,0	0,0	1,3	0,0	0,0	1,9	0,0	0,0
artes, entretem. e recreação	1,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,2	0,0	0,0	0,0
assistência médica e social	1,0	1,8	0,0	4,0	0,0	0,0	0,0	0,0	0,0	1,3	0,0	2,4	0,0	0,0	0,0
comércio atacadista	2,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	0,0	0,0	3,6	0,0
comércio varejista	1,0	3,5	9,7	0,0	0,0	0,0	0,0	0,0	16,7	3,2	12,5	4,9	0,0	0,0	0,0
Construção	4,0	6,2	9,7	8,0	0,0	0,0	0,0	0,0	0,0	7,7	12,5	7,3	0,0	0,0	0,0
educação	0,0	0,9	3,2	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,7	1,9	0,0	0,0
empresa de	25,0	2,7	4,8	8,0	35,5	0,0	100,0	0,0	16,7	16,8	0,0	17,1	1,9	10,7	0,0

eletric., gás e água																
hotel e restaurante	0,0	0,0	3,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	0,0	1,9	0,0	0,0
imob. e locadora de outros bens	0,0	0,0	0,0	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	0,0	0,0	0,0	0,0
indústria manufatureira	43,0	38,9	54,8	28,0	29,0	50,0	0,0	100,0	33,3	43,2	37,5	22,0	75,9	28,6	0,0	0,0
Informação	5,0	5,3	6,5	20,0	9,7	0,0	0,0	0,0	16,7	4,5	12,5	9,8	0,0	21,4	0,0	0,0
mineração	2,0	2,7	0,0	4,0	0,0	50,0	0,0	0,0	0,0	2,6	0,0	1,2	0,0	0,0	100,0	0,0
outros serviços	0,0	0,9	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,2	1,9	0,0	0,0	0,0
serv. apoio e ger. de res. remed.	1,0	1,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	2,4	0,0	0,0	0,0	0,0
serviços financ. e seguros	0,0	19,5	0,0	0,0	22,6	0,0	0,0	0,0	0,0	7,7	12,5	12,2	3,7	14,3	0,0	0,0
serviços profis., cient. e técnicos	0,0	0,9	1,6	0,0	3,2	0,0	0,0	0,0	0,0	1,9	0,0	0,0	0,0	0,0	0,0	0,0
transporte e armazenamento	4,0	6,2	1,6	4,0	0,0	0,0	0,0	0,0	0,0	1,3	12,5	7,3	0,0	14,3	0,0	0,0
TOTAL (%)	100															
TOTAL CASOS	100	113	62	25	31	2	1	1	6	155	8	82	54	28	2	2

Tabela E8 – Classificação NAICS, com 10 clusters para o cluster índice e 4 clusters para o cluster mercado, no ano 2007

CLASSIFICAÇÃO NAICS – ANO 2007 -MÉTODO DE PARADA														
SETORES	Cluster índice										Cluster mercado			
	1	2	3	4	5	6	7	8	9	10	1	2	3	4
adm. empresas e empreendimentos	2,3	2,9	9,6	9,7	0,0	13,3	3,6	40,0	50,0	50,0	6,3	0,0	4,9	11,1
agric., pecuária, silv., pesca e caça	1,5	1,4	0,0	0,0	0,0	0,0	3,6	0,0	0,0	0,0	1,2	0,0	0,0	0,0
artes, entretenimento e recreação	0,0	1,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,0	0,0
assistência médica e social	0,8	0,0	2,4	0,0	0,0	0,0	7,1	0,0	0,0	0,0	1,5	0,0	0,0	0,0
comércio atacadista	0,8	0,0	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	2,4	0,0
comércio varejista	3,8	8,7	1,2	3,2	0,0	0,0	0,0	0,0	0,0	0,0	2,7	14,3	7,3	0,0
Construção	3,8	10,1	8,4	6,5	0,0	6,7	50,0	0,0	0,0	0,0	10,2	0,0	4,9	0,0
Educação	1,5	1,4	0,0	0,0	0,0	13,3	0,0	0,0	0,0	0,0	1,5	0,0	0,0	0,0
empresa de eletricidade, gás e água	3,8	5,8	20,5	9,7	81,0	0,0	0,0	0,0	25,0	0,0	12,0	14,3	14,6	0,0
hotel e restaurante	0,0	2,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,0	11,1
Imob. locadora de bens	0,8	0,0	1,2	0,0	0,0	6,7	0,0	0,0	0,0	0,0	0,9	0,0	0,0	0,0
indústria manufatureira	44,4	60,9	34,9	9,7	9,5	46,7	17,9	40,0	25,0	50,0	39,5	14,3	31,7	55,6
Informação	6,0	1,4	7,2	9,7	4,8	0,0	10,7	20,0	0,0	0,0	5,1	0,0	14,6	0,0
Mineração	2,3	0,0	2,4	6,5	0,0	0,0	0,0	0,0	0,0	0,0	1,2	42,9	0,0	0,0
outros serviços	1,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,0	11,1
serviços de apoio e ger. De res. remediação	0,8	0,0	2,4	3,2	0,0	6,7	0,0	0,0	0,0	0,0	1,2	14,3	0,0	0,0
serviços fin. e seguros	18,0	0,0	1,2	41,9	4,8	6,7	3,6	0,0	0,0	0,0	11,1	0,0	9,8	0,0
serviços profis., científicos e técnicos	1,5	1,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,0	0,0	11,1
transporte e armazenamento	6,8	1,4	7,2	0,0	0,0	0,0	3,6	0,0	0,0	0,0	3,9	0,0	9,8	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	133	69	83	31	21	15	28	5	4	2	334	7	41	9

APÊNDICE F - Critério classificação BOVESPA, com variáveis de tamanho, para cada ano do período 2000-2007

Tabela F1 – Classificação e critério BOVESPA, *cluster índice*, ano 2000

CLASSIFICAÇÃO BOVESPA –C/VARIAVEIS DE TAMANHO, ANO 2000										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	4,5	18,5	0,0	18,2	13,0	0,0	0,0	0,0	0,0	0,0
construção e transporte	5,7	16,3	33,3	13,6	11,1	0,0	0,0	0,0	0,0	0,0
consumo cíclico	11,4	25,0	0,0	36,4	20,4	0,0	0,0	25,0	0,0	0,0
consumo não cíclico	5,7	8,7	0,0	9,1	7,4	10,0	0,0	0,0	0,0	0,0
financeiro e outros	14,8	13,0	0,0	9,1	22,2	20,0	100,0	0,0	100,0	100,0
materiais básicos	17,0	16,3	0,0	4,5	18,5	25,0	0,0	25,0	0,0	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	1,9	5,0	0,0	0,0	0,0	0,0
tecnologia da informação	1,1	1,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	10,2	0,0	0,0	9,1	0,0	30,0	0,0	25,0	0,0	0,0
utilidade pública	29,5	1,1	66,7	0,0	5,6	10,0	0,0	25,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	88	92	3	22	54	20	1	4	1	2

Tabela F2 – Classificação e critério BOVESPA, *cluster mercado*, ano 2000

CLASSIFICAÇÃO BOVESPA – COM VARIAVEIS DE TAMANHO, ANO 2000										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	2,9	10,7	13,7	0,0	10,0	21,4	0,0	33,3	0,0	0,0
construção e transporte	2,9	28,6	9,8	0,0	11,4	14,3	0,0	8,3	0,0	0,0
consumo cíclico	10,0	25,0	9,8	0,0	18,6	38,1	11,1	25,0	0,0	100,0
consumo não cíclico	8,6	10,7	2,0	0,0	11,4	4,8	0,0	8,3	0,0	0,0
financeiro e outros	10,0	10,7	25,5	100,0	24,3	4,8	0,0	16,7	0,0	0,0
materiais básicos	22,9	10,7	17,6	0,0	15,7	16,7	11,1	0,0	0,0	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	1,4	0,0	11,1	0,0	0,0	0,0
tecnologia da informação	1,4	0,0	0,0	0,0	1,4	0,0	0,0	0,0	0,0	0,0
telecomunicações	11,4	0,0	3,9	0,0	2,9	0,0	55,6	8,3	0,0	0,0
utilidade pública	30,0	3,6	17,6	0,0	2,9	0,0	11,1	0,0	100,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	70	28	51	3	70	42	9	12	1	1

Tabela F3 – Classificação e critério BOVESPA, *cluster índice*, ano 2001

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2001										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	4,7	15,7	6,3	12,2	0,0	18,8	0,0	0,0	100,0	50,0
construção e transporte	5,9	17,1	6,3	11,2	0,0	12,5	0,0	0,0	0,0	50,0
consumo cíclico	8,2	22,9	18,8	18,4	100,0	43,8	50,0	0,0	0,0	0,0
consumo não cíclico	5,9	10,0	0,0	8,2	0,0	6,3	25,0	0,0	0,0	0,0
financeiro e outros	8,2	12,9	6,3	26,5	0,0	12,5	25,0	100,0	0,0	0,0
materiais básicos	22,4	14,3	18,8	15,3	0,0	6,3	0,0	0,0	0,0	0,0
petróleo, gás e biocomb.	1,2	1,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	1,2	1,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	15,3	1,4	0,0	4,1	0,0	0,0	0,0	0,0	0,0	0,0
utilidade pública	27,1	2,9	43,8	4,1	0,0	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	85	70	16	98	1	16	4	1	1	2

Tabela F4 – Classificação e critério BOVESPA, *cluster mercado*, ano 2001

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS TAMANHO, ANO 2001										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	5,4	17,2	14,5	5,0	18,2	26,7	0,0	9,1	0,0	0,0
construção e transporte	6,5	20,7	14,5	0,0	15,2	13,3	0,0	9,1	0,0	0,0
consumo cíclico	12,9	17,2	15,8	0,0	36,4	46,7	10,0	36,4	0,0	50,0
consumo não cíclico	3,2	10,3	13,2	15,0	6,1	0,0	0,0	0,0	0,0	0,0
financeiro e outros	12,9	17,2	22,4	25,0	12,1	6,7	10,0	18,2	0,0	0,0
materiais básicos	18,3	13,8	17,1	30,0	12,1	6,7	10,0	0,0	0,0	50,0
petróleo, gás e biocomb.	1,1	0,0	0,0	0,0	0,0	0,0	10,0	0,0	0,0	0,0
tecnologia da informação	1,1	0,0	0,0	0,0	0,0	0,0	0,0	9,1	0,0	0,0
telecomunicações	8,6	0,0	2,6	10,0	0,0	0,0	50,0	9,1	0,0	0,0
utilidade pública	30,1	3,4	0,0	15,0	0,0	0,0	10,0	9,1	100,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	93	29	76	20	33	15	10	11	1	2

Tabela F5 – Classificação e critério BOVESPA, *cluster índice*, ano 2002

CLASSIFICAÇÃO BOVESPA –COM VARIÁVEIS TAMANHO, ANO 2002										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	1,5	16,4	10,8	18,8	28,6	0,0	0,0	0,0	0,0	0,0
construção e transporte	4,4	13,7	9,6	18,8	28,6	0,0	0,0	0,0	0,0	0,0
consumo cíclico	2,9	28,8	16,9	27,1	28,6	0,0	25,0	0,0	0,0	50,0
consumo não cíclico	7,4	11,0	7,2	4,2	28,6	0,0	0,0	0,0	0,0	0,0
financeiro e outros	11,8	6,8	24,1	18,8	28,6	100,0	0,0	0,0	100,0	0,0
materiais básicos	20,6	17,8	18,1	8,3	28,6	0,0	25,0	0,0	0,0	0,0
petróleo, gás e biocomb.	1,5	0,0	1,2	0,0	28,6	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,0	1,4	0,0	2,1	28,6	0,0	0,0	0,0	0,0	0,0
telecomunicações	19,1	0,0	2,4	2,1	28,6	0,0	0,0	100,0	0,0	50,0
utilidade pública	30,9	4,1	9,6	0,0	28,6	0,0	50,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	68	73	83	48	7	2	4	1	2	2

tecnologia da informação	0,0	0,0	0,0	0,0	0,0	1,5	5,3	0,0	0,0	0,0
telecomunicações	2,0	4,3	2,6	0,0	0,0	9,2	5,3	26,1	0,0	50,0
utilidade pública	4,0	4,3	5,3	33,3	###	32,3	5,3	13,0	0,0	0,0
TOTAL (%)	100									
TOTAL CASOS	50	23	76	15	1	65	19	23	22	2

Tabela F9 – Classificação e critério BOVESPA, *cluster* mercado, ano 2004

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2004										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	13,8	26,4	0,0	7,5	6,3	10,5	0,0	0,0	0,0	100,0
construção e transporte	6,9	9,4	2,5	24,5	14,1	5,3	0,0	0,0	0,0	0,0
consumo cíclico	8,6	30,2	2,5	24,5	15,6	26,3	33,3	100,0	25,0	0,0
consumo não cíclico	3,4	15,1	10,0	7,5	7,8	5,3	0,0	0,0	25,0	0,0
financeiro e outros	3,4	1,9	17,5	18,9	29,7	21,1	66,7	0,0	25,0	0,0
materiais básicos	13,8	11,3	27,5	11,3	15,6	21,1	0,0	0,0	25,0	0,0
petróleo, gás e biocomb.	0,0	1,9	2,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0
tecnologia da informação	0,0	0,0	0,0	1,9	1,6	0,0	0,0	0,0	0,0	0,0
telecomunicações	10,3	0,0	17,5	3,8	4,7	0,0	0,0	0,0	0,0	0,0
utilidade pública	39,7	3,8	20,0	0,0	4,7	10,5	0,0	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	58	53	40	53	64	19	6	1	4	1

Tabela F10 – Classificação e critério BOVESPA, *cluster* mercado, ano 2004

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2004										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	9,7	23,3	47,1	6,7	0,0	9,5	6,9	0,0	0,0	50,0
construção e transporte	11,3	26,7	11,8	5,3	0,0	14,3	15,5	0,0	0,0	0,0
consumo cíclico	30,6	30,0	5,9	13,3	0,0	38,1	12,1	0,0	0,0	0,0
consumo não cíclico	14,5	3,3	0,0	4,0	0,0	4,8	10,3	9,1	13,6	50,0
financeiro e outros	12,9	3,3	0,0	17,3	0,0	28,6	24,1	0,0	27,3	0,0
materiais básicos	12,9	13,3	17,6	17,3	0,0	0,0	13,8	36,4	27,3	0,0
petróleo, gás e biocomb.	0,0	0,0	0,0	0,0	0,0	0,0	1,7	9,1	0,0	0,0
tecnologia da informação	1,6	0,0	0,0	1,3	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	1,6	0,0	0,0	8,0	0,0	4,8	5,2	36,4	13,6	0,0
utilidade pública	4,8	0,0	17,6	26,7	###	0,0	10,3	9,1	18,2	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	62	30	17	75	1	21	58	11	22	2

Tabela F11 – Classificação e critério BOVESPA, *cluster* índice, ano 2005

CLASSIFICAÇÃO BOVESPA –C/VAR TAMANHO, ANO 2005										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	6,7	20,3	16,7	7,6	0,0	0,0	12,9	0,0	0,0	100,0
construção e transporte	6,7	17,4	25,0	10,6	0,0	7,1	12,9	0,0	0,0	0,0
consumo cíclico	3,8	33,3	8,3	19,7	60,0	28,6	25,8	0,0	0,0	0,0
consumo não cíclico	8,6	10,1	0,0	9,1	0,0	7,1	9,7	0,0	0,0	0,0
financeiro e outros	16,2	10,1	16,7	21,2	20,0	28,6	6,5	0,0	100,0	0,0
materiais básicos	17,1	5,8	33,3	16,7	20,0	14,3	16,1	0,0	0,0	0,0
petróleo, gás e biocomb.	1,0	0,0	0,0	0,0	0,0	0,0	3,2	0,0	0,0	0,0
tecnologia da informação	0,0	0,0	0,0	0,0	0,0	7,1	6,5	0,0	0,0	0,0
telecomunicações	11,4	1,4	0,0	4,5	0,0	7,1	0,0	100,0	0,0	0,0
utilidade pública	28,6	1,4	0,0	10,6	0,0	0,0	6,5	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	105	69	12	66	5	14	31	1	2	1

Tabela F12 – Classificação e critério BOVESPA, *cluster* mercado, ano 2005

CLASSIFICAÇÃO BOVESPA –COM VARIÁVEIS DE TAMANHO, ANO 2005										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	13,7	18,2	13,1	0,0	2,7	0,0	11,1	0,0	17,1	0,0
construção e transporte	10,5	18,2	1,6	0,0	16,2	0,0	33,3	0,0	17,1	0,0
consumo cíclico	20,0	9,1	13,1	0,0	16,2	25,0	25,9	0,0	37,1	0,0
consumo não cíclico	10,5	18,2	3,3	0,0	10,8	0,0	3,7	14,3	8,6	0,0
financeiro e outros	15,8	18,2	19,7	0,0	2,7	62,5	18,5	21,4	8,6	0,0
materiais básicos	16,8	9,1	21,3	0,0	5,4	0,0	3,7	25,0	11,4	33,3
petróleo, gás e biocomb.	1,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	33,3
tecnologia da informação	1,1	0,0	1,6	0,0	2,7	0,0	0,0	0,0	0,0	0,0
telecomunicações	4,2	0,0	4,9	0,0	5,4	12,5	0,0	25,0	0,0	33,3
utilidade pública	6,3	9,1	21,3	100,0	37,8	0,0	3,7	14,3	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	95	11	61	1	37	8	27	28	35	3

Tabela F13 – Classificação e critério BOVESPA, *cluster* índice, ano 2006

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2006										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	12,2	5,5	20,9	7,7	2,1	7,7	100,0	0,0	0,0	0,0
construção e transporte	10,6	18,2	14,9	3,8	8,5	0,0	0,0	0,0	0,0	0,0

consumo cíclico	14,6	21,8	26,9	11,5	2,1	30,8	0,0	0,0	0,0	0,0
consumo não cíclico	10,6	9,1	6,0	11,5	10,6	7,7	0,0	0,0	100,0	0,0
financeiro e outros	16,3	25,5	3,0	19,2	14,9	38,5	0,0	100,0	0,0	0,0
materiais básicos	13,0	14,5	14,9	3,8	21,3	7,7	0,0	0,0	0,0	0,0
petróleo, gás e biocomb.	0,0	1,8	1,5	0,0	2,1	0,0	0,0	0,0	0,0	100,0
tecnologia da informação	2,4	0,0	1,5	7,7	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	2,4	1,8	6,0	11,5	10,6	7,7	0,0	0,0	0,0	0,0
utilidade pública	17,9	1,8	4,5	23,1	27,7	0,0	0,0	0,0	0,0	0,0
TOTAL (%)	100									
TOTAL CASOS	123	55	67	26	47	13	1	1	1	1

Tabela F14 – Classificação e critério BOVESPA, *cluster* mercado, ano 2006

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS DE TAMANHO, ANO 2006										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	14,3	15,7	16,3	33,3	6,7	4,3	0,0	10,0	0,0	0,0
construção e transporte	0,0	9,8	16,3	33,3	6,7	10,9	0,0	20,0	0,0	0,0
consumo cíclico	28,6	14,7	23,3	0,0	13,3	6,5	0,0	34,0	14,3	0,0
consumo não cíclico	0,0	10,8	7,0	16,7	8,3	13,0	8,3	8,0	14,3	0,0
financeiro e outros	0,0	18,6	20,9	0,0	16,7	4,3	41,7	10,0	57,1	0,0
materiais básicos	0,0	14,7	7,0	16,7	13,3	15,2	25,0	16,0	0,0	50,0
petróleo, gás e biocomb.	0,0	1,0	2,3	0,0	0,0	0,0	0,0	0,0	14,3	50,0
tecnologia da informação	0,0	2,9	2,3	0,0	3,3	0,0	0,0	0,0	0,0	0,0
telecomunicações	14,3	2,9	2,3	0,0	5,0	15,2	8,3	2,0	0,0	0,0
utilidade pública	42,9	8,8	2,3	0,0	26,7	30,4	16,7	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	7	102	43	6	60	46	12	50	7	2

Tabela F15 – Classificação e critério BOVESPA, *cluster* índice, ano 2007

CLASSIFICAÇÃO BOVESPA – COM VAR TAMANHO, ANO 2007										
SETORES	Cluster índice									
	1	2	3	4	5	6	7	8	9	10
bens industriais	6,3	18,4	13,6	0,0	0,0	0,0	0,0	14,3	20,0	20,0
construção e transporte	4,2	12,6	19,7	0,0	22,2	55,6	11,4	14,3	0,0	20,0
consumo cíclico	7,4	16,1	33,3	11,1	33,3	0,0	2,9	22,9	20,0	20,0
consumo não cíclico	11,6	9,2	7,6	0,0	11,1	11,1	11,4	17,1	0,0	0,0
financeiro e outros	34,7	19,5	12,1	0,0	16,7	14,8	5,7	2,9	40,0	20,0
materiais básicos	12,6	14,9	9,1	0,0	16,7	3,7	28,6	11,4	0,0	20,0
petróleo, gás e biocomb.	0,0	2,3	0,0	0,0	0,0	0,0	2,9	2,9	0,0	0,0
tecnologia da informação	0,0	1,1	0,0	0,0	0,0	11,1	0,0	8,6	0,0	0,0
telecomunicações	4,2	4,6	0,0	5,6	0,0	3,7	17,1	0,0	20,0	0,0
utilidade pública	18,9	1,1	4,5	83,3	0,0	0,0	20,0	5,7	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	95	87	66	18	18	27	35	35	5	5

Tabela F16 – Classificação e critério BOVESPA, *cluster* mercado, ano 2007

CLASSIFICAÇÃO BOVESPA – COM VARIÁVEIS TAMANHO, ANO 2007										
SETORES	Cluster mercado									
	1	2	3	4	5	6	7	8	9	10
bens industriais	6,8	13,6	17,2	20,0	5,0	6,3	0,0	0,0	0,0	0,0
construção e transporte	5,9	23,6	15,5	0,0	12,5	21,9	0,0	25,0	20,0	0,0
consumo cíclico	10,2	13,6	22,4	40,0	7,5	43,8	0,0	37,5	0,0	0,0
consumo não cíclico	8,5	11,8	10,3	0,0	17,5	6,3	7,7	0,0	0,0	0,0
financeiro e outros	23,7	14,5	12,1	0,0	7,5	18,8	38,5	37,5	60,0	0,0
materiais básicos	15,3	8,2	19,0	20,0	12,5	3,1	30,8	0,0	0,0	50,0
petróleo, gás e biocomb.	0,8	0,9	1,7	0,0	0,0	0,0	0,0	0,0	0,0	50,0
tecnologia da informação	1,7	4,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
telecomunicações	3,4	2,7	1,7	0,0	17,5	0,0	7,7	0,0	20,0	0,0
utilidade pública	23,7	6,4	0,0	20,0	20,0	0,0	15,4	0,0	0,0	0,0
TOTAL (%)	100	100	100	100	100	100	100	100	100	100
TOTAL CASOS	118	110	58	5	40	32	13	8	5	2