

Texto para Discussão

Série Economia

TD-E / 40 - 2003

**Características das Firms Inovadoras  
no Estado de São Paulo:  
Uma Análise Empírica a partir da PAEP**  
Prof. Dr. Sérgio Kannebley Júnior



**Universidade de São Paulo**  
**Faculdade de Economia, Administração e Contabilidade**  
**de Ribeirão Preto**

Reitor da Universidade de São Paulo  
Adolpho José Melfi

Diretor da FEA-RP/USP  
Marcos Cortez Campomar

Chefe do Departamento de Administração  
André Lucirton Costa

Chefe do Departamento de Contabilidade  
José Carlos Marion

Chefe do Departamento de Economia  
Rudinei Toneto Júnior

CONSELHO EDITORIAL

*Comissão de Pesquisa da FEA-RP/USP*

*Faculdade de Economia, Administração e Contabilidade  
de Ribeirão Preto-USP  
Avenida dos Bandeirantes, 3900  
14049-900 Ribeirão Preto – SP*

A série **TEXTO PARA DISCUSSÃO** tem como objetivo divulgar i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) outros trabalhos considerados de relevância dadas as linhas de pesquisa da instituição. *A série foi subdividida em função das principais áreas de atuação da FEA-RP/USP: Economia, Administração e Contabilidade. Veja o site da CPq na Home Page da FEA-RP: [www.fearp.usp.br](http://www.fearp.usp.br). Informações: e-mail: [cpq@fearp.usp.br](mailto:cpq@fearp.usp.br).*

# Características das Firms Inovadoras no Estado de São Paulo: Uma Análise Empírica a partir da PAEP<sup>?</sup>

Sérgio Kannebley Júnior<sup>??</sup>

## Resumo

Este artigo reporta um esforço de caracterização das empresas inovadoras do estado de São Paulo. Esta caracterização é realizada com base em informações financeiras, patrimoniais e de recursos humanos das empresas constantes na PAEP (Pesquisa de Atividade Econômica Paulista). Para isso foram empregados procedimentos estatísticos não-paramétricos, que informaram, em ordem decrescente, que os três principais fatores distintivos entre empresas inovadoras e não inovadoras são: a orientação exportadora, o tamanho da empresa e a variação interindustrial. De modo menos incisivo, também contribuem para a distinção entre empresas inovadoras e não inovadoras o grau de concentração do mercado, a origem do capital, a qualificação do trabalho e grau de apropriabilidade.

**Palavras-chave:** inovação; economia industrial; métodos não-paramétricos

## Abstract

This article reports an effort to characterize São Paulo's innovative firms. This characterization is done based on PAEP (Paulista Economic Activity Research) information. It was employed statistical non-parametrical procedures, which informs that the three distinctive main factors are the export orientation, the size of the firm, and the industrial sector effect. Besides, it was showed that the market concentration, the foreign capital origin, the labor force qualification and appropriability conditions are important to distinct between innovators and non innovators firms, but in a less incisive way.

**Keywords:** innovation; industrial economics; nonparametric methods

*JEL classification:* O30; L00; C14

---

<sup>?</sup> Agradecimentos à Fundação SEADE, especificamente a Noeli Pereira, pela colaboração e disponibilização das informações para a realização desse trabalho e à assistência dos auxiliares de pesquisa Harry Richard Hamming e Paulo Rogério Rodrigues Maduro Jr. Qualquer erro, ou omissão, é de responsabilidade exclusiva do autor.

<sup>??</sup> Professor FEA-RP/USP.

## Introdução

A literatura empírica sobre os fatores que influenciam, ou que caracterizam, a atividade inovativa das empresas está bastante desenvolvida. As resenhas de Cohen e Levin (1989) e Cohen (1995) apresentam diversos argumentos teóricos e resultados obtidos em testes empíricos para este tema. Usualmente a realização de atividades inovativas, ou a intensidade com que são realizadas, é explicada em razão das características das firmas, das estruturas de mercados, da variação interindustrial, das condições de apropriabilidade e de demanda.

A grande maioria dos trabalhos expostos na literatura refere-se ao exame dos determinantes de atividades de Pesquisa e Desenvolvimento (P&D) das firmas tecnologicamente ativas em países industrializados. Kumar e Siddhartan (1997) apresentam resultados da literatura empírica que demonstram que essa caracterização de atividade inovativa não é representativa para as firmas de países em desenvolvimento. Esses resultados destacam a importância da compra de P&D externo, bem como a realização de progresso técnico por meio da importação de bens de capitais. Esses resultados foram confirmados para o caso brasileiro recentemente por meio das pesquisas de atividade tecnológica contidas na PAEP (Pesquisa de Atividade Econômica Paulista) e na PINTEC (Pesquisa Industrial de Inovação e Tecnologia).

As informações constantes na PAEP estão em concordância com essas evidências, já que demonstram que aproximadamente 78,5 % das empresas que declararam não terem realizado inovações de produto e/ou processo no período de 1994 a 1996, sejam elas incrementais ou significativas, também não efetuaram importações de bens de capital estrangeiros. Por outro lado, aproximadamente 57,4 % das empresas que responderam terem realizado inovações de produto e/ou processo importaram de bens de capital no mesmo período. Também de acordo com as informações da PAEP, os gastos com atividades de P&D interno corresponderam a apenas 16,5% do total de gastos em atividades inovativas das firmas paulistas em 1996<sup>1</sup>.

Sabendo dessas características distintas do processo inovativo em países em desenvolvimento, e particularmente no caso brasileiro, este trabalho se propõe a analisar as características das firmas inovativas, e não apenas aquelas realizadoras de atividades de P&D. Será adotada aqui uma caracterização ampla para a atividade inovativa, ficando aberto o caminho para trabalhos futuros que desejem trabalhar com categorias mais restritas de inovação. A determinação das características das firmas inovadoras em contraposição às firmas não inovadoras deve fornecer subsídios à análise de indicadores de atividade inovativa e tecnológica e também para a formulação de políticas. É essa a proposta final desse trabalho e para isso são utilizadas informações constantes na PAEP para o período de 1994 a 1996. As variáveis selecionadas para análise referem-se às diversas características financeiras, patrimoniais e de recursos humanos das empresas.

O trabalho é composto, além dessa introdução, de mais quatro seções. Na primeira seção são descritas e justificadas as variáveis a serem utilizadas nos modelos estatísticos. Na segunda é realizada uma breve análise das informações constantes na pesquisa. A seguir são apresentados a metodologia de árvores de classificação QUEST e os resultados fornecidos pela estimação das árvores. Por fim, na última seção são tecidas algumas considerações finais sobre a análise realizada.

---

<sup>1</sup> Essas informações estão em conformidade com aquelas fornecidas pela PINTEC ao relatar que aproximadamente 77 % das empresas inovadoras atribuíram alta ou média importância para a atividade inovativa de aquisição de máquinas e equipamentos, correspondendo a mais de 50 % dos gastos em atividades inovativas. Também segundo a PINTEC, as atividades internas de P&D foram classificadas como de média ou alta importância por apenas 34,1% das empresas brasileiras, sendo que os gastos relativos a essa atividade correspondem a 16,7 % do total dos gastos em atividades inovativas das empresas brasileiras no ano de 2000.

## **1. Definição das Variáveis**

Baseado na literatura teórica e empírica sobre o tema, são utilizadas oito variáveis explicativas que buscam retratar as empresas segundo suas características financeiras, patrimoniais e de recursos humanos.

### **1.1. Tamanho da Firma**

Essa variável é usualmente associada à hipótese Schumpeteriana de relação positiva, mas não necessariamente linear ou contínua, entre atividade inovativa e tamanho da firma. Vários argumentos são tecidos para se justificar o efeito positivo do tamanho da firma sobre a atividade inovativa, entre os quais: i) imperfeição do mercado de crédito (grandes firmas teriam mais facilidade para financiar projetos de P&D de risco visto que tamanho é correlacionado com a disponibilidade e estabilidade de fundos internos para investimento); ii) existência de economias de escala na função de P&D propriamente, considerando que os retornos de P&D são mais altos quando o inovador tem um maior volume de vendas sobre o qual consegue diluir mais facilmente os custos fixos da inovação, particularmente nos casos de inovação de processo; iii) P&D é dito ser mais produtivo nas grandes empresas resultado da complementaridade entre P&D e outras atividades não-manufatureiras (marketing e planejamento) existente em maior grau nestas empresas; iv) firmas maiores, diversificadas providenciam economias de escopo ou reduzem o risco associado à inovação.

Por outro lado, existem contra-argumentos relacionados à perda de eficiência da atividade de P&D. Argumenta-se que quando a firma torna-se muito grande, a eficiência em P&D é prejudicada em função da perda de controle gerencial ou, alternativamente, devido ao aumento excessivo da burocracia que acaba por desviar a atenção dos cientistas. Outro ponto a ser destacado é a diminuição da habilidade dos cientistas e empresários em capturar os benefícios de seus esforços individuais, podendo também diminuir os incentivos destes na busca da inovação.

As evidências empíricas ainda que favoreçam a hipótese Schumpeteriana, apresentam resultados ambíguos em razão da definição variável dependente. Em alguns estudos a variável dependente é o total de gastos em P&D, enquanto que em outros se trata da intensidade de P&D, dada pela razão entre o total de gastos de P&D e o total de vendas da empresa. Adicionalmente, Cohen (1995) relata problemas relacionados ao viés de seleção das amostras, compostas em sua maioria por grandes empresas e inovadoras e a desconsideração de fatores explicativos específicos à indústria e à empresa.

### **1.2. Estrutura de Mercado**

Também de acordo com a tradição Schumpeteriana postula-se uma relação positiva entre estruturas de mercado mais concentradas e a atividade inovativa, principalmente as atividades de P&D. O argumento básico é que a estrutura de mercado perfeitamente competitiva não incentiva atividades de P&D em razão da baixa apropriabilidade dos resultados dessas atividades, não favorecendo à obtenção de lucros extraordinários pelas empresas inovadoras. Esses lucros extraordinários seriam, por conseqüência, a fonte de financiamento para novas atividades, gerando um processo de retroalimentação da estrutura de mercado.

A verificação empírica desse postulado usualmente é conduzida por meio da estimação da relação entre um índice de concentração de vendas no mercado e o comportamento inovativo da empresa. Conforme foi argumentado, esta relação não está livre da ambigüidade no que diz respeito ao seu sentido causal, reconhecendo-se a possibilidade de

simultaneidade entre inovação e estrutura de mercado. Também são ambíguas as evidências empíricas produzidas pela literatura sobre o tema.

### **1.3. Orientação Exportadora**

Deve-se esperar que a exposição à competição do mercado internacional incentive a firma a investir em atividades inovativas. Isso seria feito no intuito de atender à demanda mais exigente e as pressões competitivas mais intensas enfrentadas no mercado internacional. Argumenta-se também que as exportações, ao ampliarem os mercados das firmas, aumentam os retornos da atividade inovativa na medida em que diluem seus custos. Braga e Willmore (1991) entendem que provavelmente a primeira linha de argumentação é a que melhor se adequa ao caso brasileiro em razão do Brasil ser um país em desenvolvimento. Em seu estudo encontram uma relação positiva e estatisticamente significativa entre a probabilidade de existência de atividade tecnológica e a orientação exportadora das empresas.

Também é importante lembrar que essa relação pode estar sujeita a um problema de simultaneidade, na medida em que a atividade tecnológica também pode aprimorar a capacidade competitiva, permitindo que a empresa seja capaz de enfrentar os mercados internacionais mais competitivos.

### **1.4. Aprendizado Acumulado**

Considerando os aspectos específicos à firma, uma variável representativa desses aspectos é o tempo de constituição da empresa. É possível se argumentar que empresas com maior tempo de constituição acumularam ao longo do tempo um aprendizado em atividades inovativas que reduzem o custo do investimento em novas atividades inovativas. Desse modo, é postulada uma relação positiva entre o tempo de constituição da empresa e o desempenho de atividades inovativas, baseada no argumento da acumulação de aprendizado ou conhecimento específico à atividade realizada, além da institucionalização da atividade de pesquisa em empresas grandes já estabelecidas.

Henderson (1993) pondera esta hipótese ao distinguir o comportamento das empresas novas (entrantes) e já estabelecidas frente a diferentes formas de inovação. Segundo a autora a hipótese apresentada acima se aplica mais adequadamente ao caso das inovações incrementais, na medida em que essa forma de inovação baseia-se mais fortemente no conhecimento e nas capacidades previamente adquiridas. Entretanto, frente à possibilidade de inovações radicais o extensivo conhecimento em uma tecnologia poderia ser uma desvantagem em potencial. Isso ocorreria se essas inovações fossem fundadas em princípios científicos e/ou tecnológicos diversos dos até então utilizados. Nesse caso, as firmas entrantes, em razão das diferenças nas capacidades informacionais, estariam em vantagem, realizando um esforço tecnológico superior ao das empresas já estabelecidas no mercado.

### **1.5. Qualificação do Trabalho**

Ainda que não diretamente relacionado à atividade inovativa, a qualificação do trabalhador é um fator relevante na capacidade de execução de tarefas de maior grau de complexidade existentes em firmas tecnologicamente mais avançadas. A existência de trabalhadores qualificados é complementar à execução de atividades inovativas. Sem os primeiros, muitos dos esforços empreendidos no desenvolvimento de novos processos e/ou produtos poderiam estar comprometidos pela perda de eficiência e qualidade da produção. Sendo assim, espera-se que exista uma relação positiva entre o grau de qualificação do fator trabalho e a execução de atividades inovativas.

## 1.6. Origem do Capital Controlador

As empresas multinacionais são detentoras de um conjunto de ativos intangíveis, como marcas internacionalmente reconhecidas, acesso cativo à tecnologia e capacitações gerenciais e organizacionais. Em razão disso é possível que a estratégia competitiva da empresa multinacional em um país local seja uma estratégia de rivalidade não baseada em preços. Essa estratégia de rivalidade não baseada em preços estaria apoiada em gastos de marketing, controle de qualidade e desenvolvimento de produtos, além de um conjunto de serviços ao consumidor.

De acordo com esse arcabouço competitivo é possível se conceber uma relação positiva entre a execução de atividades tecnológicas e a natureza estrangeira do capital. Esta relação deve ser mais forte em atividades relacionadas ao desenvolvimento e/ou adaptação de produtos ao mercado doméstico. É também importante notar que Braga e Willmore (1991) encontraram evidências de um efeito positivo da propriedade estrangeira do capital sobre a probabilidade das firmas possuírem um programa sistemático de desenvolvimento de produtos na indústria brasileira.

## 1.7. Apropriabilidade

A que extensão os novos conhecimentos podem ser transmitidos a relativamente baixos custos do seu criador para outros competidores e, particularmente, a que extensão esse conhecimento, incorporado em novos processos e produtos, pode ser copiado ou imitado a relativamente baixos custos, determinam se as recompensas apropriáveis são, ou não, suficientes para justificar o esforço inovativo.

Considerando que a efetividade das patentes difere sensivelmente entre as indústrias, a possibilidade de internalização do conhecimento, ao invés do licenciamento, pode ser uma estratégia recompensadora para as empresas que investem em pesquisa ou mudanças tecnológicas<sup>2</sup>. Sendo assim, uma empresa com maior proporção do valor adicionado de suas vendas deve ter uma probabilidade maior de investir em atividades de P&D. Este argumento deve ser mais apropriado ao caso de grandes empresas, já que este processo pode ser extremamente custoso ou inviável para pequenas empresas. Ainda assim, deve-se esperar uma relação positiva entre a realização de atividades inovativas e uma variável dada pela razão entre o valor adicionado e a receita líquida obtida pela empresa.

## 1.8. Efeitos Setoriais

Usualmente verifica-se uma diferença interindustrial no grau em que as indústrias se engajam em atividades inovativas. Os argumentos que tentam explicar essas diferenças estão na maior parte relacionados às diferentes oportunidades de progresso tecnológico que as indústrias percebem.

Em termos empíricos, desconsiderar a capacidade explicativa das diferenças interindustriais constitui um problema de viés por omissão de variável relevante. Este problema é freqüentemente encontrado em estudos que procuram analisar as relações entre tamanho, grau de concentração e P&D, levando a estimativas enviesadas do efeito do tamanho ou da estrutura de mercado sobre a atividade inovativa. Nesse trabalho buscar-se-á capturar estes efeitos por meio da categorização dos setores segundo as oportunidades

---

<sup>2</sup> Segundo Cohen (1995), a principal razão citada para limitação da efetividade das patentes é capacidade de os competidores inventarem “legalmente em torno das patentes”.

tecnológicas ou de acordo com a intensidade do fator de produção do setor industrial <sup>3</sup>. A definição dos setores segundo suas oportunidades tecnológicas foi realizada de modo semelhante àquela feita por Quadros *et. alii.* (2001)<sup>4</sup>.

## **2. Análise Descritiva das Características das Empresas Inovadoras do Estado de São Paulo – PAEP (1996)**

### **2.1. Base de Dados**

A PAEP foi realizada pela Fundação Sistema Estadual de Análise de Dados (FSeade). É uma pesquisa de caráter amostral tendo 1996 como ano base para a sua realização e o Estado de São Paulo como o universo de análise. A base expandida contém informações de 41.658 empresas industriais e fornece um banco de dados para a elaboração de pesquisas e análises acerca das transformações técnico-produtivas em curso na economia paulista. Produz não só dados de mensuração da atividade econômica – como, por exemplo, número de unidades, valor da produção, pessoal ocupado – mas também indicadores que permitem avaliar a extensão dos recentes processos de reestruturação e seus impactos nos diferentes setores da economia paulista. A disponibilização de microdados de empresas no ramo da indústria permite a mensuração da estrutura econômica paulista, inclusive no que tange às atividades inovativas das empresas paulistas.

Antes de iniciar a análise propriamente dita é importante destacar algumas limitações e vantagens dessa base de dados. Uma primeira limitação deve-se ao fato da pesquisa ser em um único ponto do tempo, isto é, uma “*cross-section*”. Isto não permite que sejam capturadas as relações dinâmicas da atividade inovativa postuladas pela teoria econômica. Também é um fator prejudicial à análise o fato das questões sobre a condução de atividades inovativas se referirem ao período de 1994 a 1996, enquanto que as demais variáveis que são utilizadas para a caracterização das empresas se referirem ao ano de 1996<sup>5</sup>. A grande vantagem que o trabalho com a base expandida fornece é a disponibilização de microdados das empresas que combinem as diversas informações de suas características com aquelas sobre suas atividades inovativas. Adicionalmente, a disponibilidade de informações para todos os tamanhos de empresas evita a produção de resultados *viesados* em direção às grandes empresas industriais.

### **2.2. Análise Descritiva**

Na tabela 1 abaixo são apresentadas informações sobre o número de empresas que realizaram ou não alguma atividade inovativa para a indústria de transformação <sup>6</sup>. Essa tabela também contém informações da taxa de inovação do setor discriminadas segundo a forma de

---

<sup>3</sup> A taxonomia de oportunidades tecnológicas foi criada por Pavitt (1984). Esta taxionomia está relacionada ao aparecimento e ao desenvolvimento dos paradigmas tecnológicos entre os diversos setores industriais. Segundo Dosi (1988) o aparecimento de novos paradigmas é desigualmente distribuído entre os setores, sendo também desigualmente distribuídas as dificuldades técnicas para o avanço da eficiência da produção e da performance do produto, bem como a competência tecnológica para inovar. No entanto, argumenta que essa distribuição não é aleatória, mas sim dependente da natureza tecnológica dos setores, da distância que a tecnologia empregada está do “centro revolucionário do paradigma” e da base de conhecimento subjacente à inovação em cada setor.

<sup>4</sup> Essa classificação foi realizada com base no esforço tecnológico realizado pelas empresas industriais de São Paulo relacionado à distribuição do volume de P&D realizado entre os setores. Reclassificados segundo a divisão CNAE, os setores com altas oportunidades tecnológicas são as divisões 11, 30, 32, 33 e 35, com médias oportunidades as divisões 28, 29, 31 e 34, os de média-baixa oportunidade as divisões 21, 23, 24, 25, 26 e 27, e por fim os com baixas oportunidades, as divisões 10, 13, 14, 15, 16, 17, 18, 19, 20 e 22.

<sup>5</sup> Nesse caso é necessário se admitir a existência de uma relativa inércia na estrutura das empresas e setorial que diminua o poder transformador das atividades inovativas para aquelas empresas que as realizaram no início do período.

<sup>6</sup> As informações referentes à indústria extrativa não foram apresentadas aqui em razão de sigilo estatístico.

inovação produzida no período, além de um índice de concentração setorial para o Estado de São Paulo para o ano de 1996. As empresas industriais paulistas que produziram inovações de processo ou produto, incrementais ou significativas, no período de 1994 a 1996 determinaram uma taxa média de inovação para o estado em torno de 31 % <sup>7</sup>. Os setores com as cinco maiores taxas de inovação foram os setores de Fabricação de Máquinas de Escritório e Equipamentos de Informática (30), Fabricação de Equipamentos de Instrumentação Médico-Hospitalares (33), Fabricação de Material Eletrônico e Aparelhos de Comunicações (32), Fabricação de Outros Equipamentos de Transporte (35), Fabricação de Produtos Químicos (24). Já os setores com as cinco menores taxas de inovação foram os setores de Fabricação de Celulose, Papel e Produtos de Papel (21), Fabricação de Couros e Calçados (19), Edição, Impressão e Reprodução de Gravações (22), Confecção de Artigos de Vestuário e Acessórios (18) e de Fabricações de Produtos de Madeira (20). Uma análise atenta também permite perceber a existência de uma relação positiva entre a medida de concentração industrial no Estado e a taxa de inovação setorial <sup>8</sup>. É possível perceber que existe na maioria dos setores a prevalência de empresas inovadoras em processo e produto. Em média 63 % da taxa média de inovação é na forma de inovação de produto e processo, enquanto que apenas 24 % correspondem a inovações de produto e 13 % a inovações de processo.

---

<sup>7</sup> Estritamente, na PAEP define-se a realização de INOVAÇÃO DE PRODUTO (NATUREZA INCREMENTAL) como aquela correspondente a um substancial aperfeiçoamento de um produto previamente existente. Realização de INOVAÇÃO DE PRODUTO (NATUREZA SIGNIFICATIVA) aquela correspondente à introdução de um produto inteiramente novo, com características tecnológicas ou de uso e finalidade que o distinguem daqueles produzidos até então e a realização de INOVAÇÃO DE PROCESSO como correspondente à incorporação de um novo processo de produção ou às modificações tecnológicas em processos já adotados.

<sup>8</sup> O coeficiente de correlação apurado para essa relação é de 0,11. É importante também lembrar que essa medida não é a mais apropriada para refletir a relação teórica estabelecida acima entre estrutura de mercado e atividade inovativa, já que não corresponde à estrutura nacional do setor, nem captura a contestabilidade do setor realizada pela possibilidade de comércio internacional. Porém, considerando a representatividade da indústria paulista sobre a indústria nacional não deixa de ser uma medida referencial importante.

**Tabela 1**  
**Empresas Não Inovadoras e Inovadoras segundo a Divisão CNAE**

DIVISÃO CNAE	NÃO INOVA	INOVA	PROCESSO (%)	PRODUTO (%)	PROD. e PROC. (%)	TAXA de INOVAÇÃO	CR4
15	1060	383	16	21	62	27	0.2
17	1489	458	11	27	62	24	-
18	4623	765	11	25	63	14	0.2
19	1219	327	13	31	56	21	0.1
20	802	82	6	64	30	9	0.3
21	691	215	17	29	54	24	0.4
22	1910	519	20	8	72	21	0.3
23	48	19	41	0	59	29	0.3
24	940	651	11	19	70	41	0.3
25	1654	827	13	31	56	33	0.2
26	1099	421	9	37	54	28	0.3
27	784	335	16	12	72	30	0.2
28	2569	1027	23	20	57	29	0.4
29	1841	955	11	21	69	34	0.1
30	20	52	6	27	67	72	0.2
31	823	327	10	20	69	28	0.6
32	261	201	10	19	71	43	0.3
33	317	248	4	20	76	44	0.6
34	599	341	14	19	68	36	0.5
35	64	44	9	30	61	41	0.7
36	2374	884	9	24	67	27	0.6

CR4 é a medida proporcional das receitas das quatro maiores empresas industriais da divisão.

Na tabela 2 abaixo são apresentadas informações sobre o número e o percentual de empresas inovadoras e não inovadoras segundo suas faixas de tamanho e origem do capital <sup>9</sup>. Como é possível observar o percentual de empresas não inovadoras é decrescente à medida que aumenta a faixa de tamanho independentemente da origem do capital da empresa. Ainda que o inverso não ocorra de modo exato para as empresas inovadoras, é possível perceber que o percentual de empresas tende a ser maior à medida que aumenta o tamanho das empresas. A exceção a essa tipificação é a categoria de empresas nacionais, em que de modo similar às empresas não inovadoras nacionais, o percentual de empresas inovadoras é decrescente à medida que aumenta a faixa de tamanho. No caso das empresas estrangeiras e mistas este percentual é menor na faixa de tamanho de micro empresas, atingindo um máximo na faixa de médias empresas e declinando na faixa de grandes empresas. Ainda assim, é importante notar que o número de empresas inovadoras nacionais é bastante superior ao número de empresas estrangeiras e mistas inovadoras.

<sup>9</sup> As faixas de tamanho são definidas segundo o número total de pessoas ocupadas. É considerada uma micro-empresa aquela que possui de 0 a 19 empregados, pequena empresa aquela que possui de 20 a 99 empregados, média empresa aquela com 100 a 499 empregados e grande empresa aquelas com mais de 500 empregados.

**Tabela 2**  
**Empresas Inovadoras segundo a Faixa de Tamanho e Origem do Capital**

	NACIONAL		ESTRANGEIRA		MISTO	
	número	%	número	%	número	%
NAO INOVA						
MICRO	16822	67,7	58	24,9	37	40,0
PEQUENA	6759	27,2	76	32,3	33	36,3
MÉDIA	1133	4,6	75	32,1	20	21,9
GRANDE	138	0,6	25	10,7	2	1,8
INOVA						
MICRO	3900	45,2	29	12,3	3	2,8
PEQUENA	3166	36,7	70	30,0	46	50,7
MÉDIA	1303	15,1	113	48,2	56	61,8
GRANDE	264	3,1	94	40,0	35	37,9

Na tabela 3 abaixo é apresentada uma extensão da tabulação realizada na tabela 2 acima. Nessa tabulação as empresas são classificadas primeiramente de acordo com a origem do seu capital em contraposição ao fato de terem realizado ou não atividade inovativa. Posteriormente as firmas são reclassificadas de acordo com suas faixas de tamanho em contraposição ao fato de exportarem ou não. Essas informações demonstram, como de conhecimento geral, que independentemente do fato dessas empresas inovarem ou não, bem como da origem de seu capital, que existe uma relação positiva entre a orientação exportadora e o tamanho da empresa. No entanto, é importante destacar que no caso das empresas exportadoras estrangeiras inovadoras essa relação é mais tênue do que nas exportadoras nacionais. É possível também perceber a redução do percentual de empresas não exportadoras inovadoras, em contraposição às empresas não inovadoras e não exportadoras. Assim, quando se compara a proporção entre empresas exportadoras e não exportadoras, considerando o fato de serem inovadoras ou não, esta proporção é favorável ao conjunto de empresas inovadoras, sendo que as maiores razões são obtidas no caso das empresas de capital estrangeiro ou misto<sup>10</sup>. A única exceção a essa evidência é verificada na faixa das grandes empresas de capital estrangeiro.

**Tabela 3**  
**Percentual de Empresas Inovadoras segundo a Orientação Exportadora**

	NÃO INOVA			INOVA		
	Não-Exporta (A)	Exporta (B)	(B/A)*100	Não-Exporta (C)	Exporta (D)	(D/C)*100
NACIONAL						
MICRO	98,12	1,88	1,91	95,85	4,15	4,33
PEQUENA	92,54	7,46	8,06	85,39	14,61	17,10
MÉDIA	69,06	30,94	44,80	55,24	44,76	81,03
GRANDE	35,14	64,86	184,57	24,14	75,86	314,23
ESTRANGEIRO e MISTO						
MICRO	97,43	2,57	2,64	35,91	64,09	178,51
PEQUENA	43,86	56,14	128,01	27,06	72,94	269,49
MÉDIA	26,94	73,06	271,22	24,29	75,71	311,62
GRANDE	0,00	100,00	n.d.	13,45	86,55	643,24

Na tabela 4 abaixo são apresentadas informações relativas ao ano de constituição da empresa, classificadas segundo suas faixas de tamanho para empresas inovadoras ou não. Como era de se esperar as empresas maiores são as mais antigas, independentemente do fato

<sup>10</sup> Ou seja, a diferença entre a coluna (D) e a coluna (B).

de terem promovido atividades inovativas ao longo do período de 1994 a 1996. Em termos gerais, as estatísticas descritivas fornecidas não destacam diferenças sensíveis entre a idade das empresas inovadoras e não inovadoras, a não ser para o caso das micro-empresas. As mesmas indicam que para a faixa de tamanho das micro-empresas, as empresas inovadoras têm data de constituição mais recente. Particularmente é interessante notar que a maior diferença é aquela correspondente ao valor modal para a faixa das grandes empresas. Nessa faixa de tamanho, ainda que o valor da média indique as empresas inovadoras são mais antigas que as não inovadoras, a informação modal revela o contrário.

**Tabela 4**  
**Estatísticas Descritivas – Ano de Constituição da Empresa**

ESTATÍSTICA	FAIXA DE TAMANHO			
	MICRO	PEQUENA	MÉDIA	GRANDE
<b>NÃO INOVA</b>				
Média	1983	1979	1968	1963
Mediana	1986	1982	1973	1961
Moda	1986	1986	1974	1959
Desvio Padrão	10	13	64	18
<b>INOVA</b>				
Média	1985	1978	1969	1960
Mediana	1988	1982	1970	1961
Moda	1988	1984	1974	1973
Desvio Padrão	9	16	16	22

Na tabela 5 abaixo são apresentadas medidas de tendência central para a média salarial do pessoal ocupado diretamente ligado à atividade produtiva em moeda nacional para o ano de 1996. De modo similar às tabulações anteriores essas informações são classificadas segundo as faixas de tamanho das empresas e de acordo com sua origem do capital. Inicialmente é notória a evidência de que tanto a média, quanto a mediana da média salarial mensal das firmas nacionais sejam inferiores ao das firmas estrangeiras e mistas, independentemente do fato de serem inovadoras ou não. A única exceção a essa evidência encontra-se no valor da mediana da faixa de microempresas não inovadoras<sup>11</sup>. Na comparação entre empresas inovadoras e não inovadoras, em termos médios, existe uma diferença em favor da média salarial das empresas inovadoras. Particularmente na faixa de micros e pequenas empresas predominam as evidências de médias salariais superiores para as empresas inovadoras independentemente da origem de capital. Já para as faixas de médias e grandes empresas, enquanto que as empresas nacionais inovadoras apresentam médias salariais superiores ao das empresas nacionais não inovadoras, o inverso se verifica para o valor médio das empresas de capital estrangeiro e misto nas faixas de médias e grandes empresas. No entanto, para o valor da mediana essa diferença volta a ser favorável às empresas inovadoras.

<sup>11</sup> Ainda considerando como fator distintivo a origem do capital, relativamente as maiores disparidades salariais encontram-se no conjunto de empresas inovadoras nas faixas de micro, pequena empresas (nacional com relação à estrangeira) e na faixa de pequena empresa não inovadoras (nacional com relação à estrangeira).

**Tabela 5**  
**Estatísticas Descritivas – Média Salarial em Reais para 1996**

			NACIONAL	ESTRANGEIRA	MISTO
NAO INOVA	MICRO	média	347,0	377,3	562,3
		mediana	294,8	244,8	389,5
	PEQUENA	média	426,0	1532,2	826,1
		mediana	352,6	1180,6	703,9
	MÉDIA	média	573,0	1708,7	1296,6
		mediana	501,4	911,5	992,4
	GRANDE	média	723,4	1207,5	1072,6
		mediana	588,6	1025,5	n.d.
INOVA	MICRO	média	380,8	975,2	838,9
		mediana	312,5	1271,1	839,6
	PEQUENA	média	487,9	1473,8	1175,2
		mediana	417,5	1273,5	1013,6
	MÉDIA	média	666,8	1235,6	1077,3
		mediana	562,8	1084,6	1009,7
	GRANDE	média	818,1	1169,5	965,1
		mediana	681,5	1123,2	719,0

A tabela 6 apresentada abaixo traz informações sobre a razão entre o valor adicionado e a receita líquida da empresa, discriminadas segundo as faixas de tamanho das empresas. Como é possível observar, os valores apresentados na tabela diferem sensivelmente apenas nas faixas de micro e pequena empresas, sendo estas diferenças a favor das empresas não inovadoras.

**Tabela 6**  
**Estatísticas Descritivas – Valor Adicionado/Receita Líquida**

	NÃO INOVA		INOVA	
	Média	Mediana	Média	Mediana
MICRO	0,47	0,56	0,40	0,53
PEQUENA	0,48	0,52	0,35	0,51
MÉDIA	0,48	0,51	0,49	0,52
GRANDE	0,46	0,47	0,43	0,46

Em suma, o que as evidências apresentadas na análise descritiva acima sugerem é que existe uma diferenciação mais acentuada entre empresas inovadoras e não inovadoras para o período de 1994 a 1996, quando se tomam como fatores distintivos o tamanho da empresa, o efeito setorial, sua orientação exportadora, a média salarial e a origem do capital. No que tange à origem do capital essas evidências não são plenamente conclusivas em razão das ambigüidades apresentadas nas relações entre tamanho e origem do capital, e média salarial e origem do capital. Foi possível também observar que o ano de constituição da empresa, bem como a relação entre o valor adicionado e a receita líquida das empresas não são fatores que favorecem a distinção entre empresas inovadoras e não inovadoras. Essas considerações preliminares serão examinadas a seguir por meio de análise estatística.

### 3. Árvores de Classificação para Empresas Inovadoras/Não Inovadoras

O objetivo dessa seção é produzir evidências empíricas capazes de auxiliar na classificação das empresas inovadoras, em contraposição às não inovadoras, baseadas no conjunto de variáveis apresentadas na seção 1 acima. A seguir é apresentada, brevemente, a metodologia de árvores de classificação utilizada nesse trabalho. Posteriormente são apresentados e discutidos seus resultados.

#### 3.1. Metodologia

Árvores de classificação são métodos relativamente novos empregados na exploração de relações entre dados em problemas de classificação. Uma árvore de classificação é uma regra para predição da classe de um objeto (variável dependente) a partir dos valores de suas variáveis explicativas. São procedimentos estatísticos não paramétricos, baseados em algoritmos de busca exaustiva. Os resultados apresentados são estruturas hierarquizadas e flexíveis que permitem a observação de distintas relações entre a variável dependente e diversos subconjuntos de variáveis explicativas. As metodologias mais populares são a CHAID (Chi-Square Automatic Iterated Detection) proposta por Kass (1980), a C&RT (Classification and Regression Tree), de Breiman *et. alii* (1984) e a QUEST (Quick, Unbiased, Efficient, Statistical Tree) proposta por Loh e Shih (1997).

O diagrama abaixo ilustra uma árvore de partição binária com 3 níveis de classificação. Ela é composta por nós. Desde o nó raiz ao nó final, existem diversos nós que formam os ramos de classificação. O nó raiz contém todas as informações da amostra e representa a variável dependente  $Y$ , com o número de observações para cada uma das duas categorias (0 e 1 no caso) ou classes. À medida que a árvore se expande, os dados são ramificados em subconjuntos mutuamente exclusivos. Esses subconjuntos são ramos da árvore. No nível seguinte da árvore a amostra é dividida segundo o melhor preditor de  $Y$ , a variável explicativa  $X_1$ . O nó 1 apresenta o subconjunto de informações de  $Y$ , classificada segundo uma regra do tipo  $X_1 < c$ , enquanto que o nó 2 classifica as informações de  $Y$  segundo a regra  $X_1 = c$ , em que  $c$  é um número ou categoria qualquer correspondente à variável  $X_1$ . Esses nós 1 e 2 contêm informações sobre as frequências da variável  $Y$  relacionados a cada subgrupo definido segundo as combinações  $(Y, X_1, <c)$  e  $(Y, X_1 = c)$ . Os nós 1 e 2 são os pais dos nós 3,4,5 e 6. Para o nó 1 o melhor preditor é a variável  $X_2$ , enquanto que para o nó 2 o melhor preditor é a variável  $X_3$ . Esses nós filhos, obtidos a partir da divisão segundo os preditores  $X_2$  e  $X_3$ , contêm informações sobre as frequências da variável  $Y$  relacionadas a cada subgrupo definido segundo as combinações  $(Y, X_1, X_2)$  e  $(Y, X_1, X_3)$ . Estes últimos nós por se encontrarem ao final da árvore também são chamados de nós terminais.

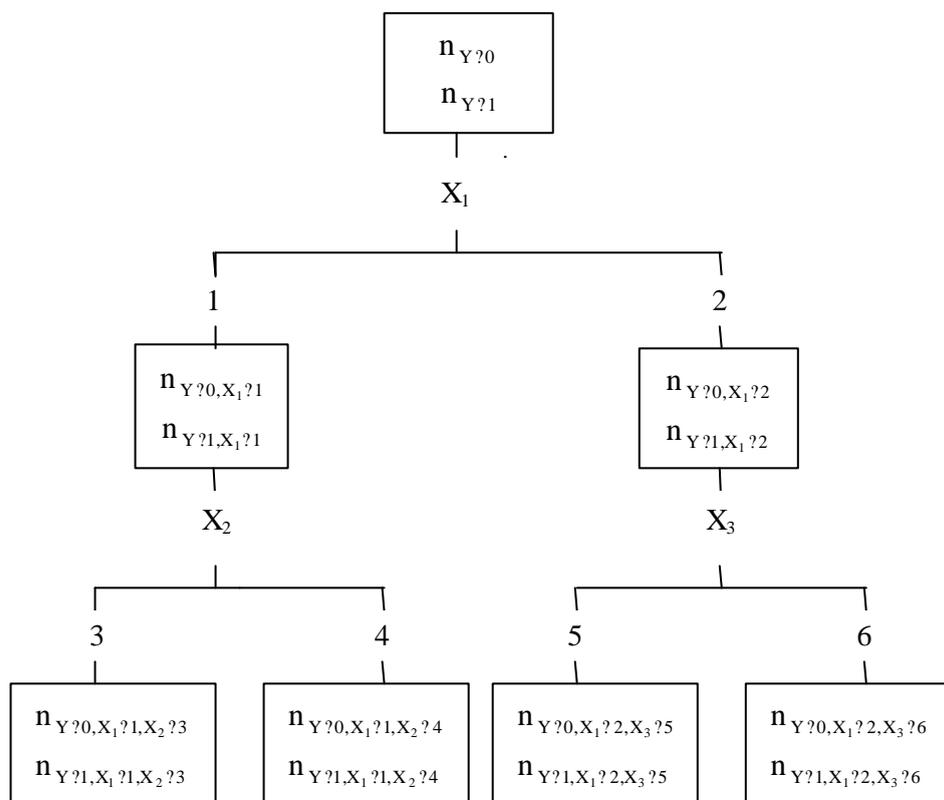
Nessa análise será empregado o método de estimação QUEST. Essa metodologia permite a partição binária dos nós, evitando a extrema complexidade da análise, possibilita a inserção de custos diferenciados de classificação para as categorias e a realização da redução da árvore (*pruning*) baseado em uma relação benefício/custo. Resumidamente, a metodologia proposta por Loh e Shih (1997) trabalha com a partição dos nós e a seleção das variáveis explicativas separadamente. O algoritmo de partição é dependente do fato da variável preditora ser ordinal, contínua ou nominal<sup>12</sup>. Nos dois primeiros casos, com uma

---

<sup>12</sup> A título de esclarecimento nesse trabalho define-se como variável categórica aquela que não apresenta medida contínua. As variáveis categóricas podem ser divididas em nominais ou ordinais. Para as primeiras, as categorias são estabelecidas segundo alguma qualificação da variável, enquanto que para as segundas, as categorias obedecem a algum ordenamento. Adicionalmente, é importante esclarecer que a metodologia QUEST somente é aplicável quando a variável dependente é categórica. No entanto, as variáveis preditoras podem ser contínuas ou categóricas.

variável X com um número J de classes superior a dois, a partição é realizada inicialmente com reclassificação da variável em duas superclasses, a partir da aplicação do algoritmo de cluster proposto por Hartigan e Wong (1979, *apud* Loh e Shih (1997)). Em seguida é aplicada a análise de discriminante quadrática para se determinar o ponto de partição <sup>13</sup>. Se a variável categórica for nominal, anteriormente a isso é necessária sua transformação em uma variável Z, por meio do mapeamento de X em vetores de variáveis *dummy* 0-1 e a projeção desses vetores sobre sua maior coordenada discriminante <sup>14</sup>. A seleção de variáveis é realizada também dependendo do fato da variável preditora ser categórica nominal ou ordinal, ou contínua. Se a variável X for nominal então são realizados testes Qui-Quadrado para a hipótese nula de independência entre X e a variável categórica dependente. Se X é contínua ou ordinal são utilizados testes F de análise de variância <sup>15</sup>.

**Diagrama 1 – Árvore de Classificação**



A estimação da árvore envolve o estabelecimento de vários parâmetros. Para a determinação da extensão da árvore são estabelecidos pelo pesquisador a extensão máxima de níveis de nós da árvore e as frequências mínimas nos nós pais e filhos. A classificação ainda é

<sup>13</sup> A análise de discriminante quadrática tem como principal vantagem a possibilidade de lidar com variâncias desiguais para as classes.

<sup>14</sup> A maior coordenada discriminante é a projeção que maximiza a razão entre a soma dos quadrados entre-classes e intra-classes. Para mais informações sobre esse método de transformação linear ver Gnanadesikan (1977).

<sup>15</sup> Se o menor valor p de probabilidade é inferior ao valor p de significância (0,05 no caso) ajustado segundo Bonferroni para comparações múltiplas, a variável preditora é escolhida para a partição do nó. Se nenhum valor p é menor que o valor p limite, valores p de probabilidade são computados para testes estatísticos que são robustos à violações distribucionais, tal como os testes F de Levene para variâncias desiguais. Para obter mais detalhes sobre o algoritmo de seleção quando nenhum preditor tem valor p de probabilidade menor que algum valor limite especificado, ver Loh e Shih (1997).

realizada em razão dos custos de má classificação e a probabilidade *a priori* para as categorias da variável dependente <sup>16</sup>. Todos esses parâmetros influem nos resultados obtidos pela árvore. Particularmente aqui, é interessante analisar a influência desses dois últimos parâmetros na possibilidade de se trabalhar com classes desequilibradas e custos distintos de má classificação.

Este ponto pode ser observado por meio da análise de discriminante quadrática apresentada em Loh e Shih (1997) para determinação do ponto de partição, com a introdução dos custos de má classificação. Adotando-se uma função de densidade de probabilidade normal univariada, a partição do nó pai para uma variável Y, com duas classes A e B, segundo uma variável explicativa X, é realizada a partir da determinação das raízes da seguinte equação quadrática <sup>17</sup>:

$$s_A^2 + s_B^2 x^2 + 2\bar{x}_A s_B^2 + \bar{x}_B s_A^2 x + \bar{x}_B^2 s_A^2 + \bar{x}_A^2 s_B^2 - 2s_A^2 s_B^2 \ln \frac{p_B s_A}{p_A s_B} = 0$$

ou

$$ax^2 + bx + c = 0 \tag{1}$$

em que

- $p_j$  = probabilidade a priori de  $j = A, B$
- $\bar{x}_j$  = média amostral de X para a classe  $j = A, B$
- $s_j^2$  = variância amostral de X para a classe de  $j = A, B$
- $a = s_A^2 + s_B^2$
- $b = 2\bar{x}_A s_B^2 + \bar{x}_B s_A^2$
- $c = \bar{x}_B^2 s_A^2 + \bar{x}_A^2 s_B^2 - 2s_A^2 s_B^2 \ln \frac{p_B s_A}{p_A s_B}$

A incorporação dos custos de má classificação à análise é realizada por meio da alteração das probabilidades *a priori*, redefinidas da seguinte forma:

$$p'_A = \frac{C_{B|A} p_A}{C_{B|A} p_A + C_{A|B} p_B} \text{ e } p'_B = \frac{C_{A|B} p_B}{C_{B|A} p_A + C_{A|B} p_B}$$

sendo

- $p'_j$  é a probabilidade *a priori* ajustada, com  $j = A, B$
- $c(A?B)$  = custo de classificar erroneamente como A caso X pertença a B
- $c(B?A)$  = custo de classificar erroneamente como B caso X pertença a A

<sup>16</sup> As probabilidades *a priori*, ou “*priors*”, especificam o quanto é provável, sem usar qualquer conhecimento *a priori* dos valores dos preditores, que um caso ou objeto pertença a uma das classes. No problema de classificação o pesquisador pode estabelecer que as probabilidades *a priori* sejam iguais às proporções de cada classe da variável dependente, equalizar esses valores para todas as classes, ou ainda estabelecer estes valores segundo uma idéia distinta sobre o fato. Já os custos de má classificação são penalidades numéricas para a classificação de um item em uma categoria ao qual este item não pertence de fato.

<sup>17</sup> Para ajudar no raciocínio imagine que A é não-inovador e B inovador, e X seja, por exemplo, o tamanho da empresa. Então o vetor de X seria particionado em  $X_A$  e  $X_B$ , gerando as médias e variâncias amostrais segundo cada uma das classes.

Assim, à medida que são estabelecidos custos de má classificação não simétricos, isto implica em ajustar as probabilidades *a priori* do modelo. Como é de conhecimento geral, as raízes dessa equação são dadas por:

$$d = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (2)$$

Sendo assim, para qualquer valor de  $\frac{p'_B s_A}{p'_A s_B} > 1$  isto implica em alterar o valor da constante *c*, e conseqüentemente o valor de *d*. Para simplicidade da análise admita o caso de variâncias iguais e médias desiguais. Com isso, a única raiz obtida é dada por:

$$d = \frac{\bar{x}_A - \bar{x}_B}{2} + \frac{s_A^2}{\bar{x}_A - \bar{x}_B} \ln \frac{p'_A}{p'_B} \quad (3)$$

Desse modo o deslocamento do ponto médio dependerá do fato de  $\bar{x}_A - \bar{x}_B > 0$  ou  $\bar{x}_A - \bar{x}_B < 0$  em conjunção com o fato de  $\ln \frac{p'_A}{p'_B} > 0$  ou  $\ln \frac{p'_A}{p'_B} < 0$ . Na presente análise,

sendo A a classe de empresas não inovadoras, e B a classe de empresas inovadoras, em que  $p_A > p_B$  e suponha que  $\bar{x}_A < \bar{x}_B$  com  $\bar{x}_A, \bar{x}_B > 0$ . Segundo a expressão (3) acima, se os custos de má classificação forem equivalentes, haverá uma tendência de deslocamento do ponto de partição em direção à classe menos numerosa de empresas minorando, a má classificação de empresas não-inovadoras e aumentando a parcela de erro na classificação de empresas inovadoras. A reversão dessa tendência deve ocorrer à medida que  $c^A|B < c^B|A$ , de modo

a tornar  $\ln \frac{p'_A}{p'_B} < 0$ . Essa discussão demonstra a importância do estabelecimento dos custos

de má classificação e, conseqüente, ajuste das probabilidades *a priori*, na determinação da classificação realizada pela árvore, na medida em que o ponto de partição é fundamental na verificação da relação entre a variável a ser predita e a variável preditora. Ao se alterar o ponto de partição, isso leva a uma alteração tanto na definição das variáveis preditivas selecionadas, bem como nos resultados de classificação da árvore.

A medida de capacidade, ou acurácia, preditiva de uma árvore é construída com base em uma matriz de classificação dada por:

Valor Predito	Valor Observado	
	Classe A	Classe B
Classe A	$N_{AC}$	$N_{AM}$
Classe B	$N_{BM}$	$N_{BC}$

em que

$N_{jC}$  é número de observações corretamente classificadas com  $j = A, B$

$N_{jM}$  é número de observações incorretamente classificadas com  $j = A, B$

$N_j = N_{jC} + N_{jM}$  com  $j = A, B$

Essa medida é igual a 1 – Risco, em que o Risco (ou erro aparente) é dado por:

$$\text{RISCO} = R(T) = \frac{N_{AM} + N_{BM}}{N_A + N_B} * 100 \quad (4)$$

Nesse trabalho o risco será calculado utilizando o procedimento de validação cruzada. Esse procedimento determina um número de subamostras aleatórias, de igual tamanho se possível. No presente caso igual a 10 amostras. Com isso, a árvore de classificação é computada 10 vezes, sendo que em cada uma dessas rodadas, uma dessas subamostras é deixada de fora da amostra e utilizada como amostra de testes para validação da árvore. O risco estimado apresentado em (4) será igual à média dos riscos obtidos em cada uma dessas estimativas.

A complexidade de análise de árvore de classificação pode ser um fator de custo para o pesquisador. Assim à medida que a árvore se expande, por um lado produz o benefício de redução do risco, mas por outro lado traz um custo com o aumento de sua complexidade. Um critério de redução da árvore é aquele que considera essa relação benefício/custo. Sendo assim, para a redução da árvore é utilizado um índice custo-complexidade que mede o risco e complexidade da árvore, dado por:

$$R_c(T) = R(T) + a|\tilde{T}| \quad (5)$$

em que  $a$  é o custo de complexidade e  $|\tilde{T}|$  é o número de nós terminais para uma árvore  $T$ <sup>18</sup>. O processo de redução é realizado por meio da estimação de seqüência de sub-árvores em que são removidos sub-ramos que promovem a menor redução do risco, relativamente ao cálculo recursivo de uma medida de complexidade. A escolha da árvore reduzida é de tal forma que o risco da menor das sub-árvores seja menor ou igual ao de risco da árvore de mínimo risco mais um erro padrão do risco.

### 3.2.2. Resultados

A amostra para a estimação das árvores é constituída de 38.176 empresas<sup>19</sup>. As estimativas de árvores foram realizadas estabelecendo como 5 a extensão máxima do nível de nós da árvore, 100 a frequência mínima nos nós pais e 50 a frequência mínima nos nós filhos. Inicialmente são utilizados como probabilidades *a priori* para as classes inovadoras e não inovadoras as respectivas parcelas dessas classes na amostra utilizada e estabelecidos custos de má classificação iguais a 1 para as classes não inovadoras e inovadoras. O nível de significância para os testes sobre as variáveis preditivas é de 5%. As variáveis selecionadas para análise e as transformações utilizadas estão apresentadas no quadro 1 abaixo. A árvore 1, apresentada a seguir, corresponde àquela estimada com processo de redução baseado no critério de risco mínimo. Ou seja, considerando os parâmetros acima descritos e o processo de validação cruzada, é selecionada a árvore que tem o risco médio mínimo. A seguir é apresentada na árvore 2 a estimação da árvore reduzida baseada no critério de custo-complexidade.

<sup>18</sup> Uma árvore poderia crescer até o momento em que cada nó terminal contivesse somente uma informação. Essa árvore teria um risco mínimo. Isto ocorreria se a fosse igual a zero e  $T$  fosse suficientemente grande, combinado à ausência de restrições para as frequências nos nós pais e filhos. Assim, à medida que  $a$  aumenta, a partir de zero, produz-se uma seqüência de sub-árvores  $T_1, T_2, T_3$  com progressivamente um número menor de nós.

<sup>19</sup> Essa amostra foi reduzida em razão de variáveis sem informação e devido à exclusão das empresas com valores para a variável de valor adicionado negativos.

**Quadro 1 – Variáveis Empregadas na Estimação das Árvores de Classificação**

Variável	Variável Empírica	Transformações
Não Inova/Inova (dependente)	Declaração das empresas para o período 1994/96	Categórica 0 = Não Inova 1 = Inova
Tamanho da Firma	Faixa de Pessoal Ocupado Total na Empresa	Categórica 1 = 0 a 29 2 = 30 a 49 3 = 50 a 99 4 = 100 a 249 5 = 250 a 499 6 = 500 a 999 7 = 1000 ou mais
Estrutura de Mercado	Medida CR4 para indústria paulista	Nenhuma
Orientação Exportadora	Empresas que exportam 1% ou mais do seu total de vendas	Categórica 0 = Não Exporta 1 = Exporta
Aprendizado Acumulado	Década de Constituição da Empresa	Categórica 1 = até 1949 2 = 1950 a 1959 3 = 1960 a 1969 4 = 1970 a 1979 5 = 1980 a 1989 6 = 1990 e depois
Qualificação do Trabalho	Média Salarial Mensal	Logarítmica
Origem do Capital Controlador	Nacional, Estrangeiro ou Misto em 1996	Categórica 1 = Nacional 2 = Estrangeiro 3 = Misto
Apropriabilidade	Relação entre Valor Adicionado e Receita Líquida da Empresa	Logarítmica
Efeitos Setoriais	Divisão CNAE <sup>20</sup> . Classificações Segundo as Oportunidades Tecnológicas (INTES3) e Segundo a Intensidade do Fator de Produção (2)	Categórica (INTENS3) 1 = Intens. em Recursos Naturais 2 = Intens. em Trabalho 3 = Intens. em Capital e Tecnologia Categórica (2) 1 = Alta Oportunidade 2 = Média Oportunidade 3 = Média-Baixa Oportunidade 4 = Baixa Oportunidade

As tabelas 7 e 8 apresentam os sumários estatísticos das árvores 1 e 2. Essas árvores possuem custos simétricos de classificação, sendo a primeira de risco mínimo e a segunda reduzida de acordo com o critério de custo-complexidade. Como é possível perceber o risco estimado para essas árvores é igual a 0,23<sup>21</sup>. Os nós terminais, em ordem decrescente, com maior percentual de empresas inovadoras são indicados pela coluna lucro nessas tabelas<sup>22</sup>. O

<sup>20</sup> As classificações aqui utilizadas foram propostas por Moreira e Najberg (1998).

<sup>21</sup> Essa estimativa de erro é inferior à estimativa ingênua de 0,5 e também, inferior à estimativa de que todas as empresas são não-inovadoras.

<sup>22</sup> É importante esclarecer que a árvore teria como objetivo a classificação estrita em cada nó. Ou seja, todas as informações, em cada um dos nós, deveriam ser de apenas uma das classes (Não-Inovadora ou Inovadora). Sendo assim, se, por exemplo,

primeiro nó com maior percentual de empresas inovadoras corresponde à categoria de empresas exportadoras, com menos de 50 empregados, atuantes em setores intensivos em recursos naturais relativamente concentrados. O segundo nó é aquele correspondente às empresas não exportadoras, com mais de 100 empregados atuantes em setores de alta e média oportunidades tecnológicas, com a presença de capital estrangeiro em suas composições societárias. O terceiro nó com maior percentual de empresas inovadoras corresponde àquele de empresas exportadoras com 250 ou mais empregados. A combinação da análise das árvores 1 e 2 demonstra que, em termos hierárquicos, a principal variável classificadora das empresas segundo suas classes de não inovadora e inovadora é o do fato dessas empresas serem ou não exportadoras, seguido pelo tamanho das empresas e pelos efeitos setoriais. Para toda a indústria são também determinantes na probabilidade de inovar, porém em menor magnitude, a origem do capital, o tempo de constituição da empresa e o grau de concentração de mercado. Os nós 3, 4, 5, 6 e 10 da árvore 1 demonstram que o percentual de empresas inovadoras é crescente à medida que aumenta o tamanho das empresas, tendo portanto, essa variável um efeito positivo sobre a probabilidade de inovar independentemente do fato da empresa ser exportadora ou não.

A partir do terceiro nível as variáveis preditivas são distintas para cada sub-ramo da árvore 1. No caso da categoria de empresas exportadoras com menos de 49 empregados o terceiro fator distintivo entre empresas inovadoras e não inovadoras é o efeito setorial, definido segundo a intensidade dos fatores de produção. Além da maior probabilidade de ser inovadora apresentada pelas empresas exportadoras com menos de 49 empregados atuantes em setores intensivos em recursos naturais relativamente concentrados, já destacada acima, esse mesmo subconjunto de empresas tem sua probabilidade de ser inovadora aumentada caso sejam atuantes em setores intensivos em capital e tecnologia e constituídas após a década de 1970. Ainda que exista uma maior probabilidade da empresa inovar na categoria de empresas exportadoras, com menos de 250 empregados atuantes e, setores de baixa, média-baixa e média oportunidades tecnológicas, conforme destacado pelo nó 19 da árvore 1, o processo de redução apresentado na árvore 2 demonstrou que essa categorização pouco contribui para a classificação das empresas entre inovadoras e não inovadoras. Por outro lado, é importante destacar que a maior probabilidade de se encontrar uma empresa não inovadora, ou inversamente, a menor probabilidade de se encontrar uma empresa inovadora, encontra-se na categoria de empresas não exportadoras com menos de 50 empregados, conforme indica o nó 5 das árvores 1 e 2.

As árvores de classificação 3 e 4 apresentam as estimativas de árvores reduzidas segundo o critério de custo-complexidade quando são estabelecidos custos de má classificação distintos, ajustando-se as probabilidades *a priori* segundo as variações dos custos. Na árvore 3 é estabelecido um custo igual a 2 para a classificação errônea de uma empresa não-inovadora como inovadora e custo igual a 1 para classificação errônea de uma empresa inovadora como não-inovadora. Ou seja, a árvore 3 deve classificar mais precisamente a classe de empresas não inovadoras. O inverso é feito na árvore 4. Ou seja, é estabelecido um custo igual a 2 para a classificação errônea de uma empresa inovadora como não-inovadora e custo igual a 1 para classificação errônea de uma empresa não-inovadora como inovadora. Os sumários estatísticos dessas árvores estão apresentados, respectivamente, nas tabelas 9 e 10 e demonstram estas afirmações por meio da observação das matrizes de

---

um nó da árvore contém um percentual da classe Inovadora superior a 50% esse nó está dizendo que mais provavelmente sob aquela partição esse nó representaria a classe de empresas inovadoras segundo classificação proposta pelo ramo ao qual este nó pertence. Dessa forma, a medida de risco do nó é a quantidade complementar ao percentual de empresas Inovadoras no nó. É nesse sentido que a tabela de lucros indica os nós em que o grau de homogeneidade é maior, significando que nesses nós a classificação é mais clara, e, portanto, menos sujeita a riscos.

classificação. Essas alterações referem-se às mudanças nos pontos de partição referentes à faixa de pessoal ocupado das empresas e à década de constituição.

Na árvore 3, basicamente, o que os novos nós informam que as grandes empresas exportadoras (com 500 ou mais empregados) têm uma probabilidade maior de inovar, ou caso não pertença a setores de alta oportunidade tecnológica, ou caso pertencentes a estes setores, tenham data de constituição anterior a 1950. Também é notada uma terceira alteração nessa árvore 3, com relação à árvore 2, que é a redefinição do ponto de partição para a classificação das empresas de acordo com o efeito setorial definido segundo as oportunidades tecnológicas. É possível perceber que, para empresas não exportadoras com mais de 100 empregados, a probabilidade de ser inovadora é maior para aquelas empresas atuantes em setores de baixa oportunidade tecnológica. No entanto, para as empresas pertencentes a setores de média-baixa à alta oportunidade tecnológica, com mais de 100 empregados, importa na definição da probabilidade ser inovadora o fato dessa ser estrangeira ou de capital misto.

Na árvore 4, em que o custo de má classificação é dobrado para a má classificação de empresas inovadoras como não inovadoras, o sub-ramo das empresas exportadoras foi redefinido, simplificando-o com relação aos resultados apresentados na árvore 2. Basicamente, destaca-se a informação de que a probabilidade das empresas serem inovadoras é maior no caso das empresas com menos de 100 empregados caso as mesmas pertençam a setores intensivos em recursos naturais ou setores intensivos em capital <sup>23</sup>. No nó referente às empresas não-exportadoras a alteração dos custos de má classificação na árvore 4 implicou em mudanças nos pontos de partição dos nós filhos das faixas de tamanho. Privilegiou-se a classificação das empresas não exportadoras com menos de 49 empregados, aquelas justamente com a maior proporção de empresas não-inovadoras. Analisando este sub-ramo sob a ótica das empresas inovadoras, é possível perceber que a maior probabilidade de uma empresa ser inovadora está no conjunto de empresas pertencentes a setores de baixa oportunidade tecnológica, com menor razão entre o valor adicionado e a receita líquida da empresa. Na faixa de empresas com 50 a 99 empregados, não exportadoras, os nós que apresentam maior proporção de empresas inovadoras são aqueles de média salarial superior. No entanto, é importante observar que os resultados da árvore não foram capazes de definirem novos nós terminais em que a proporção de empresas inovadoras fosse superior a 50 %.

Desse modo, o que estas estimativas permitem concluir é que a orientação exportadora é o principal fator distintivo para a realização ou não da atividade inovativa, seguido pelo tamanho da empresa. No entanto, independentemente da orientação exportadora, a probabilidade da empresa ser inovadora é aumentada à medida que aumenta o tamanho da empresa. Comparando os aumentos de proporções de empresas inovadoras nos nós filhos referentes às faixas de tamanho nota-se um acréscimo relativo superior naqueles nós filhos provenientes dos nós pais não exportadores. Esta observação sugere um efeito interado entre tamanho e orientação exportadora negativo. Ou seja, o aumento da probabilidade de inovar é maior à medida que aumenta o tamanho caso a empresa seja não exportadora. Posteriormente a análise deve ser compartimentalizada segundo as categorias de exportadora e não exportadora. No caso das exportadoras duas categorias emergem da classificação. As empresas com até 49 empregados, atuantes em setores ou intensivos em recursos naturais, ou em setores intensivos em capital e tecnologia, e as demais empresas exportadoras de tamanho superior. Quanto às empresas não exportadoras, a probabilidade de uma empresa ser inovadora possui uma relação direta com o tamanho, tendo destaque nesse ramo a presença de empresas de capital estrangeiro ou de capital misto em setores de média-baixa, média e alta tecnologia.

---

<sup>23</sup> Ou seja, essa nova partição condensou a informação contida nos sub-ramos determinados a partir dos nós 7 e 8 das árvores 1 e 2, de que a probabilidade de uma firma ser inovadora é maior caso pertença aos setores intensivos em recursos naturais ou aos setores intensivos em capital e tecnologia.

**Tabela 7 - Sumário Estatístico da Árvore de Classificação 1**

Nós	Frequência	Percentual	Lucro	matriz de classificação		
30	64	0,2	0,91	Valor Observado		
27	70	0,2	0,73	Valor Predito	Não Inova	Inova
10	917	2,4	0,69	Não Inova	27430	7693
19	1058	2,8	0,62	Inova	1123	1930
13	556	1,5	0,57			
18	388	1,0	0,55	Risco	0,2309	
20	327	0,9	0,48			
28	1364	3,6	0,45			
17	223	0,6	0,30			
29	54	0,1	0,24			
5	32919	86,2	0,21			
15	236	0,6	0,17			

**Tabela 8 - Sumário Estatístico da Árvore de Classificação 2**

Nós	Frequência	Percentual	Lucro	matriz de classificação		
18	64	0,2	0,91	Valor Observado		
15	70	0,2	0,73	Valor Predito	Não Inova	Inova
4	2302	6,0	0,63	Não Inova	27259	7537
9	556	1,5	0,57	Inova	1294	2086
14	388	1,0	0,55			
16	1364	3,6	0,45	Risco	0,2313	
13	223	0,6	0,30			
17	54	0,1	0,24			
5	32919	86,2	0,21			
11	236	0,6	0,17			

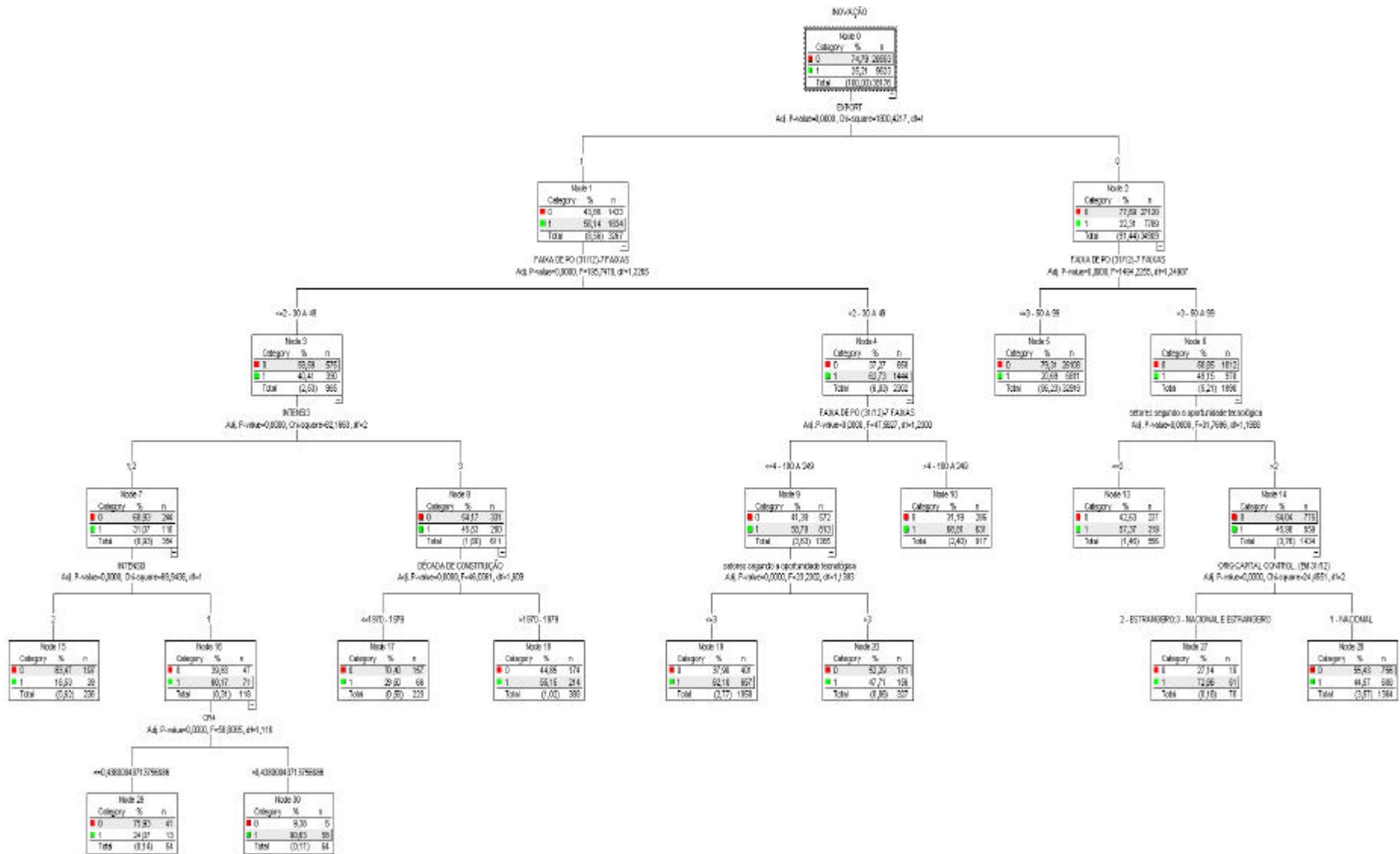
**Tabela 9 - Sumário Estatístico da Árvore de Classificação 3**

Nós	Frequência	Percentual	Lucro	matriz de classificação		
7	313	0,6	0,70	Valor Observado		
11	55	0,1	0,64	Valor Predito	Não Inova	Inova
9	70	0,1	0,59	Não Inova	25269	6135
13	98	0,2	0,58	Inova	3284	3488
3	2778	6,1	0,36			
14	1822	4,2	0,31	Risco	0,3254	
12	121	0,3	0,29			
5	32919	88,5	0,12			

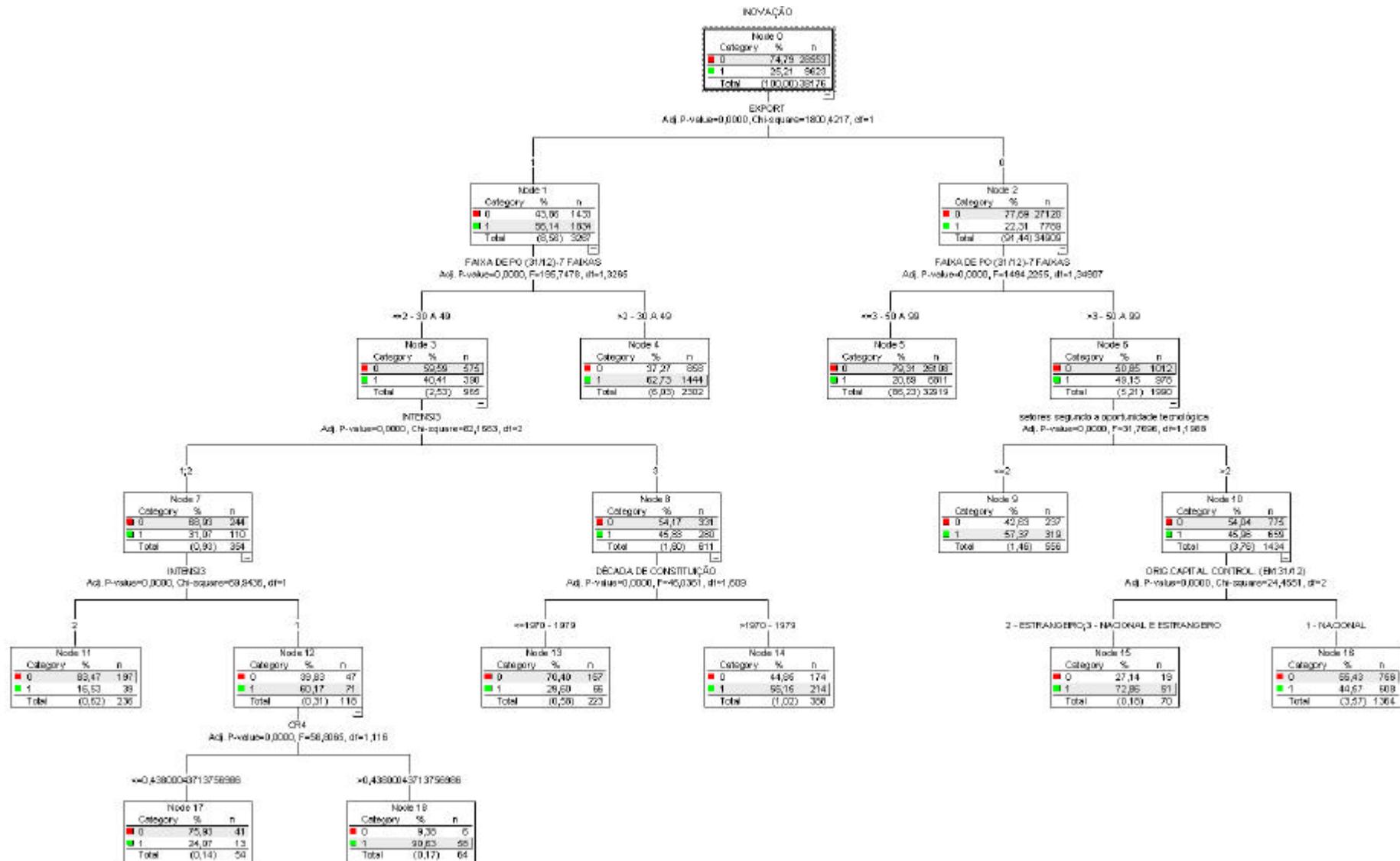
**Tabela10 - Sumário Estatístico da Árvore de Classificação 4**

Nós	Frequência	Percentual	Lucro	matriz de classificação		
4	1681	5,8	0,80	Valor Observado		
8	1262	4,0	0,68	Valor Predito	Não Inova	Inova
12	1990	6,2	0,66	Não Inova	28441	9199
14	239	0,7	0,62	Inova	112	424
17	348	1,0	0,60			
16	1252	3,7	0,57	Risco	0,1412	
15	1265	3,4	0,46			
18	362	0,9	0,40			
7	324	0,8	0,39			
10	29453	73,3	0,32			

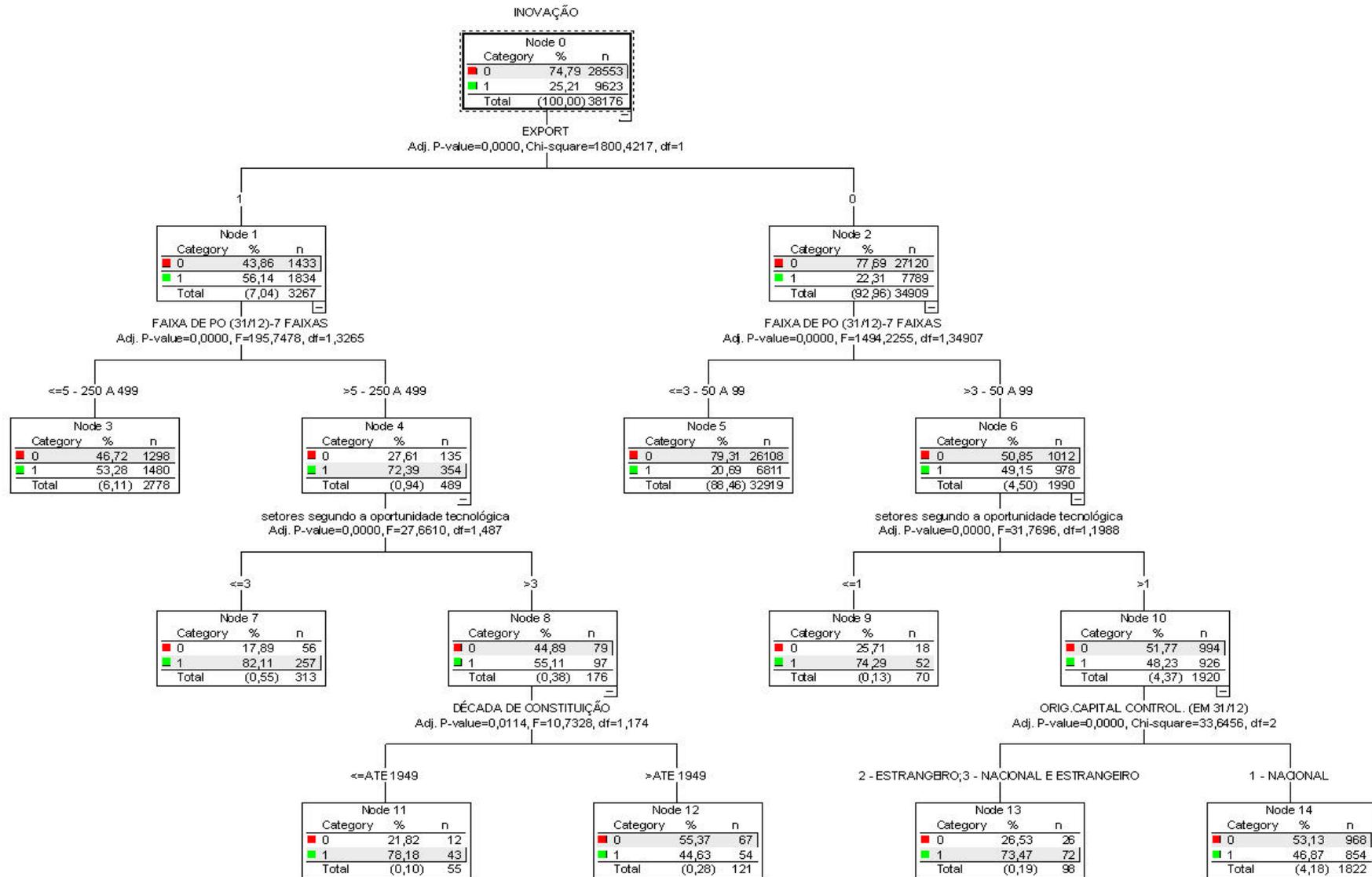
# Árvore 1: Não Inovadoras/Inovadoras



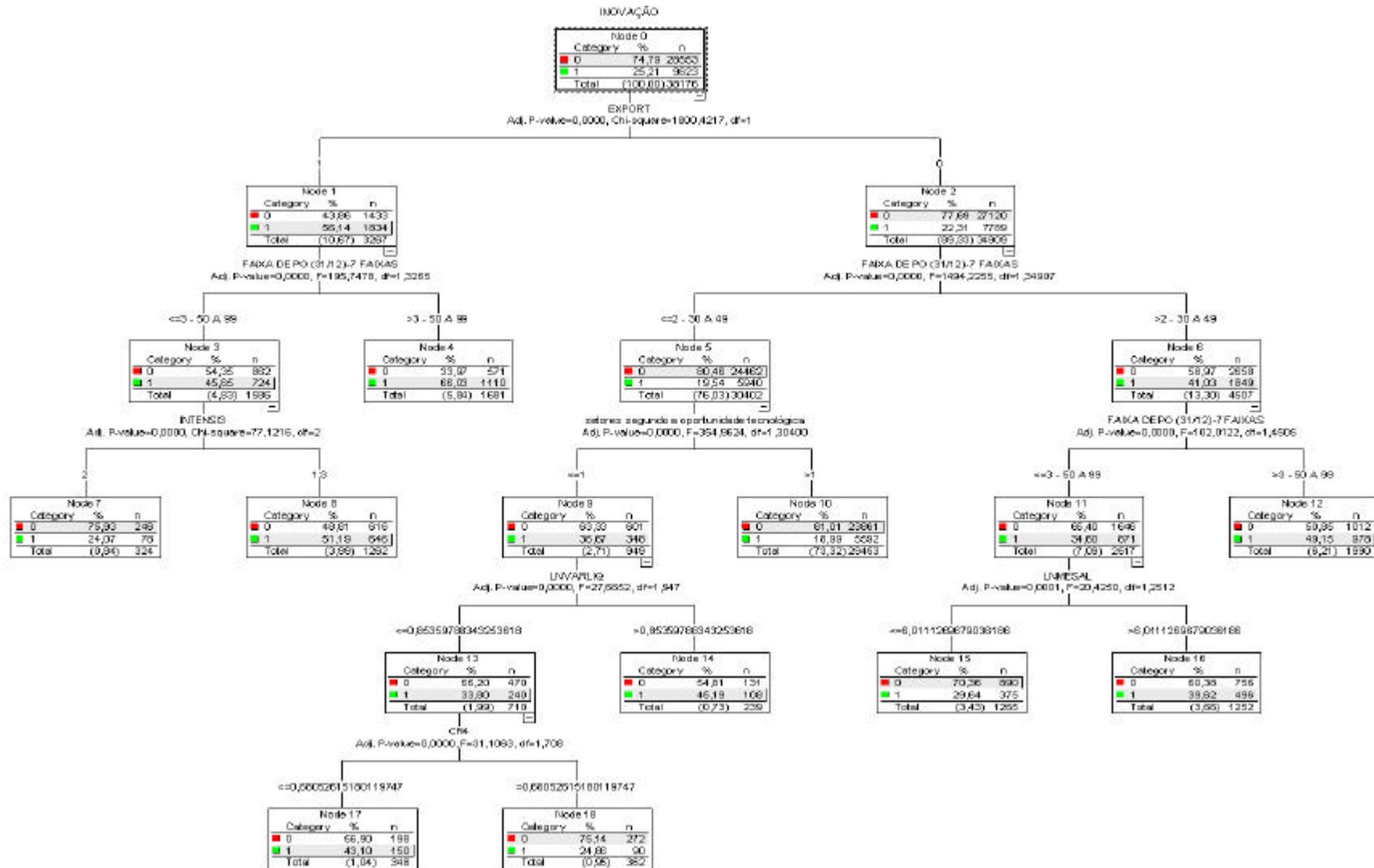
## Árvore 2: Não Inovadoras/Inovadoras



### Árvore 3: Não Inovadoras/Inovadoras



## Árvore 4: Não Inovadoras/Inovadoras



## Considerações Finais

Este trabalho procurou realizar uma classificação das empresas inovadoras segundo suas características patrimoniais, financeiras e recursos humanos, a fim de auxiliar na análise de informações sobre a atividade inovativa no estado de São Paulo, ou com algum grau de restrição, para a indústria brasileira. Isto ocorre na medida em que as classificações aqui estabelecidas são úteis para se definir as características relevantes para a distinção das empresas inovadoras em contraposição àquelas não inovadoras.

Trata-se de um primeiro trabalho nessa linha sobre o tema, abrindo perspectivas para trabalhos posteriores em que poderão ser analisadas categorias mais restritas, ou particulares, de atividade inovativa. Para sua realização foram utilizadas informações constantes na PAEP para o período de 1994 a 1996 e realizadas estimações de árvores de classificação. A hierarquização e o formato apresentado pelas árvores referendaram a maioria das proposições teóricas acima apresentadas na definição das variáveis da pesquisa. Um resultado interessante apresentado pelas árvores de classificação é aquele relativo à variável CR4, representativa da estrutura de mercado. Diferentemente da explicação usualmente oferecida pela teoria, em que possivelmente essa variável ajudaria na definição da probabilidade de grandes empresas atuantes como líderes em setores industriais concentrados, esta variável está associada às micro e pequenas empresas industriais. No caso dessas empresas serem exportadoras, os resultados obtidos denotam pressões competitivas oferecidas por estes setores mais concentrados sobre estas empresas. Apresenta evidência ambígua a distinção entre empresas inovadoras realizada pela variável representativa do aprendizado acumulado. Enquanto que o conhecimento tácito mostra-se um fator significativo para as grandes empresas exportadoras, conforme apresentado nos resultados da árvore 3, o inverso não se verifica para as pequenas empresas exportadoras atuantes em setores intensivos em capital e tecnologia, segundo os resultados da árvore 2, possibilitando a interpretação sugerida por Henderson (1993) apresentada acima.

Os resultados, em termos gerais, demonstram que a probabilidade de uma empresa ser inovadora está diretamente associada à orientação exportadora da empresa. Adicionalmente, em ordem decrescente, são fatores que alteram positivamente essa probabilidade o seu tamanho, a participação dessas empresas em setores intensivos em tecnologia, ou com maiores oportunidades tecnológicas, a origem não nacional do capital e grau de concentração setorial e a idade da empresa.

A distinção causal entre inovação e inserção internacional não é possível de ser realizada nesse estudo, ficando essa questão em aberto e a ser examinada em estudos posteriores. No entanto, pelo lado das empresas não exportadoras é possível interpretar as evidências de classificação das empresas como estratégias de proteção, ou obtenção de poder de mercado. Esta interpretação é oferecida pela relação clara entre a probabilidade de inovar e o tamanho da empresa. É interessante notar a maior probabilidade em inovar de empresas não exportadoras com capital estrangeiro pertencentes a setores com maiores oportunidades tecnológicas. Estas evidências sugerem a interpretação de uma possível estratégia competitiva não baseada em preços, em que o desenvolvimento de produtos e o acesso a novas tecnologias fariam parte de um pacote diferenciador do produto. Também é interessante destacar a probabilidade de inovar de empresas com mais de 100 empregados em setores de média-baixa e baixa oportunidades tecnológicas. A maior chance dessas empresas inovarem possivelmente está relacionada às estratégias de conservação do seu poder competitivo nos seus setores, frente à possibilidade de competição externa.

## **Bibliografia**

- BRAGA, H.; WILLMORE, L. **Technological imports and technological effort.** Journal of Industrial Economics. Blackwell Publishers, 39 (4): 421-432, Jun./1991.
- COHEN, W. M. **Empirical studies of innovative activity.** In: STONEMAN, P. (ed.), Handbook of the economics of innovation and technological change. Oxford: Blackwell, 1995.
- COHEN, W. M.; LEVIN, R. C. **Empirical studies of innovative activity and market structure.** In: SCHMALENSEE, R.; WILLIG, R. (EDS.), Handbook of Industrial Organization. Amsterdam: North Holland, 1989.
- DOSI, G. **Sources, procedures, and microeconomic effects of innovation.** Journal of Economic Literature. American Economic Association, 26 (3): 1120-1171, Sep./1988.
- GNANADESIKAN, R. **Methods for statistical data analysis of multivariate observations.** New York: John Wiley & Sons, Inc, 1977.
- HENDERSON, R. **Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry.** The RAND Journal of Economics. RAND Corporation, 24 (2): 248-270, Summer/1993.
- KASS, G. V. **An exploratory technique for investigating large quantities of categorical data.** Applied Statistics. Royal Statistical Society, 29 (2): 119-127, 1980.
- KUMAR, N.; SIDDHARTHAN, N. S. **Technology, market structure and internalization: issues and policies for developing countries.** London: Routledge, 1997.
- LOH, W. Y., SHIH, Y. S. **Split selection methods for classification trees.** Statistica Sinica, 7: 815-840, 1997.
- MOREIRA, M. M.; NAJBERG, S. **Abertura comercial: criando ou destruindo empregos?** Pesquisa e Planejamento Econômico. IPEA, 28 (2): Ago./1998.
- QUADROS, R., FURTADO, A., BERNARDES, R e FRANCO, E. **Technological innovation in Brazilian industry: an assessment based on the São Paulo innovation survey.** Technological forecasting and social change, 67 (2-3): 203-219, Jun./2001.
- PAVITT, K. **Sectoral patterns of technical change: towards a taxonomy and a theory.** Research Policy, 13: 343-373, 1984.

## **Apêndice - Classificação CNAE – Indústria de Transformação**

- 15 FABRICAÇÃO DE PRODUTOS ALIMENTÍCIOS E BEBIDAS
- 16 FABRICAÇÃO DE PRODUTOS DO FUMO
- 17 FABRICAÇÃO DE PRODUTOS TÊXTEIS
- 18 CONFECÇÃO DE ARTIGOS DO VESTUÁRIO E ACESSÓRIOS
- 19 PREPARAÇÃO DE COUROS E FABRICAÇÃO DE ARTEFATOS DE COURO, ARTIGOS DE VIAGEM E CALÇADOS
- 20 FABRICAÇÃO DE PRODUTOS DE MADEIRA
- 21 FABRICAÇÃO DE CELULOSE, PAPEL E PRODUTOS DE PAPEL
- 22 EDIÇÃO, IMPRESSÃO E REPRODUÇÃO DE GRAVAÇÕES
- 23 FABRICAÇÃO DE COQUE, REFINO DE PETRÓLEO, ELABORAÇÃO DE COMBUSTÍVEIS NUCLEARES E PRODUÇÃO DE ÁLCOOL
- 24 FABRICAÇÃO DE PRODUTOS QUÍMICOS
- 25 FABRICAÇÃO DE ARTIGOS DE BORRACHA E PLÁSTICO
- 26 FABRICAÇÃO DE PRODUTOS DE MINERAIS NÃO-METÁLICOS
- 27 METALURGIA BÁSICA
- 28 FABRICAÇÃO DE PRODUTOS DE METAL-EXCLUSIVE MÁQUINAS E EQUIPAMENTOS
- 29 FABRICAÇÃO DE MÁQUINAS E EQUIPAMENTOS
- 30 FABRICAÇÃO DE MÁQUINAS PARA ESCRITÓRIO E EQUIPAMENTOS DE INFORMÁTICA
- 31 FABRICAÇÃO DE MÁQUINAS, APARELHOS E MATERIAIS ELÉTRICOS
- 32 FABRICAÇÃO DE MATERIAL ELETRÔNICO E DE APARELHOS E EQUIPAMENTOS DE COMUNICAÇÕES
- 33 FABRICAÇÃO DE EQUIPAMENTOS DE INSTRUMENTAÇÃO MÉDICO-HOSPITALARES, INSTRUMENTOS DE PRECISÃO E ÓPTICOS, EQUIPAMENTOS PARA AUTOMAÇÃO INDUSTRIAL, CRONÔMETROS E RELÓGIOS
- 34 FABRICAÇÃO E MONTAGEM DE VEÍCULOS AUTOMOTORES, REBOQUES E CARROCERIAS
- 35 FABRICAÇÃO DE OUTROS EQUIPAMENTOS DE TRANSPORTE
- 36 FABRICAÇÃO DE MÓVEIS E INDÚSTRIAS DIVERSAS
- 37 RECICLAGEM