



**UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

JAIR SAMPAIO SOARES JÚNIOR

**TEORIAS SOCIAIS IMPLÍCITAS NOS ÍNDICES E
SISTEMAS DE INDICADORES: UMA
CONTRIBUIÇÃO ESTATÍSTICA AO ESTUDO DO
DESENVOLVIMENTO**

Salvador

2010

JAIR SAMPAIO SOARES JÚNIOR

**TEORIAS SOCIAIS IMPLÍCITAS NOS ÍNDICES E
SISTEMAS DE INDICADORES: UMA
CONTRIBUIÇÃO ESTATÍSTICA AO ESTUDO DO
DESENVOLVIMENTO**

Tese apresentada ao Núcleo de Pós Graduação em Administração, Escola de Administração, Universidade Federal da Bahia, como requisito parcial para obtenção do título de Doutor em Administração.

Orientador: Prof. Dr. Rogério Hermida Quintella

Salvador

2010

Escola de Administração - UFBA

S676 Soares Júnior, Jair Sampaio

Teorias sociais implícitas nos índices e sistemas de indicadores: uma contribuição estatística ao estudo do desenvolvimento / Jair Sampaio Soares Júnior. – 2010.

320 f. : il.

Orientador: Prof. Dr. Rogério Hermida Quintella.

Tese (doutorado) - Universidade Federal da Bahia, Escola de Administração, 2010.

1. Mineração de dados (Computação). 2. Sistemas especialistas. (Computação). 3. Desenvolvimento social. 4. Indicadores sociais.

I. Universidade Federal da Bahia. Escola de Administração. II. Quintella, Rogério Hermida. III. Título.

CDD 006.33

Dedico esta tese a minha amada filha Carolina,
fonte de inspiração da minha vida.

AGRADECIMENTOS

Escrever esta tese foi uma tarefa desafiadora e só foi possível chegar ao final graças ao apoio de muitas pessoas. Gostaria de usar este espaço para agradecer a todos que compartilharam deste desafio.

Ao amigo e colega, professor doutor Rogério Quintella, faço um agradecimento especial por sua permanente atenção, valorosas contribuições e ensinamentos transmitidos durante minha formação de acadêmico pesquisador. Serei eternamente grato à confiança e compreensão em mim depositada.

Aos meus irmãos, meus avós, tios, primos e amigos pela infinita compreensão em ser privados da minha companhia e contabilizarem muitas horas sem a minha presença ao longo do desenvolvimento desta pesquisa.

Aos meus professores, Antônio Pinho, Élvia Fadul, Robinson Tenório, Paulo Penteadó, Maria do Carmo Lessa, Tânia Fischer, Reginaldo Souza Santos, entre muitos outros. Sou muito grato ao seu carinho, dedicação e compreensão dos meus sonhos acadêmicos, meus devaneios de pesquisador e minha ansiedade.

À Direção, professores, funcionários e demais colaboradores do Núcleo de Pós-Graduação em administração – NPGA da Escola de Administração da Universidade Federal da Bahia – EAUFBA.

À Secretaria do NPGA da EAUFBA, em especial Dacy, Anaélia, Eliana, Cristina, Geovana, André e Adriano, por sua paciência e permanente boa-vontade.

Aos meus colegas e amigos da Superintendência de Estudos Econômicos e Sociais da Bahia – SEI e do Instituto Brasileiro de Geografia e Estatística – IBGE, pelos dados e idéias extremamente importantes para execução desta pesquisa.

Aos meus colegas de doutorado, por terem me mostrado o valor da diversidade e da distância que não separa. Em especial à Vitor Athayde Filho (*in memoriam*), por ter propiciado muitos momentos de reflexão sobre a abordagem econômica do desenvolvimento social.

A todos os demais professores dos quais tive o privilégio de interagir, aos meus queridos alunos, muitos deles hoje bacharéis, mestres e doutores, que sempre admiraram meu trabalho, em fim, aos amigos e colegas que juntei ao longo desta caminhada.

SOARES JÚNIOR, Jair Sampaio. Teorias Sociais Implícitas nos Índices e Sistemas de Indicadores: uma Contribuição Estatística ao Estudo do Desenvolvimento. 320 il. 2010. Tese (doutorado) 2011. Tese (Doutorado). Núcleo de Pós-Graduação em Administração, Escola de Administração, Universidade Federal da Bahia.

RESUMO

Dados públicos armazenados, nunca antes disponibilizados à população, hoje podem ser acessados livremente através da Internet. Paralelamente, impulsionado pelo aumento da capacidade computacional de armazenamento e processamento de dados, o Knowledge Discovery in Databases – KDD tem se tornado um método amplamente discutido para extração de conhecimento das bases públicas de dados. Este trabalho se propõe a contribuir para o estudo do desenvolvimento humano ao avaliar o potencial do KDD como método a ser empregado na mensuração do desenvolvimento social a partir de informações públicas. Por outro lado, a evolução do pensamento científico sobre o desenvolvimento converge cada vez mais para uma percepção transdisciplinar, complexa e intangível. Nessa perspectiva, os atuais métodos utilizados na construção de índices e sistemas de indicadores sociais, mostram-se insuficientes para representar o fenômeno e o KDD destaca-se em meio ao estado da arte das tecnologias de pesquisa empregadas no estudo desse tema como um método promissor, uma vez que contempla a modelagem de conceitos sociais e permite identificar e mensurar diversas relações entre os fatores associados ao fenômeno. A partir da teoria das medições, que se baseia na filosofia da ciência e na estatística, são exploradas as potencialidades analíticas do método na elaboração de índices sociais, tanto em nível teórico-metodológico – em que se discutem os construtos, modelos e indicadores – quanto em nível prático, através da sua aplicação ao contexto do desenvolvimento social no Brasil. Para isso, no nível teórico-metodológico, são analisados quarenta e três dos mais conhecidos trabalhos nacionais e internacionais utilizados na mensuração do desenvolvimento social. No nível prático, são utilizados dados públicos do Instituto Brasileiro de Geografia e Estatística, Ministério da Saúde, Ministério da Educação e Ministério da Justiça de todos os 5.560 municípios brasileiros. Os resultados da pesquisa apontam que, de fato, o KDD apresenta-se como um método com grande potencial analítico em relação aos métodos tradicionais, revelando-se, também, como adequado à abordagem multivariada do fenômeno, uma vez que é capaz de refletir, na esfera das investigações da realidade social, a complexidade do fenômeno, em afinidade com as formulações teóricas mais recentes.

Palavras-chave: Knowledge Discovery in Databases, Mineração de Dados, Índices, Indicadores, Modelos, Sistema de Indicadores, Desenvolvimento, Desenvolvimento Social.

SOARES JÚNIOR, Jair Sampai. il 2011- Implicit Theories in Social Indices and Indicators Systems: a Statistical Contribution to the Study of Development 320 il. 2010. Doctorate Thesis, Nucleus of Post Graduation in Administration, Federal University of Bahia , Business School.

ABSTRACT

Public data stored, never before available to the population today can be accessed freely via the Internet. In parallel, driven by increased computing storage and processing of data, Knowledge Discovery in Databases - KDD has become a widely discussed method for extracting knowledge from the public databases. This paper aims to contribute to the study of human development to assess the potential of KDD as a method to be employed in the measurement of social development from public information. Moreover, the evolution of scientific thinking on development converges increasingly to a perception disciplinary, complex and intangible. From this perspective, the current methods used in the construction of indexes and systems of social indicators, to be insufficient to represent the phenomenon and KDD stands out amid the state of the art search technologies employed to study this subject as a promising method, since it includes the modeling of social concepts and to identify and measure various relationships between factors associated with the phenomenon. From the theory of measurement based on philosophy of science and statistics, are exploring the potential of the analytical method in developing social indicators, both in theoretical and methodological - in which we discuss the constructs, models and indicators - as on a practical level, through its application to the context of social development in Brazil. For this, the theoretical and methodological, are analyzed Forty-three of the best known national and international work used in the measurement of social development. On a practical level, are used public data from the Brazilian Institute of Geography and Statistics, Ministry of Health, Ministry of Education and Ministry of Justice in all 5,560 Brazilian cities. The survey results indicate that, in fact, the KDD is presented as an analytical method with great potential in relation to traditional methods, revealing, too, as appropriate to the multivariate approach the phenomenon, since it is capable of reflecting, in the sphere of research of social reality, the complexity of the phenomenon, in affinity with the more recent theoretical formulations.

Keywords: Knowledge Discovery in Databases, Data Mining, Content, Content, Templates, System Indicators, Development, Social SUMÁRIO

LISTA DE ILUSTRAÇÕES

Quadro 1 –	Estrutura dos capítulos 2; 4; 6 e 7 da tese	16
Figura 1 –	Tipos de teorias segundo a possibilidade de decomposição e o critério de falseabilidade	39
Figura 2 –	Representação do Modelo Teórico Conceitual	41
Figura 3 –	Diagramas de Caminho Hierárquico e Circular	46
Quadro 2 –	Distribuição dos índices e sistemas de desenvolvimento mundiais por dimensão de análise, 2008	61
Quadro 3 –	Distribuição dos índices e sistemas de desenvolvimento brasileiros por dimensão de análise	62
Figura 4 –	Dimensões dos Índices de Desenvolvimento Nacionais e Mundiais	64
Gráfico 1 –	Índices e Sistemas de Indicadores Nacionais estudados, Brasil, 1991-200	66
Figura 5 –	Modelo Conceitual Hierárquico do Desenvolvimento Social disponível em 46 Índices e Sistemas de Indicadores nacionais	73
Figura 6 –	Gestão do Conhecimento, Sistemas de Informação, Sistemas de Apoio à Decisão, <i>Knowledge Discovery in Databases</i> , <i>Data Mining</i> e suas interfaces.	85
Figura 7 –	Etapas do processo KDD	91
Figura 8 –	Etapas do Processo <i>KDD</i>	92
Figura 9 –	Processo <i>KDD</i>	93
Figura 10 –	Fases e Subfases do processo <i>KDD</i>	94
Figura 11 –	Esquema do processo de extração de conhecimento	95
Figura 12 –	Base de dados antes da Descoberta de Conhecimento – DCBD	98
Figura 12 –	Base de dados depois da Descoberta de Conhecimento – DCBD	98
Figura 14 –	Fases e etapas em um processo de <i>Knowledge Discovery in Database – KDD</i> .	111
Figura 15 –	Modelo gráfico hipotético das relações entre os indicadores de desenvolvimento social	119
Figura 16 –	Modelo empírico comum do Desenvolvimento Social – Brasil, 2000	162
Quadro 4 –	Quadro resumo da abordagem dos capítulos 2; 4; 6 e 7 da tese	164

LISTA DE TABELAS

Tabela.1 –	Índices e Sistemas de Indicadores Nacionais de Desenvolvimento por região geográfica de elaboração – Brasil, 2010	67
Tabela 2 -	Métodos de Agregação utilizados nos Sistemas de Indicadores Nacionais - Brasil, 2009	69
Tabela 3 –	Resultado da sumarização dos 15 indicadores do Desenvolvimento Social para os 5.483 municípios brasileiros, Brasil – 2000	143
Tabela 4 –	Resultado da associação dos 15 indicadores selecionados para compor o Modelo Empírico do Desenvolvimento Social para os 5.483 municípios brasileiros, Brasil – 2000.	144
Tabela 5 –	Análise multivariável dos cinco construtos de Desenvolvimento Social Brasil, 2000	149
Tabela 6 –	Correlação dos 15 indicadores de desenvolvimento social com as dimensões originais – Brasil, 2000	152
Tabela 7 –	Correlação dos nove indicadores de desenvolvimento social com as dimensões originais – Brasil, 2000	153
Tabela 8 –	Autovalores e percentagens de variância explicada pelos construtos, com base em 9 indicadores	154
Tabela 9 –	Correlação dos indicadores de desenvolvimento com os fatores rotacionados	155

SUMÁRIO

1	INTRODUÇÃO	12
2	DESENVOLVIMENTO: AS TRAJETÓRIAS DE UM CONCEITO	19
2.1	PANORAMA HISTÓRICO DO DESENVOLVIMENTO: DO SÉCULO XVII AO SÉCULO XIX	20
2.2	DESENVOLVIMENTO NO SÉCULO XX: A PERSPECTIVA DAS AGÊNCIAS INTERNACIONAIS	21
2.3	ACEPÇÕES CONTEMPORÂNEAS DE DESENVOLVIMENTO	24
2.4	DESENVOLVIMENTO SUSTENTÁVEL	29
2.5	O BEM-ESTAR COMO VARIÁVEL OBJETIVA DO DESENVOLVIMENTO	32
2.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO	34
3	ASPECTOS TEÓRICO-METODOLÓGICOS DA MENSURAÇÃO DO DESENVOLVIMENTO SOCIAL	37
3.1	TERMOS TEÓRICOS, CONSTRUTOS, CONCEITOS E INDICADORES	37
3.2	MODELOS MATEMÁTICOS E AS CIÊNCIAS SOCIAIS	43
3.3	VARIÁVEIS, ESCALAS DE MEDIÇÃO, ESTRUTURAS HIERÁRQUICAS E CIRCULARES	45
3.4	MÉTODOS TRADICIONALMENTE UTILIZADOS PARA MENSURAR O DESENVOLVIMENTO SOCIAL	47
3.4.1	Definições	48
3.4.2	Tipos de Indicadores	51
3.4.3	Índices e Sistemas de Indicadores Sociais	52
4	ANÁLISE DA ESTRUTURA DE RELAÇÕES DOS INDICADORES ASSOCIADOS AO DESENVOLVIMENTO	54
4.1	BREVE HISTÓRICO DOS ÍNDICES E SISTEMAS DE INDICADORES DE DESENVOLVIMENTO	56
4.2	AS MÉTRICAS DO DESENVOLVIMENTO E SUA DINÂMICA	58
4.2.1	Sistemas e Índices Internacionais de Desenvolvimento Selecionados	60
4.2.2	Sistemas e Índices Nacionais de Desenvolvimento Selecionados	62
4.3	PRINCIPAIS SISTEMAS DE INDICADORES E ÍNDICES DE DESENVOLVIMENTO SOCIAL NO BRASIL	65
4.4	MODELO CONCEITUAL DO DESENVOLVIMENTO SOCIAL	71
4.4.1	Dinâmica Populacional: Crescimento Populacional e Fecundidade	74
4.4.2	Sistema de Produção (Trabalho e Rendimento): Distribuição do Rendimento; Ocupação e Rendimento	75
4.4.3	Longevidade (Saúde): Esperança De Vida Ao Nascer; Mortalidade; Desnutrição e Morbidade	76

4.4.4	Socialização (Educação): Escolarização; Analfabetismo e Escolaridade	77
4.4.5	Segurança: Homicídios e Morte no trânsito	78
5	KNOWLEDGE DISCOVERY IN DATABASES	80
5.1	GESTÃO DO CONHECIMENTO E SISTEMAS DE INFORMAÇÃO	81
5.2	SISTEMAS DE INFORMAÇÃO	83
5.2.1	Sistema De Apoio À Decisão	83
5.2.2	Características dos Sistemas de Apoio a Decisão	84
5.3	ASPECTOS TEÓRICOS DO MÉTODO KDD	86
5.4	FUNCIONAMENTO DO KDD	88
5.4.1	Etapas da descoberta de conhecimento em bancos de dados	89
5.5	MINERAÇÃO DE DADOS – DATA MINING	97
5.5.1	O Conceito de <i>Data Mining</i>	99
5.5.2	Diferentes abordagens	100
5.5.3	Objetivos da Mineração de Dados	102
5.5.4	Funções da Mineração de Dados	104
6	O PROTOCOLO KNOWLEDGE DISCOVERY IN DATABASE E OS MODELOS HIPOTÉTICOS	110
6.1	SÍNTESE DOS PROCEDIMENTOS	110
6.2	CRITÉRIOS PARA ADOÇÃO DO KDD	112
6.3	ETAPAS DO KDD EMPREGADAS	116
6.3.1	Prospecção	117
6.3.1.1	<i>Objetivos</i>	117
6.3.1.2	<i>Levantamento</i>	118
6.3.1.3	<i>Reunião, Seleção e Criação</i>	128
6.3.1.4	<i>Consistência</i>	131
6.3.1.5	<i>Compatibilização</i>	131
6.3.2	Mineração	132
6.3.2.1	<i>Transformação</i>	133
6.3.2.2	<i>Funções e Algoritmos</i>	133
6.3.2.3	<i>Avaliação dos resultados</i>	139
6.3.2.4	<i>Disseminação dos resultados</i>	140
7	KDD E MODELO HIPOTÉTICO DO DESENVOLVIMENTO SOCIAL	142
7.1	SUMARIZAÇÃO DA BASE DE DADOS	143
7.2	RESULTADO DA ASSOCIAÇÃO – MATRIZ DE CORRELAÇÕES	144
7.3	A ANÁLISE DO MODELO EMPÍRICO DO DESENVOLVIMENTO SOCIAL BRASILEIRO	146
7.3.1	Avaliação da consistência interna dos construtos do modelo empírico do desenvolvimento social	147
7.3.2	Análise Fatorial dos Indicadores associados o termo teórico Desenvolvimento Social	150

8	CONSIDERAÇÕES FINAIS	160
8.1	LIMITES E POSSIBILIDADES DO PROTOCOLO DE MÉTODOS KDD	165
8.2	PRINCIPAIS CONTRIBUIÇÕES	167
8.4	SUGESTÕES PARA PESQUISAS FUTURAS	170
8.4.1	Perspectivas	172
	REFERÊNCIAS	175
	APÊNDICES	187
	APENDICE A – Índices nacionais	187
	APENDICE B – Indicadores Internacionais autor/título	189
	APENDICE C – Indicadores Internacionais com resumo	231
	APENDICE D – Resumo cronológico de 20 índices de desenvolvimento	232
	APENDICE E – Lista de municípios	235
	APENDICE F – Padronizações (Transformação de Box&Cox)	253
	APENDICE G – Padronizações (Dados originais)	257
	APENDICE H – Padronizações (Transformação x^λ)	261
	APENDICE I – Box plot	266
	APENDICE J – Taxa de crescimento da população <i>versus</i> taxa de fecundidade total	272
	APENDICE K – Sistema de produção (trabalho e rendimento)	273
	APENDICE L – Longevidade (Saúde)	277
	APENDICE M – Socialização (Educação)	279
	APENDICE N – Violência (Segurança)	282
	APENDICE O – População residente por município <i>versus</i> número de domicílios por município	284
	APENDICE P – Transformações de Box&Cox	286
	APENDICE Q – Estatística descritiva dos principais indicadores	289
	APENDICE R – Histogramas	296
	APENDICE S – Dados Completos com correlação superior à 0,	302
	APENDICE T – Dados sem as observações com correlação superior à 0,5	304
	APENDICE U – Dados transformados x^λ .	306
	APÊNDICE V - Mensuração do desenvolvimento social no Brasil através de equações estruturais	310
	APÊNDICE X – Modelo circular do desenvolvimento social brasileiro	319

1 INTRODUÇÃO

O mundo contemporâneo assiste ao crescimento acentuado de dois fenômenos que motivam este trabalho. O primeiro deles é a difusão das tecnologias digitais e o segundo, o crescimento da parcela de sua população que vive sob os efeitos negativos do desenvolvimento sobre o meio ambiente e o bem-estar das pessoas. A humanidade, hoje, produz dados e informações em uma velocidade até recentemente inimaginável o que parece exceder, em muito, a capacidade de efetivamente utilizá-la. A transformação do crescente “oceano” de dados em conhecimento útil à humanidade é, provavelmente, um dos maiores desafios administrativos de nosso tempo. As discrepantes velocidades com que crescem o volume de dados e o volume de conhecimentos, encontra um importante paralelo com os problemas sociais da “era da informação” já que indubitavelmente, crescimento da riqueza econômica nem sempre é acompanhado de melhorias no bem-estar das sociedades.

É necessária a inteligência para se transformar dados e informação em conhecimento e é igualmente necessária a inteligência para se transformar riqueza em bem-estar humano. Desse tema trata esta tese.

Desde o final do século passado, a contribuição da Tecnologia da Informação (TI) na esfera pública passou de um papel restrito de suporte administrativo, para uma participação mais estratégica nas tomadas de decisão, auxiliando na formulação e análise de políticas governamentais.

O processo de globalização, a Internet e, no Brasil, a consolidação da democracia, aliados a disseminação mais rápida, ampla e com sistemas mais amigáveis (*user friendly*) para atender a crescente diversidade dos níveis de usuários, tornaram os cidadãos mais exigentes, ao tempo em que o mercado tornou-se mais competitivo. O cidadão passou a demandar mais do poder público em defesa de seus direitos. O aumento da procura por informações e a necessidade legal de maior transparência nas ações do gestor público culminaram na crescente disponibilização de informações dos principais órgãos de governo na esfera federal conduzindo, gradativamente as unidades da federação e, mais recentemente os municípios, a também estruturarem e disponibilizarem mais informações.

Entre os movimentos recentes da área de Gestão do Conhecimento na esfera pública está a implantação de aplicações e sistemas que permitem análises e suporte à tomada

de decisão a partir de Bases de Dados Públicas. Essa crescente busca por informação relevante de difícil detecção por métodos tradicionais de análise para ser utilizada pelos tomadores de decisão, levou os pesquisadores a investirem no *Knowledge Discovery in Databases* (KDD) – processo não trivial e estruturado de identificação de padrões existentes através da modelagem de fenômenos do mundo real, a fim de extrair de informações e conhecimentos potencialmente úteis e implicitamente contidos em bases de dados.

Nesta tese o que se pretende é analisar a transformação de dados públicos em conhecimento de valor social através do uso de ferramentas computacionais não tradicionais, que permitam a extração de conhecimento implícito em grandes bases de dados.

Dentre a miríade de ferramentas, métodos, conceitos e teorias, alguns são especificamente direcionados à utilização empresarial, outros, porém, tem tido ampla utilização em questões relacionadas à gestão de interesses públicos. Estas são tratadas adiante.

O crescimento da demanda por informações governamentais, evidenciada pela necessidade de investigação de temas emergentes e maior aprofundamento daqueles já investigados, a busca pela redução do intervalo de divulgação de temas já existentes, a necessidade de maior desagregação territorial e, principalmente, a necessidade legal¹ de maior transparência nas ações do gestor público culminaram na crescente disponibilização de informações dos principais órgãos de governo na esfera federal conduzindo, gradativamente as unidades da federação a também estruturarem neste sentido. Os governos passaram a promover transformações organizacionais para acompanhar e monitorar as transformações sociais daí resultantes, o que fez ressurgir o interesse e as possibilidades de aplicação de novas métricas para subsidiar a tomada de decisão por parte de governos e organismos públicos no sentido de avaliarem o esforço e os impactos de políticas públicas, assim como orientar a alocação do Gasto Social.

Atualmente, diversos importantes trabalhos² que tratam da mensuração das transformações ocasionadas pelo processo de desenvolvimento das nações, o fazem considerando o monitoramento de casos de ‘mudança social acelerada’ e, paralelamente, de um conceito quase que coloquial de bem-estar. Grande parte desses sistemas de indicadores ou índices de desenvolvimento tem por finalidade não somente mostrar, de forma sucinta, a

¹ A exemplo da Lei de Responsabilidade Fiscal - LRF

² Nesta tese foram analisadas as principais formas de mensuração encontradas nos Índices e Sistemas de Indicadores de Desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos, totalizando 397 trabalhos internacionais e 64 trabalhos nacionais.

relação entre os fatos e as mudanças, mas também subsidiar o monitoramento dos processos de transformações de natureza econômica, política, social e ambiental entre outras.

É neste contexto que se define o **problema de pesquisa** desta tese como:

Qual a construção teórico-hipotética, susceptível de matematização, com a qual se pode representar o fenômeno desenvolvimento social no Brasil por meio dos seus principais índices e sistemas de indicadores?

Diante da crescente disponibilização de Bases de Dados pelo poder público e da propagação dos Índices e Sistemas de Indicadores de Desenvolvimento Social, sejam através da Internet ou mídias específicas, este trabalho tem como objetivo principal **contribuir com o estudo do Desenvolvimento por meio da aplicação do método *Knowledge Discovey in Databases* (KDD) em bases de dados públicas, para identificar o modelo empírico comum aos instrumentos brasileiros de mensuração do desenvolvimento social.**

Para atender a esse objetivo geral identificam-se os seguintes objetivos específicos:

- a) Analisar a trajetória do conceito de desenvolvimento, buscando compreender a sua utilização nas principais formas de mensuração encontradas nos índices e sistemas de indicadores de desenvolvimento utilizados no Brasil e no mundo;
- b) Estruturar os conceitos e metodologias para construção de índices e sistemas de indicadores sociais que possam mensurar conceitos relevantes de um sistema social;
- c) Identificar e analisar as principais formas de mensuração encontradas nos Índices e Sistemas de Indicadores de Desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos;
- d) Aplicar o KDD para confrontar o modelo conceitual do desenvolvimento social sobre os dados disponibilizados pelo IBGE e Ministério da Saúde para 5.483 municípios brasileiros, a fim de detectar e estabelecer o modelo empírico comum aos índices e sistemas de indicadores de desenvolvimento social brasileiros.

O objeto do estudo aqui proposto é a aplicação do método KDD à aferição de modelos empíricos relacionais entre as variáveis mais relevantes para a mensuração do desenvolvimento social no Brasil.

Diante das considerações anteriores, o campo temático do presente trabalho pode ser definido como o da mensuração de fenômenos sociais, mais especificamente do desenvolvimento social, através do uso de ferramentas computacionais que permitam, de forma automática e inteligente, analisar, interpretar e relacionar grande volume de dados e assim poder utilizá-los de maneira útil, num processo conhecido KDD.

A relevância do tema pode ser ilustrada pela crescente demanda social por transparência nas ações públicas e disponibilização de sistemas de indicadores que favoreçam o monitoramento e avaliação de metas sociais, a exemplo das Metas do Milênio propostas pelo Programa das Nações Unidas para o Desenvolvimento – PNUD.

Apesar das tentativas de construção de índices e sistemas de indicadores para apreensão do fenômeno do desenvolvimento social, aparentemente, nenhum procedimento totalmente comprovado está disponível na literatura. A ausência de índices ou sistemas de indicadores de desenvolvimento adequados provém principalmente da dificuldade de identificar indicadores que representem os conceitos relacionados com os principais construtos teóricos do fenômeno desenvolvimento social. Por isso, é difícil saber se os resultados dos índices são meramente acidentais, ou realmente refletem algo significativo.

O material utilizado para embasar a discussão aqui apresentada foi reunido através de ampla pesquisa para a identificação de experiências de Sistemas de Indicadores e Índices construídos no Brasil e no Mundo nos últimos 45 anos. Foram utilizadas modernas e avançadas técnicas de recuperação da informação o que permitiu agregar uma série de facilidades que agregam novos valores para as buscas em *offline* e *online*, na forma de análises quantitativas e qualitativas.

Deve-se ressaltar, por fim, que o presente estudo não tem a pretensão de propor uma análise sociológica ou política aprofundada do fenômeno desenvolvimento, mas de contribuir para o avanço no uso de métodos quantitativos na área de Administração Pública.

No presente capítulo, é apresentada uma visão geral da problemática discutida nesta pesquisa e seus objetivos principais, a motivação para o seu desenvolvimento, a relevância do tema, os objetivos propostos, bem como uma descrição da organização do trabalho. O restante de trabalho está estruturado em quatro partes (Quadro 1): na primeira, é feita uma revisão da literatura (Capítulos 2 e 3) sobre o desenvolvimento e os principais métodos de mensuração de fenômenos sociais, na segunda parte, são analisados os principais índices e sistemas de indicadores sociais e produzido um modelo hipotético de influências

entre os indicadores selecionados (Capítulo 4), na terceira é apresentado o método KDD (Capítulo 5) e na quarta e última parte (Capítulos 6 e 7), são analisados os resultados da aplicação do KDD na aferição do ajuste do modelo de empírico aos dados dos municípios brasileiros, por fim, são feitas as considerações finais desta pesquisa.

Quadro 1 – Estrutura dos capítulos 2; 4; 6 e 7 da tese

Parte	Capítulo	Dimensão	Método	Unidade observacional	Fonte	Modelo
I	2 DESENVOLVIMENTO: AS TRAJETÓRIAS DE UM CONCEITO	Termo Teórico	Revisão bibliográfica	Teorias do Desenvolvimento	Bibliografia sobre Desenvolvimento	Teórico
II	4 ANÁLISE DOS ÍNDICES E SISTEMAS DE INDICADORES DE DESENVOLVIMENTO	Conceitos	Informetria	Instrumentos de mensuração	Documentos dos Índices e Sistemas de Indicadores	Conceitual
IV	6 O PROTOCOLO KNOWLEDGE DATA DISCOVERY E OS MODELOS HIPOTÉTICOS	Indicadores	KDD - Prospecção	Bases de Dados	IBGE; MS; MJ	Hipotético
IV	7 MODELOS EMPÍRICOS DO DESENVOLVIMENTO SOCIAL	Variáveis	KDD - Mineração	Jazida de Dados	Elaboração própria	Empírico

Fonte: Elaboração própria

O objetivo do Capítulo 2 – Desenvolvimento: as trajetórias de um conceito – é analisar a trajetória do conceito de desenvolvimento, buscando compreender a sua utilização nas principais formas de mensuração encontradas nos índices e sistemas de indicadores de desenvolvimento utilizados no Brasil e no mundo. A linha de raciocínio adotada neste capítulo procura, assim, apresentar a discussão do termo teórico de desenvolvimento em uma perspectiva histórica, considerando as suas dimensões políticas que ultrapassam, portanto a tradicional abordagem da simples mensuração, enveredando por diferenças conceituais, paradigmáticas e de tradições teóricas. São apresentadas as definições para o conceito a partir das três correntes predominantes na abordagem do desenvolvimento e discutidas as suas principais características. O foco neste capítulo não é a descrição da trajetória do saber sobre mensuração do desenvolvimento, mas a investigação das bases conceituais e metodológicas desse processo. Para avaliar os instrumentos de mensuração, buscou-se seus princípios, requisitos e fundamentos, pois empiricamente pode-se inferir que as bases do processo de mensuração do desenvolvimento são intuitivas. O capítulo oferece o respaldo mínimo

necessário ao entendimento dos Indicadores de Desenvolvimento (domínio) que serão estudados pelo processo de *KDD* neste trabalho.

O Capítulo 3, intitulado Aspectos teórico-metodológicos da mensuração do desenvolvimento social, tem o objetivo de estruturar os conceitos e metodologias para construção de índices e sistemas de indicadores sociais que possam mensurar conceitos intangíveis e relevantes de um sistema social. Para consecução desse objetivo foram estudados os aspectos teórico-metodológicos da mensuração de fenômenos sociais e sua conexão com as principais propostas de formulação e construção de Índices Sociais disponíveis na literatura, a fim de subsidiar a construção de categorias analíticas para o estudo dos instrumentos de avaliação do desenvolvimento.

O Capítulo 4, intitulado Análise dos índices e sistemas de indicadores de desenvolvimento, consiste em identificar as relações (conceito ordenador) entre os indicadores utilizados nos instrumentos de mensuração do desenvolvimento mais reconhecidos nacional e internacionalmente. Para alcançar este objetivo foram selecionados, a partir de categorias analíticas previamente elaboradas dentre Índices e Sistemas de Indicadores Sociais mais conhecidos, os sistemas de avaliação do desenvolvimento social mais reconhecidos no contexto nacional e internacional e, em seguida, foram analisados os indicadores propostos nos sistemas de avaliação selecionados e avaliada a relevância desses indicadores no que se refere ao desenvolvimento social. Por fim, são analisados o corpus teórico, os conceitos ordenadores e os indicadores empregados nos instrumentos nacionais de avaliação do desenvolvimento selecionados.

O Capítulo 5 – Método KDD no estudo do desenvolvimento social – prevê uma revisão dos conceitos fundamentais, pertinentes e relevantes sobre o tema, identificando as principais propostas de conformação de etapas do processo de KDD disponíveis na literatura. Além do mais, são tratadas suas diferentes abordagens, evidenciando a forte influência da formação ou da área de origem dos modelos na perspectiva por eles utilizada. O capítulo traz uma proposta de conformação de etapas do processo de KDD para bases de dados públicas.

O Capítulo 6 – O protocolo *Knowledge Data Discovery* e os modelos empíricos – é dedicado a descrever a estratégia metodológica do KDD, utilizada na elaboração do Modelo Empírico comum aos Índices e Sistemas de Indicadores de Desenvolvimento Social brasileiros. Por tratar-se de uma abordagem recente e pouco explorada na literatura brasileira, procurou-se destacar aspectos específicos de maneira a permitir ao leitor, se necessário, replicar o estudo na mesma ou em outra base de dados. O capítulo traz uma proposta de

conformação de etapas do processo de *KDD* para bases de dados públicas. Propõe-se um modelo comum de mensuração para o **fenômeno** desenvolvimento social e, por fim, delimita-se e estrutura-se uma Base de Dados de porte e relevância social para tal fim.

O Capítulo 7 – Apresenta uma análise articulada da estrutura de relações entre os principais indicadores associados ao desenvolvimento social – tem o objetivo de estabelecer um modelo empírico entre os indicadores sociais com base nos conhecimentos teóricos e práticos que se tem dos instrumentos de mensuração do Desenvolvimento Social, a fim de detectar e estabelecer conjecturas explicativas sobre este fenômeno no Brasil. Para consecução do objetivo proposto, apresenta-se inicialmente os resultados do estudo empírico, bem como da aplicação do *KDD*. São confrontados, neste ponto, o modelo conceitual e o modelo empírico, resultando numa proposta de representação do modelo empírico comum aos instrumentos brasileiros de mensuração do desenvolvimento social.

Por fim, o capítulo 8 – Considerações Finais – está estruturado em quatro seções: a primeira apresenta uma síntese das principais conclusões e resultados da pesquisa. São apontadas as limitações relacionadas essencialmente à escassez de trabalhos correlatos na literatura. Os limites, perspectivas, recomendações para estudos futuros e barreiras na aplicação do *Knowledge Discovery in Databases* no estudo dos instrumentos de mensuração do Desenvolvimento são apresentados na seqüência.

2 DESENVOLVIMENTO: AS TRAJETÓRIAS DE UM CONCEITO

La vision du futur sur laquelle une civilisation règle ses projets, fonde ses buts idéaux et ses espérances. (GORZ, 1988, p.22)

O objetivo deste capítulo é apresentar aspectos relativos às interpretações dominantes do conceito Desenvolvimento procurando agrupá-los segundo a similitude de suas abordagens, visando subsidiar o levantamento de questões que envolvem a instrumentalização desses conceitos e suas formas de mensuração.

A análise proposta pretende oferecer a especificação do termo teórico ou construto que melhor define o conceito intangível Desenvolvimento, evidenciando, como o referencial teórico conceitual determina a arquitetura e a construção dos principais indicadores e índices para mensuração desse fenômeno.

Reconhecendo o grau de complexidade que envolve o tema, pretende-se, inicialmente, identificar aspectos relativos às interpretações e ideologias dominantes nos conceitos de “desenvolvimento/crescimento”.

O processo de desenvolvimento abrange uma complexidade de relações. Sua análise, portanto, não pode ser restrita apenas à dimensão econômica como, via de regra, apresenta-se a questão, tanto na mídia quanto em parte considerável da literatura especializada. A ênfase nessa dimensão tem origens históricas que já evidenciam a fragilidade dessa abordagem.

Para atender ao objetivo do capítulo, foram definidos os seguintes objetivos específicos: 1. Identificar as principais diferenças e similitudes nas abordagens teóricas sobre desenvolvimento e reuni-las por corrente de pensamento; 2. Analisar o termo teórico desenvolvimento e seus desdobramentos.

As seções que se seguem mostram a evolução do conceito através de uma revisão teórica que possibilita a abordagem aprofundada das métricas do desenvolvimento que se faz a partir do capítulo 6 dessa tese.

2.1 PANORAMA HISTÓRICO DO DESENVOLVIMENTO: DO SÉCULO XVII AO SÉCULO XIX

O conceito de desenvolvimento é quase tão antigo quanto a civilização. Sua larga utilização nas sociedades ocidentais, desde as civilizações greco-romanas até o final do século XIX como um construto genérico que designa os mais variados aspectos relacionados ao bem-estar da humanidade fez, porém, com que o conceito se aproximasse de uma doutrina. Para fins do presente estudo optou-se por adotar um recorte temporal a partir do ideário iluminista do século XVII e de suas metamorfoses até os dias atuais.

Ao final daquele século a idéia predominante de desenvolvimento era dependente de fenômenos naturais e ‘positivos’, sendo sua expansão contida apenas pela ‘consciência de limite’³. Após a longa disputa entre chamados Modernos e Antigos, finalizada com a vitória dos Modernos, Leibniz (1646-1716) inaugura o conceito de progresso infinito⁴.

Outros pensadores modernos a exemplo de Condorcet, Kant, Hegel e Marx, por sua vez, concebem e interpretam o progresso de maneira distinta daquela decorrente da idéia de ‘consciência de limite’, ou seja, alinham-se com o conceito de progresso potencialmente infinito. No trabalho desses autores é possível perceber também certa proximidade com o pensamento Agostiniano ao conceber a história como totalidade⁵, uma firme marcha da civilização, um processo contínuo mesmo que inconstante e não linear, na direção do bem estar comum. Nas palavras de Rist: “um evoluir constante, alicerçado na crença da perfectibilidade humana e motivado pela busca incessante de um bem-estar” (RIST, 2005, p.70).

Em meados do século XVIII, opondo-se radicalmente à visão rousseuniana do bom selvagem⁶, Buffon (1707-1788) defende a idéia de que na natureza existe um protótipo geral em cada espécie e de que esse protótipo aperfeiçoa-se em função do clima e dos hábitos da sociedade em que se insere. Em sua obra, *A História Natural*, o autor defende a idéia de que o europeu é aquele para quem a civilização irá chegar e afirma: “*Du fait même de leur*

³ Um ponto de inflexão da curva de crescimento, para se adequar às leis naturais ou ao plano de Deus.

⁴ A noção de “progresso infinito” está relacionada com a concepção linear e progressista da história, e para a qual a cultura ocidental deve ser disseminada a todos os povos.

⁵ A noção de história como totalidade empregada por Agostinho distingue-se do modo de pensar as totalidades sistêmicas em raciocínio teórico modelar conforme é apresentado no capítulo 3 desta tese.

⁶ Rousseau glorifica os valores da vida natural e ataca os vícios da sociedade civilizada. O mito do “bom selvagem” está associado à idéia de maior perfectibilidade do “nativo” e a pureza do seu estado natural, contrapondo-se à falsidade e ao artificialismo do homem civilizado.

supériorité, les peuples policés sont responsables d'un monde en devenir” (BUFFON apud DUCHET, 1972, p.54). Desse modo, Buffon dá quase que ‘força de lei’ ao que ele percebe como uma realidade histórica.

Ainda no século XVIII, o Marquês de Condorcet (1743-1794) lançou as bases do pensamento que iria predominar na segunda metade do Século XX, ao defender a idéia de que a Europa acabaria respeitando a independência de suas antigas colônias e, em seguida, deveria contribuir com elas através da civilização⁷ de seus povos (RIST, 2001). Um outro paradigma de desenvolvimento, porém, já havia alcançado seu apogeu no século XIX, sob a forma de Evolucionismo Social. O termo desenvolvimento, nesse novo paradigma, prevaleceu sobre conceitos tais como ‘modernização’ ou ‘liberação’. Tal predominância parece já decorrer da necessidade de uma concepção mais ampla para representar as múltiplas dimensões necessárias ao bem-estar da humanidade.

Pode-se agora destacar um aparente alinhamento das teorias da modernidade no sentido de perceber os grupos sociais como uma única espécie e, portanto, devendo apresentar trajetórias similares de desenvolvimento. Entretanto, a recorrente analogia com os processos naturais de desenvolvimento não explica nem consegue lidar adequadamente com a multiplicidade cultural presente nas sociedades, particularmente nos séculos XX e XXI, conforme é discutido a seguir.

2.2 DESENVOLVIMENTO NO SÉCULO XX: A PERSPECTIVA DAS AGÊNCIAS INTERNACIONAIS

A crescente difusão de conceitos e políticas de desenvolvimento permite que as últimas décadas possam ser aqui nominadas como era do desenvolvimento. A origem dessa *era*, em que supostamente ainda vivemos, pode ser localizada no final da primeira metade do Século XX, possivelmente na política internacional do presidente norte-americano Harry S. Truman, quando este decide destacar a importância de se ampliar a “ajuda técnica” oferecida a alguns países da América Latina e a algumas outras nações menos favorecidas:

I believe that we should make available to peace loving peoples benefits of our store of technical knowledge in order to help them realize their

⁷ O termo Civilização foi empregado até a primeira Guerra Mundial para representar o processo de ocidentalização promovido, principalmente, pelos países da Europa.

aspirations for a better life. And, in cooperation with other nations, we should foster capital investment in areas needing development. (TRUMAN, 1949).

As mencionadas idéias do Marquês de Condocert podem assim ser confrontadas com uma nova situação internacional. Observa-se que o ano de 1960 é considerado ‘o ano da descolonização’. Processo, até então, legitimado meramente como uma ‘empreitada generosa’ visando oferecer uma oportunidade às sociedades consideradas mais atrasadas para ‘avançarem’ no caminho da civilização⁸.

Vários países tornaram-se independentes a partir daquele momento e o suposto processo de “descolonização” percorreu o mundo, mudando a relação de forças até mesmo no plenário das Nações Unidas e aparentemente colocando em um plano mais elevado os problemas da desigualdade no mundo e a necessidade de desenvolvimento no então chamado ‘terceiro mundo’.

A questão do desenvolvimento ganhou novos contornos e vieses pouco depois, em dezembro de 1964, com a criação da Conferência das Nações Unidas sobre Comércio e Desenvolvimento – UNCTAD. Esta organização foi criada com o objetivo de tentar preencher a lacuna não coberta pelas instituições surgidas no pós-guerra.

Outro marco internacional para a questão do desenvolvimento se dá em 1968, com a criação do chamado Clube de Roma, reunindo profissionais de diversas áreas e de diferentes países na formação de uma livre associação de cientistas, tecnocratas e políticos que tinham o objetivo de refletir e procurar soluções para diversos problemas mundiais. Esse novo fórum publicou em 1972 um relatório denominado Limites do Crescimento, em que reconhece a finitude dos recursos naturais e a seriedade dos problemas que já então assolavam o meio ambiente comprometendo a sobrevivência humana na Terra. Dessa forma, ganham espaço e legitimidade algumas das ‘novas’ dimensões do fenômeno desenvolvimento.

Ainda naquele ano, na Conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento – CNUMAD, realizada em Estocolmo, o crescente movimento internacional em torno do tema terminou por dar destaque ao problema da possível falta de continuidade do abastecimento de energia no planeta.

Na década seguinte, a ação dos países centrais ampliou o domínio das questões financeiras na agenda internacional do desenvolvimento através da promoção de idéias e práticas neoliberais de exaltação do mercado e redução do estado. Como consequência, o

⁸ Os europeus que colonizaram a África denominaram por cooperação técnica ou cooperação cultural, as ações civilizatórias dirigidas às ex-colônias na fase de descolonização.

‘terceiro mundo’ volta a perder espaço na agenda política mundial e seus pleitos de desenvolvimento já não encontram o mesmo eco nas prioridades dos países de economia mais desenvolvidas.

De modo geral, as instituições de Bretton Woods, a Organização para a Cooperação e o Desenvolvimento Econômico – OCDE e a Organização Mundial do Comércio ganham relevância em relação a UNCTAD, que pretendeu ser, nos anos 1970, o principal foro de negociação de uma “nova ordem econômica internacional” (ALMEIDA, 1994).

Em 1987, a Comissão Mundial para o Meio Ambiente e Desenvolvimento publica o relatório denominado *Nosso Futuro Comum*, instrumento utilizado para difundir o recém criado conceito da ‘sustentabilidade do desenvolvimento’, também conhecido como Relatório Brundtland.

Já na década de 1990, a Conferência das Nações Unidas sobre o Meio Ambiente e Desenvolvimento (Eco-92), realizada no Rio de Janeiro, trouxe o tema de volta à agenda política internacional. Tratava-se, agora, de avançar com um conceito de desenvolvimento bem mais amplo, agregando a preservação ambiental e o uso racional dos recursos naturais ao imperativo da resolução dos problemas sociais e do crescimento econômico. Nessa conferência, apesar dos conflitos de interesses presentes, foi elaborado um programa de ação de longo prazo e amplo potencial de impacto internacional: a Agenda 21. Essa agenda lança algumas bases para se praticar, em termos globais, o ‘novo’ conceito de desenvolvimento sustentável.

Os resultados da Eco-92 desdobraram-se em termos de instrumentos de mensuração a partir de agosto de 1994, com a realização da Conference and Workshop on Indicators of Sustainability – CWIS. Segundo HART (1994), embora já existissem algumas iniciativas de concepção de metodologias para aplicação de instrumentos de mensuração, a CWIS possibilitou alguns avanços entre os quais se destacam: a) reconhecimento de diferentes definições de sustentabilidade; b) necessidade de uma metodologia adequada para o uso efetivo de indicadores de sustentabilidade e c) constatação dos interesses governamental e privado no tema.

Ainda naquela década, instituições como o World Resources Institute e o World Bank fomentaram a divulgação da ‘economia verde’ (*green national accountig*), sugerindo alterações metodológicas, que incorporassem no Sistema de Contas Nacionais e

consequentemente no cálculo do PIB, variáveis relacionadas com a utilização dos recursos naturais (matas, rios, lagos etc.). Para medir a ‘economia verde’, em 1995, o World Bank realizou uma avaliação de 192 países e concluiu que o capital físico, também medido pelo PIB convencional, corresponde, em média, a apenas 16% da riqueza total produzida por estas nações. Já o capital humano, chega a alcançar 64% e o capital natural, também em média, representa 20% do total. A partir dessa perspectiva, segundo Marzall e Almeida (1999), o PIB de alguns países europeus passou a refletir a diminuição dos recursos naturais e os custos da poluição gerada.

Entre o final da década e o início do novo milênio, a ‘globalização’ tornou-se um fenômeno mais importante nos debates e reflexões sobre problemas internacionais de ordem política, econômica, social, cultural e ambiental. Entre outros fatores, a extraordinária expansão do conteúdo informacional na vida cotidiana de muitas sociedades passou a vigorar como um padrão hegemônico e determinante da chamada Sociedade da Informação e do Conhecimento (CASTELLS, 2000), favorecendo a proliferação da ideia ocidental de desenvolvimento, assumindo uma nova complexidade e incitando desafios para os pesquisadores do campo.

Com o objetivo de elucidar algumas questões contemporâneas acerca do conceito de desenvolvimento, o próximo item traz algumas definições relevantes para, em seguida, apresentar uma análise estruturada das atuais correntes predominantes na abordagem do desenvolvimento.

2.3 ACEPÇÕES CONTEMPORÂNEAS DE DESENVOLVIMENTO

A obra seminal de Rostow (1971) recupera o darwinismo social, para explicar o construto desenvolvimento como um processo de sucessão evolutiva em estágios, onde as sociedades humanas partem de um modelo rudimentar até chegarem em um modelo de civilização ocidental industrializada de consumo, considerada única e universal. Na abordagem de RIBEIRO (2005), o sentido da palavra desenvolvimento, em várias áreas do conhecimento, converge para “um estado, processo, bem-estar, progresso, crescimento econômico, crescimento humano ou equilíbrio ecológico”. Já Comissão SUD (1990) define o termo como “um processo de amadurecimento e desenvolvimento da personalidade que libera a população do medo e da exploração”. Ambas as compreensões, parecem divergir e avançar

em relação à concepção de estudiosos mais tradicionais do tema para os quais, o desenvolvimento usualmente é visto como um fenômeno que interessa fundamentalmente aos países desenvolvidos (RIBEIRO, 2005).

Uma visão institucional internacional como a do PNUD (1991), por outro lado, traz ainda, uma outra interpretação do que seja desenvolvimento: “ampliar a gama de escolhas à população que permita tornar o desenvolvimento mais democrático e participativo. [...] acesso a (*sic*) renda [...] participação nas decisões e gozo de liberdades humanas, econômicas e políticas”. Segundo Souza (2009), esta abordagem se configura como uma visão tecnocrática de desenvolvimento que decorre de ações planejadas e ordenadas realizadas por governos e organizações de grande influência e poder.

Essas três definições até aqui apresentadas estão vinculadas a modelos identificados a partir das três correntes predominantes na abordagem do termo teórico ou construto desenvolvimento. A primeira, e mais usual, trata o termo como sinônimo de crescimento econômico. A segunda nega que o termo seja apropriado e reza que seu emprego não passa de reles ilusão, manipulação ideológica, crença ou mito. A terceira, mais difícil e desafiadora via, porém, “tenta explicar que o desenvolvimento nada tem de quimérico e nem pode ser amesquinhado como crescimento econômico” (VEIGA, 2004, p.17). Descreve-se sucintamente, a seguir, cada uma dessas escolas de pensamento.

a) Desenvolvimento como Crescimento Econômico

Na perspectiva econômica, não se pode deixar de mencionar Schumpeter (1911, 1939, 1942) que utilizou o termo desenvolvimento como ‘evolução’, ‘desdobramento’, ‘revelação’, ‘inovação’.

Para essa escola parece não haver necessidade de uma teoria do desenvolvimento, bastando aplicar ao conceito a economia moderna, disciplina anistórica e universalmente válida. Conforme Veiga (2005, p.80): “Os fundamentalistas de mercado implicitamente consideram o desenvolvimento como algo redundante. O desenvolvimento viria como decorrência natural do crescimento econômico graças ao efeito cascata (*trick down effect*)”.

Até algumas décadas atrás, alguns dos principais representantes da escola de pensamento econômico consideravam que o crescimento seria capaz de promover o

desenvolvimento⁹ da sociedade e com isso, supostamente, solucionar direta ou indiretamente os problemas provenientes das mais variadas dimensões, principalmente os relacionados com a diminuição da pobreza de nações e regiões. Diante da impossibilidade do controle populacional, políticas de crescimento seriam uma alternativa para o combate à pobreza. Cogitadas há séculos (SMITH, 1988; RICARDO, 1978), essas políticas ainda encontram respaldo na literatura recente sobre desenvolvimento (PRESTON, 1996; BUSTELO, 1998) e baseiam-se na idéia de que uma economia, à medida que cresce, torna maior a disponibilidade de recursos para a população beneficiando assim os mais pobres.

Os autores que fazem parte desta corrente de pensamento defendem a idéia de que em sociedades modernas urbanizadas, boa parte do bem-estar está associada à renda de que as pessoas dispõem para ter acesso a bens e serviços a serem adquiridos no mercado. Posicionamento que reflete diretamente na formulação das políticas sociais e ambientais. Entretanto, a despeito da importância teórica dessa linha de pensamento o próprio Banco Mundial reconhece que a obtenção do crescimento econômico por si só, não garante a diminuição da pobreza.

b) Desenvolvimento como Ficção

No discurso dos autores que fazem parte dessa segunda corrente está presente a crítica de que o desenvolvimento não pode ser reduzido a uma compreensão quantitativa do mundo. RIVERO (2002) defende a idéia de que aqueles que “acreditam no desenvolvimento” ignoram os processos qualitativos histórico-culturais, o progresso não-linear da sociedade, as abordagens éticas, e até prescindem dos impactos ecológicos.

Outro ponto que caracteriza os autores que fazem parte desta cética corrente está relacionado ao fato de acreditarem que existe uma reduzida, ou quase nula, possibilidade de os países periféricos ascenderem na rígida hierarquia capitalista mundial.

Segundo Rist (2001), cabe uma reflexão crítica à visão econômica do desenvolvimento ao considerar que as sociedades fazem parte de uma mesma espécie, “cujo tempo transcorre em um mesmo ritmo para todas elas”: Como seria possível uma sociedade que iniciou seu desenvolvimento tardiamente, alcançar aquelas que já se encontram em um estágio avançado de desenvolvimento?

⁹ Leia-se progresso

As críticas quanto à utilização do termo desenvolvimento convergem assim no entendimento de que imaginá-lo como uma simples representação assintótica do crescimento não é apropriado, uma vez que esse raciocínio relaciona de forma simplista um fenômeno natural a um fenômeno social. Rist (2001) ratifica dizendo que, ao invés de facilitar a compreensão do fenômeno, a metáfora o obscurece, “naturalizando a história”. Seria como se cada lugarejo estivesse destinado a tornar-se uma grande cidade ou cada espécie animal estivesse destinada a transformar-se em uma outra espécie, já que esta última apresenta maior sucesso evolutivo.

Com o triunfo da perspectiva Darwinista no século XIX o Evolucionismo Social vem se inserir nesse debate até que se estabelece, já no século seguinte um novo paradigma no qual a sociedade teria a possibilidade de alcançar sua ‘perfeição’. A aceitação deste paradigma no imaginário coletivo aparentemente implicou, entre outras coisas, na aceitação de uma suposta primazia da superioridade ocidental sobre as outras civilizações. Desta forma, no cerne desta concepção estaria a idéia de que existe uma ‘história natural da humanidade’ e que, portanto, o ‘desenvolvimento’ das sociedades, da riqueza e do conhecimento corresponderia a um princípio ‘natural’ auto-dinâmico¹⁰.

A aproximação dos teóricos do desenvolvimento com as idéias darwinistas permitiu, no plano político, legitimar a colonização em curso principalmente na África e na Ásia do fim do século XIX, época em que este movimento tinha a justificativa de proporcionar às sociedades colonizadas a condição de se tornarem avançadas conforme os critérios estabelecidos por este processo de dominação. Os benefícios dessa intervenção materializavam a idéia de que uma mesma ‘natureza humana’ implica em uma mesma ‘natureza social’.

Em direta oposição a esta visão, Furtado (2001), entre tantos, defende a proposição de que os países dependentes serão sempre subdesenvolvidos: “a transição do subdesenvolvimento para o desenvolvimento é dificilmente concebível, no quadro da dependência”.

Corroborando a idéia dos pensadores desta corrente, Veiga (2005, p.79) afirma que: “A renúncia à idéia de desenvolvimento deve-se ao fato de ter funcionado como armadilha ideológica inventada para perpetuar assimétricas relações entre as minorias dominantes e as maiorias dominadas, nos países e entre os países”.

¹⁰ Para Furtado (2001), o desenvolvimento, visto nesta perspectiva, é um mito, uma fantasia, algo inalcançável no arcabouço de um sistema que destrói recursos naturais, aquele que agrava disparidades de renda e tende ainda a produzir uma homogeneização cultural danosa.

Diante da ‘rejeição à crença no desenvolvimento’, parte dos autores desta corrente prefere o uso do termo ‘desenvolvimento econômico’ em detrimento do termo mais abrangente.

Talvez como fruto da evidente discordância dessas correntes, surge uma terceira escola discutida à seguir.

c) Via “alternativa”

A terceira via reúne parte dos autores da corrente cética que acredita que o conceito multidisciplinar de ‘desenvolvimento’ possui complementaridades com a idéia unidisciplinar de ‘desenvolvimento econômico’. Veiga (2005, p.81), por exemplo, considera que Furtado tem a melhor concepção do conceito:

O crescimento econômico, tal qual o conhecemos, vem se fundando na preservação dos privilégios das elites que satisfazem seu afã de modernização; já o *desenvolvimento* se caracteriza pelo seu projeto social subjacente. Dispor de recursos para investir está longe de ser condição suficiente para preparar um melhor futuro para a massa da população. Mas quando o projeto social prioriza a efetiva melhoria das condições de vida dessa população, o crescimento se metamorfoseia em desenvolvimento.

Para essa corrente, embora ‘crescimento econômico’ possa ser transformado em ‘desenvolvimento’, não existe uma igualdade semântica entre os termos. Para que o ‘desenvolvimento’ ocorra, é necessário que o crescimento econômico seja pensado no escopo de um projeto social voltado para o bem-estar da sociedade.

Em face da fragmentação das Ciências Sociais e da lacuna teórica existente nas distintas abordagens do desenvolvimento, imputa-se comumente ao conceito três princípios básicos: eficiência, equidade e liberdade. Cada uma dessas finalidades, ainda que conceitualmente clivadas, estão associadas a três planos distintos: a ‘eficiência’ visa à dimensão econômica, a ‘equidade’ à dimensão social e a ‘liberdade’ ao fim político (UNICAMP, 2005).

O que se percebe, portanto, é que para entender desenvolvimento como desejam os defensores da “via alternativa”, é necessário um olhar integrado e multidisciplinar. O estudo do fenômeno requer um novo paradigma e, portanto, a construção de categorias analíticas novas, que integrem o aporte de outros campos da ciência.

Esta seção **apresentou as principais diferenças e similitudes nas abordagens teóricas sobre o construto desenvolvimento e reunido-as por corrente de pensamento**. A seção seguinte provê uma breve descrição dessas correntes teóricas à luz do ainda emergente conceito da sustentabilidade.

2.4 DESENVOLVIMENTO SUSTENTÁVEL

O termo desenvolvimento sustentável surgiu no final da década de 70 e foi consagrado em 1987 pela já mencionada Comissão Brundtland. Essa comissão elaborou a mais amplamente aceita definição de desenvolvimento sustentável:

O desenvolvimento sustentável é um processo de transformação no qual a exploração dos recursos, a direção dos investimentos, a orientação do desenvolvimento tecnológico e a mudança institucional se harmonizam e reforça o potencial presente e futuro, a fim de atender às necessidades e aspirações futuras [...] é aquele que atende às necessidades do presente sem comprometer a possibilidade de as gerações futuras atenderem as suas próprias necessidades (BECKER, 1993, p.49).

Esta definição tornou-se, desde então, ponto de partida para o marco teórico inspirador dos trabalhos subseqüentes (MATOS, 1997).

Em função do aumento do interesse político mundial nas questões ambientais e do acirramento dos conflitos sócio-ambientais, houve um aumento da motivação para o debate sobre o tema nas últimas décadas. Como argumenta Sachs (2004), é nesse contexto que surge a proposta de um Desenvolvimento Sustentável como alternativa desejável – e possível – para promover a inclusão social, o bem-estar econômico e a preservação dos recursos naturais.

Embora tenha existido uma ampla aceitação do termo ‘sustentável’ na comunidade internacional, é fato, também, que não existe um consenso em torno de uma definição clara e única para o conceito. Sob uma perspectiva ambiental, a noção de sustentabilidade “se origina de teorizações e práticas ecológicas que tentam analisar a evolução temporal de recursos naturais, tomando por base a sua persistência, manutenção ou capacidade de retorno a um presumido estado de equilíbrio, após algum tipo de perturbação” (RAYNAUT, LANA E ZANONI, 2000, p.74).

Segundo Raynaut (1997, p.370), a utilização dessa noção pode levar à interpretação de uma história que não comporta outro movimento que não “a reprodução incomensurável de um equilíbrio impossível” da natureza e da sociedade.

Veiga (2005) reúne as respostas à questão do que é sustentável em três grupos distintos: 1 - Os que acreditam que as teorias econômicas podem ser aperfeiçoadas para responder a problemática ambiental; 2 - Aqueles que acreditam ser possível conciliar crescimento econômico e conservação ambiental e 3 - Os que estudam as complementaridades entre ‘desenvolvimento econômico’ e o ‘desenvolvimento’.

A ambigüidade e falta de clareza que são atribuídos ao conceito desenvolvimento sustentável na visão de Almeida (1999), dificultam, segundo o mesmo autor, a operacionalização do conceito. Para ele esta dificuldade contribui para uma aceitação mundial e generalizada do princípio de se cultivar um caráter sustentável ao desenvolvimento, o que seria uma impossibilidade teórica em sua percepção. Para outros, a exemplo de Naredo (1987), a sustentabilidade requer meramente um padrão de vida dentro dos limites impostos pela natureza. Utilizando uma metáfora econômica, dever-se-ia, portanto, viver dentro da capacidade do capital natural.

O desenvolvimento sustentável como qualquer outro imperativo econômico convencional, pressupõe a conjugação da maximização da produção econômica com os imperativos sociais (minimização do sofrimento humano atual e futuro) e ecológicos (proteção da ecosfera) (WACKERNAGEL e REES, 1996).

Portanto, para compreender a estreita relação entre desenvolvimento sustentável e as temáticas do crescimento econômico e do meio ambiente, é fundamental recorrer aos objetos de estudo de ao menos três campos de saber:

- a) econômico, social e do comportamento humano;
- b) geografia humana e ciências regionais;
- c) ciências biológicas, físicas e químicas.

Para Veiga (2005, p.187) “É evidente que estes três âmbitos se relacionam, interagem e se sobrepõem, afetando-se e condicionando-se mutuamente”. Corroborando as idéias de Veiga, Sunkel (2001, p.296) considera que “a evolução e transformação da sociedade e da economia no processo de desenvolvimento alteram de várias maneiras o

mundo natural. E esse relacionamento recíproco se materializa, se articula e se expressa por meio de formas concretas de ordenamento territorial”.

É nessa mesma perspectiva que Sachs (2004) explicita alguns princípios básicos dessa nova visão sobre o construto desenvolvimento:

- a) satisfação das necessidades básicas;
- b) solidariedade com as gerações futuras;
- c) participação da população envolvida;
- d) preservação dos recursos naturais e do meio ambiente em geral;
- e) elaboração de um sistema social garantindo o emprego, segurança social e respeito a outras culturas; e
- f) programas de educação.

Outros autores ampliaram a gama de dimensões analíticas do desenvolvimento sustentável. Por exemplo: Darlot (2000) também analisa a questão da sustentabilidade, porém, adiciona a ela a sofisticação de um corte em cinco sub-dimensões:

- Sociocultural;
- Técnico-agronômica;
- Econômica,
- Ecológica e
- Político-institucional.

Por outro lado, o trabalho de Camino e Muller (1993) também aponta para a ampliação do escopo do construto sustentabilidade, nele incluindo:

- Justiça social;
- Viabilidade econômica;
- Sustentabilidade ambiental,
- Democracia;
- Solidariedade e
- Ética.

Em uma outra abordagem Lopes (2001) sugere que no debate da sustentabilidade do desenvolvimento, deve-se recorrer ao estudo multidimensional, enfocando as interligações entre as dimensões social, econômica e ambiental e destas com outras dimensões, tais como política, cultural, institucional e democrática.

2.5 O BEM-ESTAR COMO VARIÁVEL OBJETIVA DO DESENVOLVIMENTO

Como se observa, a maioria das abordagens aqui apresentadas considera a relação e a interação entre as partes no todo e entre o todo e seu ambiente como elemento secundário, embora isto pareça mais evidente naquelas abordagens que supõem a existência dos sistemas social, econômico e ambiental como instâncias separadas e independentes.

A definição das dimensões adotadas na análise dos autores citados é, portanto, bastante idiossincrática, convergindo em alguns pontos e divergindo em outros, sendo as dimensões selecionadas relacionadas e derivadas do contexto e dos objetivos de cada estudo. Assim, na discussão sobre desenvolvimento sustentável, é possível identificar uma clara necessidade de uma abordagem multidimensional, que contemple a operacionalização do conceito enfocando as interligações entre o grupo das dimensões mais tradicionalmente consideradas (social, econômica e ambiental) com aquelas que emergem como fruto da crescente difusão do conceito por distintos campos de conhecimento (política, cultural, institucional, etc.). Dessa forma, deve-se, em princípio, buscar a viabilização das métricas de uma convergência que permita harmonizar as preocupações com o bem-estar das pessoas e do ambiente.

Nesse sentido observa-se que apesar de originalmente utilizado em estudos sobre desenvolvimento humano, é relativamente recente o emprego do termo bem-estar em trabalhos sobre sustentabilidade. Trabalhos como os de Hodge et al. (1995), Imbach et al. (1999) e Prescott-Allen (1999; 2001) trazem como hipótese subjacente a idéia¹¹ de que o desenvolvimento sustentável é aquele que ocorre através de uma combinação do bem-estar humano e bem-estar do ecossistema. As pessoas vivem e fazem parte do ecossistema e, conseqüentemente, a análise do bem-estar de ambos não pode ser feita separadamente. Isto é,

¹¹ Para Souza (2009) essa idéia pode ser entendida como um artefato, uma conclusão enganosa derivada da ineficácia do método eleito, pois não passa visão processual, sistêmica e modelar.

o ambiente natural não faz sentido sem as dimensões humanas (político, econômico e social), assim como o bem-estar da humanidade não pode ser desassociado do ambiente.

O termo bem-estar, empregado no âmbito do desenvolvimento sustentável, é tratado por Holtz (1995) a partir das dimensões: Econômica, Desenvolvimento Humano, Qualidade de Vida, Saúde e Felicidade. As três primeiras dimensões possuem uma vasta produção literária e indicadores mundialmente aceitos, a exemplo de: i) Renda; ii) Índice de Desenvolvimento Humano – IDH e iii) Índice de Qualidade de Vida – IQV. Já as dimensões Felicidade e Saúde são tratadas pelo Autor de maneira subjetiva.

O discurso de bem-estar, por sua vez, parece manter certa afinidade com a análise da pobreza proposta por SEN (1998) no enfoque das capacidades. Neste enfoque, o Autor assinala que é relevante que as pessoas disponham de certos meios para alcançar seus objetivos de bem-estar. Defende também, a idéia de que esses meios podem ser condições físicas de vida, capital humano, liberdade política, direitos sociais, auto-respeito e riqueza.

Quem estiver privado dos meios considerados básicos é classificado, portanto, como pobre porque não é capaz de satisfazer suas necessidades se assim o desejar. Este critério valoriza a importância da liberdade das pessoas para decidir a utilização dos meios de que dispõe do modo como lhes convier. As críticas a esse modo de considerar a pobreza recaem basicamente sobre duas dificuldades: definir o que sejam as capacidades mínimas e avaliar a extensão da liberdade necessária para sua utilização. Para Ribeiro (2005, p.6) “a sedução implícita na discussão do desenvolvimento está na crença da possibilidade de eliminar a pobreza”.

Diante das acepções de desenvolvimento e sustentabilidade apresentadas, pode-se dizer que sustentável é o desenvolvimento que propicia ou permite o alcance da condição de manutenção harmônica do bem-estar do homem (econômico, social e político) e do ambiente (ecossistema e espaço). Com base nesse referencial e reconhecendo a importância da aderência dos indicadores ao conceito, bem como a relevância de uma abordagem multidimensional para o fenômeno, apresentam-se criticamente no próximo capítulo os principais indicadores e índices de desenvolvimento disponíveis na literatura.

2.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

A discussão apresentada neste capítulo, inspirada em algumas reflexões sobre as Teorias do Desenvolvimento, buscou oferecer uma abordagem baseada no diálogo com escolas de pensamento distintas. Esse exercício permitiu identificar as principais diferenças e similitudes nas abordagens sobre desenvolvimento e reuni-las por correntes ou escolas de pensamento. Constatou-se, também, que as escolas convergem ao aceitar o substantivo bem-estar como o estado desejável de harmonia, tanto para o ambiente como para o homem.

Foi possível observar que para estudar desenvolvimento é necessário um olhar integrado e não a simples superposição de percepções disciplinares – a abordagem interdisciplinar. De fato, a complexidade do fenômeno do desenvolvimento e, particularmente, do desenvolvimento sustentado, exige a construção de categorias analíticas novas que integrem o aporte de diferentes campos de conhecimento.

A revisão teórica realizada neste capítulo deixa claro que, a despeito da ampla discussão presente na literatura, não há consenso sobre termo teórico ou construto que melhor define o conceito intangível desenvolvimento ou mesmo sobre o desenvolvimento sustentável. As diversas iniciativas voltadas à mensuração do fenômeno motivaram a formulação de um dos objetivos deste capítulo que é analisar o termo teórico desenvolvimento e seus desdobramentos.

Atualmente, os trabalhos¹² que tratam à mensuração do desenvolvimento por meio de indicadores e índices estão, em grande parte, associados ao monitoramento da mudança social e de um conceito quase que coloquial de bem-estar. Destarte, os indicadores de desenvolvimento têm por finalidade não somente mostrar, de forma bastante precisa, a relação entre os fatos e as mudanças, mas também subsidiar o monitoramento dos processos de transformações de natureza econômica, política, social e ambiental, entre outras.

‘Desenvolvimento’ e ‘sustentabilidade’, como se viu, são conceitos subjetivos e de vital importância à sociedade mundial contemporânea. Daí decorre a necessidade de novos indicadores que sejam mais do que simples indicadores unidimensionais (ambientais, econômicos, sociais e políticos, isoladamente). De modo similar, indicadores de desenvolvimento sustentável devem ser mais do que indicadores de crescimento. Devem

¹² A relação dos 64 trabalhos nacionais e 397 trabalhos internacionais estudados nesta tese está disponível nos apêndices A e B.

expressar eficiência, suficiência, equidade e qualidade de vida. Crescimento interpretado na ótica da maior parte dos estudiosos leva apenas a ‘ter mais’, não necessariamente em viver melhor, em viver com bem-estar social ou ambiental.

A construção de indicadores, por sua vez, não envolve apenas a interpretação dos conceitos e das estatísticas, mas requer o enfrentamento de sua problemática sob uma ótica multidisciplinar. A elaboração de técnicas de análise e mensuração do desenvolvimento trará sempre polêmicas em virtude do caráter crítico que lhe é pertinente. Deve-se, porém, lutar por uma utilização mais madura dessas análises, seja por parte dos poderes públicos, seja por parte da imprensa ou mesmo do público em geral.

Por outro lado, para a mensuração do desenvolvimento, é importante o uso de conceitos que possam dar conta de sua condição subjetiva, complexa e, como já dito, multidimensional. Este trabalho de mensuração não prescinde do conhecimento da realidade social, econômica, política e ambiental, bem como, da adoção de critérios e valores adequados para cada situação. Independentemente das polêmicas políticas e midiáticas, as discussões sobre desenvolvimento constituem um campo importantíssimo, difícil e controverso, aberto a múltiplos entendimentos e concepções. Além da multiplicidade de percepções sobre o fenômeno, o campo de estudos abrange, também, vários modos de mensurar e tratar as manifestações do desenvolvimento, os quais devem ser adequados às distintas realidades e às opções conceituais que estiverem sendo consideradas. Sachs (2004), por exemplo, alerta que, durante as tentativas de mensuração, devem ser observadas diversas questões de natureza conceitual e metodológica.

É importante salientar que os indicadores de desenvolvimento devem ser vistos não apenas como instrumentos conjunturais e de uso eventual ou figurativo, mas como elementos indispensáveis para fundamentar a tomada de decisão nas esferas governamental e pública.

Nos últimos anos, têm surgido por todo o mundo iniciativas e projetos com vistas à definição de indicadores de desenvolvimento para um variado leque de finalidades de gestão, no âmbito dos desenvolvimentos local, regional e nacional. A abordagem multidimensional e interdisciplinar, fundamentada nos autores estudados, possibilitou a identificação de dimensões comuns de análise nestas iniciativas que nortearão a avaliação de aderência dos indicadores estudados à representação do processo de mudança social em busca de um estado de bem-estar.

Ao fim deste capítulo, é possível vislumbrar possíveis desdobramentos que certamente culminarão em uma etapa mais propositiva para a discussão aqui apresentada. Este estudo mostrou que a investigação sobre índices e indicadores de desenvolvimento não se encerra em fórmulas, técnicas ou abordagens quantitativas. Com base em um debate conceitual, os capítulos 4 e 7 desta tese oferecem a análise do conjunto de indicadores que são adotados pelos principais índices de desenvolvimento existentes.

Por fim, foi evidenciado que um bom indicador de desenvolvimento dificilmente será fácil de ser entendido. Para que possam descrever a complexidade do fenômeno, certamente os indicadores e índices terão que ter uma complexidade interna correspondentemente alta. Diante disso, parece necessário que cientistas das áreas social, econômica e ambiental, apoiados na infra-estrutura existente nas suas respectivas ciências, assumam a responsabilidade por sua construção, bem como da comunicação dos resultados.

3 ASPECTOS TEÓRICO-METODOLÓGICOS DA MENSURAÇÃO DO DESENVOLVIMENTO SOCIAL

Essentially, all models are wrong, but some are useful. George E.P. Box (1959)

Este capítulo tem o objetivo de **apresentar uma proposta de estruturação dos conceitos e metodologias para elaboração de índices e sistemas de indicadores sociais** que possam mensurar conceitos intangíveis e relevantes de um sistema social. Para consecução desse objetivo foram estudados os aspectos teórico-metodológicos da mensuração de fenômenos sociais e sua conexão com as principais propostas de formulação de Índices Sociais disponíveis na literatura, a fim de subsidiar a construção de categorias analíticas para o estudo dos instrumentos de medição do desenvolvimento, que é apresentada no Capítulo 4.

3.1 TERMOS TEÓRICOS, CONSTRUTOS, CONCEITOS E INDICADORES

Esta tese trata da mensuração de conceitos intangíveis relacionados com o fenômeno Desenvolvimento Social. Estas mensurações, por sua vez, podem assumir papéis relevantes em sistemas de indicadores, principalmente no que tange à influência no nível e também ao tipo de desenvolvimento medido.

A discussão sobre conceitos intangíveis tem sido tema recorrente em trabalhos voltados para elaboração de índices ou sistemas de indicadores na pesquisa social, pois os resultados obtidos por estes instrumentos têm sido amplamente utilizados como ferramenta indispensável para gestão pública nas três esferas de governo, no monitoramento e avaliação de programas de governo e na análise de políticas públicas, entre outras aplicações.

Como o acesso a esses conceitos intangíveis não é unívoco, faz-se necessário identificar os conceitos a eles subjacentes, até chegar a conceitos operacionais (indicadores) de mensuração direta, ou, ao menos, indiretamente aferíveis. Para subsidiar esta abordagem recorre-se à teoria das medições, mais especificamente aos elementos que a compõem, visto que podem ser úteis na mensuração de conceitos intangíveis. Entre os diferentes processos discursivos, isto é, modos de raciocínio que orientam a pesquisa científica a expor

rigorosamente seus resultados, a teoria das medições tem uma relevante participação, sobretudo nas ciências sociais, pois é um campo de conhecimentos que, segundo French (1986), trata de modelos numéricos capazes de estruturar as características qualitativas de fenômenos sociais.

O uso de conceitos que rompam com os limites das estruturas teóricas leva a enriquecer a análise das teorias por um ângulo diferente e, em certa medida, novo (ZEMELMEN, 1992, p. 208). Nessa perspectiva, não basta distinguir entre as teorias falsas ou verdadeiras, tendo em vista que se trata de uma mesma perspectiva de explicação; é preciso compreender que se faz necessário examiná-las desde a problemática da separação entre as funções teóricas e epistemológicas.

Bruyne et al. (1977, p. 29) definem teoria como um artefato de ordem simbólica, construída especificamente para esse fim, comporta conceitos que se referem aos fenômenos e possui caráter semântico. Os conceitos têm o papel de articular outros conceitos, possuindo, assim, um caráter sintático. Esses níveis semânticos e sintáticos são indissociáveis de um sistema teórico. Já a função analítica é operatória e serve de enlace entre proposições de tipo sintáticas, que constituem a forma lógica que a hipótese assume para submissão a teste empírico.

Assim, as teorias podem ser vistas não apenas sob o aspecto da falseabilidade, mas também por sua capacidade para delimitar campos de observação. Além da falseabilidade, Zemelman (1992) propõe reavaliar a abordagem dos corpos teóricos, segundo sua capacidade para se decompor em construtos¹³, conceitos e indicadores, segundo as exigências do conjunto. Por esta via se determina um novo marco de referência para avaliar a qualidade das teorias decompostas, enquanto outras, não.

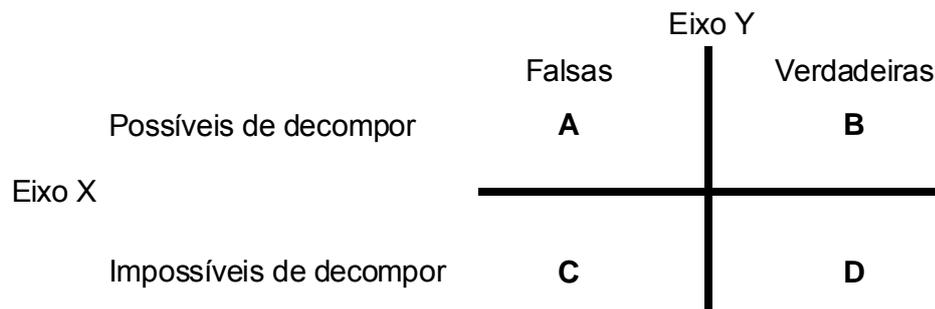
Nessa perspectiva, a combinação do critério verdade-falsidade com a capacidade de decomposição dos corpos teóricos em indicadores para delimitar campos de observação, que sirvam de base para a construção de teorias-explicativas, pode ser observada na Figura 1, a seguir.

Como pode ser observado na Figura 1, as teorias são possíveis de decompor, podendo ser falsas ou verdadeiras e não são possíveis de decompor, sendo falsas ou verdadeiras. É muito provável que nas Ciências Sociais o tipo “A” (Figura 1) seja o mais

¹³ Maior subdivisão de um termo teórico (Teoria). O termo construto é empregado nesta tese para definir um conjunto teórico que age como um “bloco construtor” podendo representar um conjunto de conceitos simples e conceitos complexos.

frequente, enquanto o “B” seja o tipo ótimo e o tipo “C” é nulo. Por fim, o tipo “D” é aquele que, apesar de ser verdadeiro, não contribui para enriquecer as formas de raciocínio para além da natureza restrita do seu teste. As teorias passíveis de decomposição e falsas (Tipo “A”) fornecem maior possibilidade de explicação do fenômeno Desenvolvimento Social e contribuem para enriquecer as formas de racionalização e podem ser submetidas à prova.

Figura 1 – Tipos de teorias segundo a possibilidade de decomposição e o critério de falseabilidade



Zemelman (1992) considera que os tipos A e B, cujos elementos conceituais são possíveis de decompor, constituem a base para o processo de acumulação de conhecimento científico, porque a possibilidade de decomposição contribui para o enriquecimento do que o autor chama de “razão cognitiva antes da expansão do acervo teórico”.

Há na teoria, pelo menos, dois aspectos fundamentais que se destacam e se distinguem analiticamente: o aspecto *conceitual* que explica o sentido e o aspecto *proposicional* relativo à formulação lógica. A formulação lógica, por sua vez, significa o aspecto sintático “assumidos pelos sistemas teóricos como articulações de proposições segundo regras de derivação lógica” (BRUYNE, 1977, p.110). A *formulação* é o pré-requisito da testabilidade da teoria e obedece ao princípio de redução. Em compensação, a *explicitação* corresponde ao aspecto significativo dos sistemas teóricos e manifesta a dependência de toda teoria para com sua problemática.

Essa distinção analítica, como demonstra Bruyne (1977), permite mostrar a articulação interna dos elementos da teoria e sua dependência com os polos epistemológico, morfológico e técnico em sua feição mais dinâmica¹⁴. Neste capítulo são considerados os polos epistemológico e morfológico.

Nessa perspectiva, uma teoria funda um corpo de hipóteses, com vistas à prova experimental e, é na sistematização da teoria que se encontra sua expressão mais alta. De modo que o caráter sistêmico da teoria é que revela sua força demonstrativa e operatória. Estas são consideradas como a melhor garantia da eficácia e da compreensão científica de um sistema social. Com efeito, a teoria das medições referida nesta seção, se vale de modelagem e técnicas estatísticas de análise multivariada, que consiste na atribuição a dimensões, a propriedades ou qualidades a certos conceitos, de uma ordem de natureza classificatória, serial, quase serial ou métrica e estabelece correspondências entre as dimensões de cada conceito e números disponíveis, segundo determinados padrões.

Essas dimensões (construtos) podem ser decompostas até o nível dos conceitos operatórios (indicador) que podem ser traduzidas em um jogo de valores possíveis, expressos sob a forma de escala que irão constituir o conceito como variável na conotação estatística. “Esta operação, no estágio da coleta das informações, confere a cada indicador da variável um perfil sobre a escala que ele permite observar” (BRUYNE; HERMAN; SCHOUTHEETE, 1977, p. 80-81). Dessa forma, chega-se a uma representação simbólica (Figura 2) de “medida” e o conjunto dessas medidas indicará a aparência de cada dimensão (construto). Ressalta-se que o conteúdo estatístico referente à variável e dimensão neste trabalho ganha outros contornos, como será explicitado na próxima seção.

¹⁴ No campo da prática científica distinguem-se quatro pólos metodológicos: epistemológico, teórico, morfológico, técnico. Embora dispostos desta maneira, “eles não configuram momentos separados da pesquisa, mas aspectos particulares de uma mesma realidade de produção de discursos e de práticas científicas”. Além disso, asseguram a cientificidade das práticas da pesquisa e exercem funções específicas, tais como: o pólo epistemológico (métodos) garante a objetivação, encarrega-se de renovar continuamente a ruptura dos objetos científicos com os do senso comum, decide as regras de produção e de explicitação dos fatos, da compreensão e validade das teorias. Dentre os processos discursivos que orbitam esse pólo cita-se: dialética, fenomenológica, a lógica hipotético-dedutiva; o pólo teórico (quadro de referências) orienta a elaboração de hipóteses e conceitos, propõe regras de interpretação dos fatos determina o movimento da conceitualização, aproxima-se dos “quadros de referência”, os quais desempenham função paradigmática implícita e possuem os seguintes quadros de referência: positivista, compreensivo, funcionalista e estruturalista; o pólo morfológico (quadro de análise) enuncia as regras de estruturação de formação do objeto científico, define certa ordem entre seus elementos o pólo técnico, suscita um conjunto variado de métodos de ordenação dos elementos constitutivos dos objetos científicos: a tipologia, o tipo ideal, o sistema, os modelos estruturais; o pólo técnico (modos de investigação) controla a coleta de dados, “exige precisão na constatação mas, sozinho, não garante exatidão”, tem como referência modos de investigação particulares: estudos comparativos, experimentações e simulação. (BRUYNE et al., 1974)

Figura 2 – Representação do Modelo Teórico Conceitual



Fonte: elaboração própria

Como pode ser observado na Figura 2, a construção mais abstrata é o termo teórico, que compreende os construtos e conceitos, aglutinando em si os aspectos de uma definição suficientemente ampla. Os construtos também são construções abstratas, mais próximas do fenômeno que o termo teórico, mas ainda dele afastadas, consciente e deliberadamente criadas para atender a um propósito científico. Já o conceito é uma construção mais próxima da realidade que o construto. O conceito capta ou apreende fatos ou fenômenos, expressando-os por um sistema semântico, gramatical ou simbólico, de modo a torná-los inteligíveis e processáveis. O conceito não é o fenômeno, mas pode analisar e comunicar as implicações deste. Em resumo, vai-se do fenômeno ao abstrato, partindo-se dos indicadores e passando pelos conceitos, construtos e termos teóricos.

Para a teoria das medições, o vigor da pesquisa baseia-se em técnicas de investigação sobre a causalidade entre objetos de estudo e seus antecedentes. Significa, então, que os procedimentos recomendados nesta teoria compreendem que, ao descrever um sistema social, utiliza-se uma representação numérica de suas relações qualitativas, na qual a magnitude relativa dos números expressa a estrutura de preferências ou crenças de indivíduos sobre o sistema. Por conseguinte, os números podem ser escolhidos de modo arbitrário, desde que haja uma relação racional que permita operações aritméticas dentro da estrutura de preferências. French (1986, p. 77) demonstra que é possível representar por variáveis

numéricas qualquer relação qualitativa que expresse uma estrutura de preferências consistentes.

A explicitação conceitual de uma teoria se faz importante porque manifesta a unidade material que leva a operações referenciais sobre os objetos de investigação. No que tange à medição do Desenvolvimento Social de uma nação, por exemplo, a pesquisa desse fenômeno exige considerar a avaliação de conceitos intangíveis. Por tais condições, vale mencionar que toda mensuração de Desenvolvimento porta conceitos intangíveis e suas relações que refletem, em certa medida, a estratégia de desenvolvimento da nação. Entretanto, além de observar esse aspecto de relevância teórica, cumpre destacar que, como foi discutido no Capítulo 2, o tema Desenvolvimento é recorrente na agenda dos Governos dos países ocidentais.

Como se discutiu no capítulo anterior, as formulações teóricas recentes sobre o desenvolvimento convergem para a noção de que se trata de um fenômeno multidimensional e complexo¹⁵. A noção de complexidade leva a um outro aspecto que o relaciona ao que os epistemólogos denominaram de mudança paradigmática na ciência ou corte epistemológico. Aprofundar-se nesta discussão foge ao escopo desta tese, por tal motivo detêm-se a pensar o Desenvolvimento Social a partir da perspectiva das formas de mensuração apresentadas nos Índices e Sistemas de Indicadores, a ser tratada no próximo capítulo.

Os componentes ou fatores (sociais, políticos, econômicos e ambientais) associados ao conceito de desenvolvimento manifestam-se, simultaneamente, articulando-se entre si e influenciando nas condições do bem-estar das pessoas e do ambiente. Sob esse ponto de vista, percebe-se que o Desenvolvimento medido por Índices e Indicadores de Desenvolvimento parece situar-se numa teia de relações que o afetam e que condicionam suas características. Dessa forma, para que um Índice ou Sistema de Indicadores Sociais seja considerado satisfatório, é esperado que a formulação do objeto contemple a concepção relacional, nos seus aspectos teóricos ou nos metodológicos.

¹⁵ Entende-se complexidade como “uma propriedade de alguns fenômenos que não podem ser compreendidos por meio de simplificações reducionistas, tão caras à ciência positivista tradicional. Simplificar, no caso, significa isolar, compreender por meio de uma única disciplina e extinguir da análise as relações que constituem o objeto” (ESPÍRITO-SANTO; OLIVEIRA; T. RIBEIRO, 2008).

3.2 MODELOS MATEMÁTICOS E AS CIÊNCIAS SOCIAIS

A Modelagem Matemática pode ser definida como a arte de transformar situações do meio circundante em modelos matemáticos (BIENBEMGUT, 1997, p. 34). Medina ET al. (2005, p.43) definem Modelização¹⁶ como sendo a representação “de um objeto, seus movimentos e suas relações, tendo como referência uma construção teórica: a teoria do objeto-modelo”. Por sua vez, Le Moigne (1990) define “modelização” como a ação de elaboração e de construção intencional, pela composição de símbolos, tornando inteligível um fenômeno complexo. Assim, além da esquematização do objeto, é necessário um segundo momento, de agregação da racionalidade que une suas múltiplas dimensões ou aspectos. Bunge (1974) nomeia este processo de construção de modelo teórico.

Swetz (1992, p.65) define Modelo matemático como uma estrutura matemática que descreve aproximadamente as características de um fenômeno em questão. Para Biembengut (1997, p. 89), um modelo é proveniente de aproximações realizadas para se poder entender melhor um fenômeno e, nem sempre, tais aproximações condizem com a realidade e podem ser definidas como um conjunto de símbolos e relações matemáticas que traduzem, de alguma forma, um fenômeno em questão ou um problema de situação real.

Segundo Bunge (1969, p.541), os conceitos e esboços de teoria se afinam mediante a sua tradução matemática, que tem como meta a construção de um modelo. Esse modelo implicará forçosamente no começo uma grande simplificação, tornando-se, pouco a pouco, mais complexo para melhor representar a realidade.

Bunge (1969) adverte que a matematização não é uma mera tradução da teoria em outra linguagem: a matematização é uma operação que faz parte da construção da teoria. Por trás de uma teoria, por assim dizer, formulada na linguagem vulgar ou semi-científica, pode haver muitas teorias precisas. As vantagens da matematização não se limitam a possibilitar uma máxima precisão. Ela aumenta a potência dedutiva da teoria (ou seja, a capacidade de deduzir novos enunciados); permite contrastações empíricas mais finas; facilita a identificação de inconsistências e a comparação da teoria com outras teorias rivais.

¹⁶ O termo “Modelização” advém da escola francesa, “modélisation”, no livro de Astolfi (2001), remete a Bernard Walliser, *Systèmes et modeles*, Seuil, 1977 e também a Jean-louis Lê Moigne, *La Théorie du système général. Théorie de la modélisation*, Paris, PUF, 2^aed, 1984. Para os autores, “Modelagem” é o termo usado na pesquisa matemática, já “Modelização”, que advém da escola francesa é usada na pesquisa de ensino de Física, generalizando para o ensino de Ciências Naturais.

No contexto desta tese, os modelos matemáticos podem ser definidos como construções teóricas hipotéticas, susceptíveis de matematização, com as quais se pretende representar um setor (área) da realidade, para efeito de estudo e de verificação da teoria. Os modelos podem ser representados por expressões numéricas ou fórmulas, diagramas, gráficos ou representações geométricas, equações algébricas tabelas, e outros.

De acordo com esta definição, Bravo (1981) distingue nos modelos as seguintes características:

- 1º Teórico-hipotéticos: em sua origem não são um conjunto de equações sobre a realidade, senão um conjunto de enunciados teóricos sobre as relações entre as variáveis que caracterizam uma dimensão da realidade. Ademais, são construções hipotéticas, enquanto são elaborações *a priori*, supostas e que ainda não foram testadas;
- 2º Susceptíveis de matematização: embora os modelos não sejam originalmente um conjunto de equações, eles podem expressar as relações entre as variáveis que compreendem mediante um conjunto de equações, que é o que transforma um modelo teórico em modelo matemático;
- 3º Representativos da realidade: os modelos devem constituir representações típicas, exemplos ou imagens da realidade. Estes modelos não constituem, sobretudo nas Ciências Sociais, imagens fiéis e sim, imagens simplificadas da realidade. É impossível representar conceitualmente a realidade de uma maneira perfeita (BOX, 1959). Todos os modelos se limitam a dar uma imagem sintética da realidade, a partir das variáveis mais importantes e que influenciam nos fenômenos sociais.
- 4º Finalidade de estudo e de investigação: são instrumentos de investigação e estudo da realidade, passíveis de teste a partir dos dados obtidos para as variáveis que os constituem. O grau de exatidão com que se ajustam ou representam a realidade.

Por outro lado, os modelos matemáticos, considerando-se o seu caráter hipotético-dedutivo, também podem ser empregados para inferir problemas e derivar consequências desses, que podem ser testadas empiricamente e, eventualmente rejeitadas.

Sabe-se que os modelos também podem constituir instrumentos de estudo teórico, Desde que sejam utilizados como ponto de partida, pode-se desenvolver dedutivamente a partir de seus princípios e postulados toda a teoria sobre a realidade a qual se referem.

3.3 VARIÁVEIS, ESCALAS DE MEDIÇÃO, ESTRUTURAS HIERÁRQUICAS E CIRCULARES

Uma variável é um conjunto de medidas associadas a um conceito, construto, aspecto, propriedade ou fator discernível em um objeto de estudo. A medida pode ser quantitativa – associada a um número – ou qualitativa – associada a uma categoria. As variáveis ligam as hipóteses e os construtos latentes, não-manifestos, ao experimento. É pela alocação de variáveis a hipóteses que os construtos e conceitos de uma teoria podem ser testados.

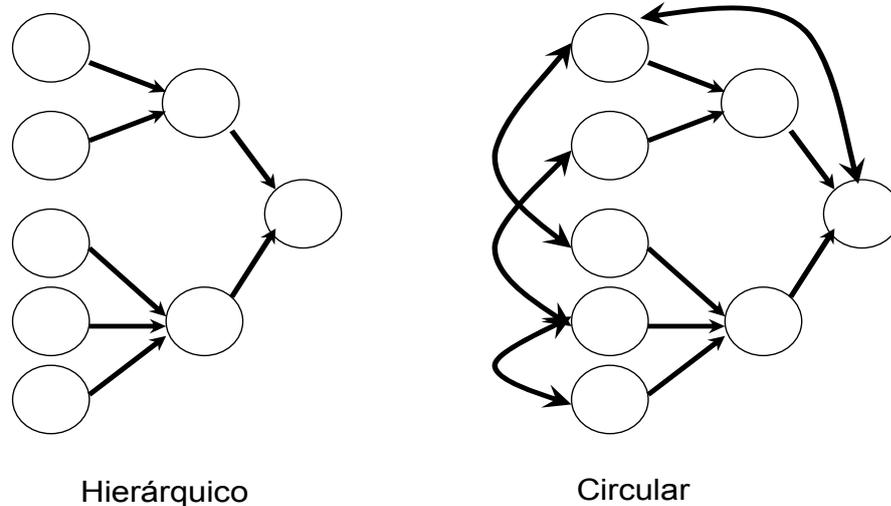
Variáveis admitem relacionamentos, classificações e tipologias. As relações entre as variáveis podem ser: i) assimétricas, nas quais o comportamento de uma variável afeta, condiciona, determina ou causa o comportamento da outra; ii) simétricas, nas quais duas variáveis não se afetam mutuamente, ou por serem indicadores do mesmo conceito, ou por serem efeitos da mesma causa, ou por serem totalmente independentes, mas fortuitamente associadas; iii) recíprocas, surgindo uma circularidade ou tautologia entre ambas.

Conforme a delimitação do objeto, as variáveis podem ser exógenas ou endógenas (fora ou dentro do objeto) e independentes ou dependentes (antecedem ou sucedem o fenômeno). Variáveis exógenas e/ou independentes podem ser associadas a causas ou usadas na predição de fenômenos descritos por variáveis endógenas e/ou dependentes. Variáveis dependentes e independentes podem manter entre si diversos tipos de relações de causalidade. (LAKATOS; MARCONI, 1991).

As variáveis podem ser classificadas em: i) quantitativas discretas, representáveis por números descontínuos; ii) quantitativas contínuas, representáveis por quaisquer números; iii) categóricas nominais, associadas a atributos independentes entre si; e iv) categóricas ordinais, associadas a atributos que mantêm entre si uma estrutura de graduação.

As relações são representadas por um modelo gráfico (Figura 3), o diagrama de caminho (path-diagram), ou por um modelo matemático. Modelos podem usar estruturas hierárquicas, representadas por regressões múltiplas, ou relações circulares, que exigem equações estruturais. Para esta tese, o estudo dos Índices e Sistemas de Indicadores interessam as estruturas hierárquicas.

Figura 3 – Diagramas de Caminho Hierárquico e Circular



Fonte: Elaboração do autor

Estruturas hierárquicas utilizam a lógica da decomposição: um critério mais complexo de mensuração é dividido em subcritérios menos complexos, arranjados de modo que um subcritério só afete o critério ao qual está subordinado, não havendo subordinação múltipla. Os critérios de nível hierárquico inferior devem ser mutuamente exclusivos e devem fornecer uma descrição exaustiva do critério de nível superior. Para cada critério superior deve haver, ao menos, dois critérios inferiores para que se justifique a necessidade de uma estrutura hierárquica (KEENEY, 1992).

A estrutura é construída a partir de uma teoria que separe e classifique o conhecimento sobre o objeto, ou por instrumentos de representação de relações, como mapas cognitivos, diagramas sistêmicos, árvores de realidade ou diagramas de influência. Esses instrumentos têm sido apresentados por diversos autores, com algumas semelhanças: i)

conexões entre causas e efeitos; ii) explicitação dos nexos causais subjacentes às relações de causa e efeito; iii) complexidade crescente; e iv) representação de relações circulares.

Segundo ISA (2007), a busca por indicadores de desenvolvimento tem motivado a construção de iniciativas distintas, visando à sua formulação e sistematização. Nesse sentido, diversos modelos lógicos ou mentais foram desenvolvidos, destacando-se, entre outros:

- i. Modelos com base na economia (modelo contábil) – visam a uma vinculação com contas nacionais. Esse modelo aplica os conceitos e procedimentos de valoração ambiental da economia e requer uma base significativa de informações sobre o inventário de recursos naturais (OCDE, 2002);
- ii. Modelos de vinculação com o bem-estar ambiental e humano (modelo pirâmide) – visam a expressar, a partir da cadeia de indicadores locais, indicadores setoriais de recursos de saída, os indicadores síntese de uma imagem geral do atual caminho, em direção à sustentabilidade (OCDE, 2002);
- iii. Modelos sistêmicos ou de integração de causa e efeito – visam a expressar as relações entre a qualidade ambiental (estado) e os seus fatores causais (de pressão), bem como as iniciativas geradas ou induzidas, em resposta às constatações propiciadas por esses indicadores (OCDE, 2002).

3.4 MÉTODOS TRADICIONALMENTE UTILIZADOS PARA MENSURAR O DESENVOLVIMENTO SOCIAL

O processo de implementação de sistemas de indicadores e índices de desenvolvimento social é complexo e falível. Em primeiro lugar, os dados disponíveis (estatísticas oficiais e registros administrativos), por mais abrangentes que sejam, são retratos parciais e enviesados da realidade, espelham aquilo que a visão de mundo e a formação teórica dos técnicos permitem ver ou priorizam enxergar.

Assim, as soluções visualizadas e as especificações dos Índices estão determinadas, *a priori*, pelas limitações dos dados disponíveis e, em última instância, pelas limitações dos conhecimentos científicos aportados pelas diferentes disciplinas acerca dos fenômenos sociais, fenômenos inerentemente complexos.

Muito esforço e muitas pesquisas vêm sendo aplicados na área de Administração Pública sem a adequada reflexão metodológica e sem o compromisso concreto de tornar o processo de construção de indicadores mais transparente e efetivo.

Visando a contribuir para o entendimento do estudo de instrumentos de mensuração do desenvolvimento social, buscou-se apresentar na subseção seguinte as definições para indicador e índice sociais, contempladas durante a realização desta pesquisa.

3.4.1 Definições

O termo indicador, que vem sendo utilizado cada vez com maior frequência em estudos e análises da realidade social, tem apresentado uma multiplicidade de definições, que variam, inclusive, em função da situação político-social em que são engendradas. Em um breve levantamento histórico, pode-se detectar, *grosso modo*, duas linhas significativas de preocupações. A primeira, desenvolvida sobretudo nos Estados Unidos, com maior ênfase a partir de 1965, ligada a uma preocupação de gerar certos dados sociais, em adição aos dados econômicos. Objetiva medir o que se definiu como bem-estar social da sociedade, envolvendo questões, tais como, saúde, ambiente físico, renda, possibilidades de ascensão social, segurança, educação, ciência e arte, participação social (US DHEW, 1969). A construção dos indicadores seria útil para intervir na realidade, no sentido de eliminar ou amenizar os desvios com relação ao bem-estar assim definido.

A segunda linha desenvolveu-se no âmbito das Nações Unidas, a partir da década de 50, como fruto da preocupação com análises comparativas entre diversas nações, do ponto de vista dos níveis de vida, além da mera comparação em termos econômicos. Houve uma certa variedade de posições entre os técnicos do próprio organismo internacional, devida à heterogeneidade reinante no corpo de profissionais da ONU. A linha mais recente desenvolvida por esse organismo serviu como documentação básica para caracterização, neste texto, do termo indicador social (ONU, 1973).

O termo Indicador diferencia-se do dado estatístico na medida em que é obtido a partir de uma relação estabelecida entre dados primários (médias, taxas, porcentagens, índices, etc.), construído a partir de um modelo conceitual. Este é uma construção intelectual que, partindo da observação do real, baseada em uma concepção de mundo, privilegia, para

estudo, certos aspectos da realidade que considera como essenciais naquele momento histórico.

A produção e o uso de indicadores de forma organizada e sistemática têm, pois, implicações políticas e não apenas metodológicas, na medida em que parte de uma opção baseada em uma concepção de mundo, objetivando intervir na realidade. Os indicadores são instrumentos para o conhecimento e análise das condições e tendências sociais, para a formulação de previsões e de propostas de intervenção na realidade social, e para o próprio acompanhamento e avaliação dos meios e resultados das políticas adotadas.

São usuários dos indicadores a sociedade como um todo, além dos administradores e planejadores. Nesse sentido, os indicadores devem ser disponíveis para qualquer cidadão e formulados de modo inteligível. A razão por que isso se faz essencial: para que as políticas reflitam os reais interesses da sociedade, é necessário que os vários grupos sociais participem em sua formulação e acompanhem a sua execução.

Para as atividades de planejamento, que se definem, sobretudo, pela capacidade de prever para intervir, dentro de um amplo espectro de possibilidades e riscos, os indicadores devem ser construídos de modo a permitir a análise da situação da sociedade, a formulação de políticas e a gestão dos serviços públicos.

Nas discussões “sobre indicadores parece ter-se desenvolvido na literatura o consenso de se usar a expressão **indicadores sociais** para significar diferente de, ou em contraste a indicadores econômicos” (HAMBURGER, 1975). Essa oposição teve sua validade, num dado momento histórico, como forma de eliminar ênfases economicistas, que mascaravam a realidade social, mas deve ser evitada no desenvolvimento do conceito, na medida em que o fenômeno social é uno e indissolúvel. O fenômeno social nunca é apenas “econômico”, ou “social” ou “cultural”, etc. Isto não significa, porém, que não admita, até certo nível, uma abordagem limitada a alguns de seus aspectos (econômico, social, religioso, etc.), desde que se mantenha uma visão global e que se selecionem, para estudos, os aspectos essenciais do fenômeno considerado, em sua correlação com o todo.

Os intervalos de tempo de observação de qualquer indicador não podem ser fixados sem que, antes, se conheça o ritmo em que se processam as mudanças dos fenômenos sociais enfocados no contexto sociopolítico, econômico e ambiental. O ritmo das mudanças depende, por sua vez, tanto da natureza dos fenômenos como das políticas adotadas no sentido de influir sobre eles.

Explicitadas essas premissas básicas, coloca-se, ainda, a questão da operacionalidade dos indicadores ao nível da administração pública¹⁷, uma vez que este trabalho analisa, no Capítulo 4, os principais sistemas de indicadores e índices de desenvolvimento social produzidos por instituições públicas.

De acordo com Martin (1975), o Comissariado Geral do Planejamento Francês, na elaboração do VI Plano, explicita que cada função pública é coberta por uma hierarquia de três níveis de indicadores:

1. indicadores de meios: medem todos os tipos de meios que são mobilizados para a ação do sistema de intervenção, meios característicos para cada campo social considerado (por exemplo, o número de médicos formados é um indicador de meios do sistema saúde);
2. os indicadores de produto: são os que quantificam as saídas do sistema de intervenção (por exemplo, o número de médicos formados é um produto do sistema do ensino médico e o número de doentes tratados é um produto do sistema de saúde);
3. os indicadores de situação: são de grande interesse desta pesquisa. Estes indicadores se situam ao nível da realidade social, e podem ser ou não o resultado (direto ou indireto) de intervenções de um ou vários sistemas administrativos. Idealmente, seria necessário que esses indicadores pudessem exprimir a significação social dos resultados físicos obtidos. Porém, por falta de informações adequadas e de instrumentos de análise dos processos sociais, os indicadores de situação tendem a ser, habitualmente, expressos apenas em termos físicos (esperança de vida, por exemplo, ainda que seu caráter significativo não seja evidente).

Os indicadores de meios e de produto são úteis para a avaliação do esforço estatal no desempenho de suas funções. Nesta medida, como um importante instrumento da Administração por Objetivos, devem estar intimamente ligados à concepção, elaboração e execução do orçamento-programa.

¹⁷ Área de administração em que esta pesquisa se insere.

A partir desses indicadores, algumas instituições¹⁸ elaboraram medidas de eficácia da ação governamental. A eficácia interna da administração pode ser obtida através de uma relação entre indicadores de produto e de meios.

Os indicadores de eficácia externa são obtidos por uma combinação de indicadores de situação e de produto, e de indicadores de situação e de meios. As duas combinações são significativas, desde que se evitem correlações simplistas, pois dificilmente há uma correspondência direta linear entre serviços específicos prestados e mudanças de situações sociais. As mudanças nas quais há influência da ação governamental ocorrem sempre como resultado de um conjunto de programas, que se inter-relaciona com uma realidade dinâmica e complexa.

3.4.2 Tipos de Indicadores

Dados e indicadores são bases para a definição e ou adaptação de estratégias de intervenção social e política por parte de grupos e organizações da sociedade civil. Além disso, dados e indicadores são utilizados como instrumentos de disputa em torno da composição de imagens e de quadros analíticos e interpretativos da realidade que se opõem às interpretações produzidas no âmbito dos espaços de poder. (DIEESE e o DIESAT)

Os indicadores sociais diferenciam-se entre si, de acordo com vários sistemas classificatórios:

- a) Segundo a área temática da realidade social a que se referem. Indicadores de saúde; educacionais; mercado de trabalho; demográficos; habitacionais; segurança pública e justiça; infra-estrutura urbana; renda e desigualdade. Há classificações temáticas ainda mais agregadas, usadas na denominação dos Sistemas de Indicadores Sociais, como os Indicadores Socioeconômicos, de Condições de Vida, de Qualidade de Vida, Desenvolvimento Social ou Indicadores Ambientais (ONU, 1988).
- b) A complexidade metodológica na construção do indicador ou pelo menos, a quantidade de informação considerada para sua definição, também pode ser

¹⁸ Entre as principais instituições destacam-se SEI/BA; IPECE/CE; TCE/BA; TCU; IPARDES; Fundação João Pinheiro; PNUD; FEE/RS

usada como critério de diferenciação de dois conjuntos de indicadores: indicadores simples ou compostos.

c) A natureza do ente indicado, se recurso (indicador-insumo), realidade empírica (indicador-produto) ou processo (indicador-processo) (Carley, 1986).

d) Segundo os seis aspectos relevantes da avaliação dos programas sociais:

i) Indicadores de Desempenho

ii) Indicadores de Eficiência

iii) Indicadores de Eficácia

iv) Indicadores de Efetividade

v) Impacto

vi) Efeito. São agrupados em dois tipos: Efeitos Sociais e Efeitos Institucionais.

3.4.3 Índices e Sistemas de Indicadores Sociais

Os índices sociais são criados a partir da aglutinação de dois ou mais indicadores simples, relativos a distintas dimensões da realidade social. Conforme Jannuzzi (2001, p.22), indicadores de natureza composta têm o mérito de serem simples e de oferecerem alto poder de síntese.

Os Índices Sociais são, portanto, vistos como propícios a situações em que se deseja ter uma avaliação geral sobre um conceito que está sendo investigado em determinados grupos sociais. Jannuzzi (2001) entende que, quando se trata de uma questão complexa, em que se relacionam múltiplos fatores, existem vantagens se trabalhar com um índice, em vez de se dispor de sistemas constituídos por uma série de indicadores separados. A razão para isso é que o índice é capaz de dar uma idéia genérica sobre o fenômeno estudado, de forma mais rápida e objetiva.

Estudos sobre desenvolvimento têm sido realizados nos últimos tempos com base em índices sociais, tanto no âmbito internacional como no nacional. Internacionalmente, o exemplo mais famoso desse tipo de indicador é o Índice de Desenvolvimento Humano (IDH), que foi proposto pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) no começo da década de 1990. Criado por Mahbub ul Haq em colaboração com Amartya Sen, o IDH apresenta-se como medida geral e sumária do desenvolvimento humano, ainda que não ambicione abranger todos os aspectos desse desenvolvimento – uma vez que tal conceito é mais rico e complexo do que um índice pode esperar capturar.

Cabe mencionar o balanço realizado por Anand e Sen (2003a: 205) a respeito da discussão sobre o potencial analítico oferecido por índices sociais para o estudo do desenvolvimento. De modo sucinto, após analisarem os prós e os contras referentes a tal método, os autores afirmam que a capacidade de se refletir conceitos amplos e complexos – como o de desenvolvimento humano ou mesmo a perspectiva da privação – em um índice útil e não complicado, é uma meta que ainda não foi devidamente atingida.

Diferentemente dos índices sociais, que primam por condensar as múltiplas dimensões de um conceito social complexo em um único número, os sistemas de indicadores sociais objetivam representar as várias dimensões do conceito separadamente. Cada uma delas deve ser refletida através de indicadores unidimensionais. A idéia é que o sistema seja composto por róis de indicadores mais simples que, sendo analisados de forma conjunta, se proponham a oferecer uma visão abrangente do fenômeno estudado. Por outro lado, por exibir indicadores separados para cada dimensão formadora do conceito, essa abordagem propicia também que os aspectos componentes do fenômeno sejam estudados individualmente.

Da mesma forma como foi observado no caso dos índices sociais, há diversos estudos que se baseiam na elaboração de sistemas de indicadores sociais para a análise do desenvolvimento. O capítulo seguinte apresenta e analisa as principais formas de mensuração encontradas nos Índices e Sistemas de Indicadores de Desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos.

4 ANÁLISE DOS ÍNDICES E SISTEMAS DE INDICADORES DE DESENVOLVIMENTO

Se quisermos apreender a essência de um complexo de noções abstratas, devemos, por um lado, investigar as relações mútuas entre os conceitos e as afirmações feitas a seu respeito e, por outro, investigar como eles se relacionam com as experiências. Albert Einstein

Este capítulo tem o objetivo de apresentar e analisar as principais formas de mensuração encontradas nos Índices e Sistemas de Indicadores de Desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos. Para consecução do objetivo proposto para este capítulo, foram definidos os seguintes objetivos específicos: 1. Levantar, por meio de pesquisa bibliográfica e documental, os mais conhecidos Índices e Sistemas de Indicadores de Desenvolvimento disponíveis na literatura nacional e internacional; 2. Comparar os pressupostos conceituais dos mais importantes Índices ou Sistemas de Indicadores internacionais e nacionais; 3. Selecionar e analisar, entre os Índices e Sistemas nacionais, os mais conhecidos instrumentos de mensuração que contemplam a dimensão social do desenvolvimento; 4. Identificar e representar graficamente, entre os instrumentos selecionados na etapa anterior, o termo teórico, os respectivos construtos e demais conceitos empregados.

Apresentam-se inicialmente os procedimentos empregados para realizar o levantamento bibliográfico e documental dos mais conhecidos e representativos Índices ou Sistemas de Indicadores disponíveis na literatura nacional e internacional. Em seguida são analisadas, em duas etapas, as principais métricas do desenvolvimento social no Brasil e no mundo. Na primeira etapa, são comparados os pressupostos conceituais que permeiam a construção de 43 Sistemas e Índices mundiais e nacionais. Na segunda etapa, são analisadas as metodologias empregadas na elaboração de 46 Sistemas ou Índices nacionais que contemplam a dimensão social do desenvolvimento. Por fim, é apresentado o modelo conceitual do termo teórico Desenvolvimento Social, elaborado a partir da síntese das relações e construtos identificados na etapa anterior que, em seguida, é confrontado, no Capítulo 7, com a estrutura empírica dos indicadores sociais dos 5.561 municípios brasileiros.

O inventário dos Sistemas de Indicadores aqui utilizado foi construído a partir de pesquisa com o uso de “palavras-chave” nos principais “motores” de busca na Rede Mundial de Computadores (Internet) e levantamento bibliográfico em bibliotecas de universidades e de instituições de planejamento, pesquisa e estatística, entre os anos de 2004 e 2008. Todo o

material inventariado nesta etapa da pesquisa foi armazenado em bases de dados modeladas especificamente para a elaboração desta tese, o que possibilitou a identificação e posterior entrevista com os gestores públicos e pesquisadores de universidades¹⁹ e demais instituições²⁰ consultadas.

O levantamento, em duas etapas, das experiências mundiais e nacionais de elaboração dos principais Índices e Indicadores de Desenvolvimento existentes, cujos resultados também foram armazenados em base de dados específica, totalizou 64 Sistemas ou Índices nacionais e 397 Sistemas ou Índices internacionais. Tanto os Sistemas de Indicadores como os Índices estudados foram analisados sucintamente, no que tange a termo teórico, construto, conceitos abordados pelo Sistema ou Índice, forma de cálculo dos indicadores, origem, fontes de dados, procedimentos de conversão de escala, métodos e técnicas empregadas para agregação ou aglutinação dos indicadores em um único índice, entre outras informações²¹ de menor relevância para este estudo.

Visando melhorar a eficiência da recuperação da informação sobre os trabalhos estudados nas respectivas bases de dados, foi feita a “prospecção” das informações disponíveis, utilizando técnicas informétricas²² de indexação temática automática de palavras dos textos que compõem a amostra de 461 documentos, resultante do levantamento bibliográfico e documental realizado. Foi empregada a Lei²³ de Zipf-Booth, com o intuito de relacionar a frequência de ocorrência de palavras nos textos e determinar o conteúdo semântico dos mesmos, por meio de tais procedimentos (BRAGA, 1996).

A análise crítica dos mais conhecidos instrumentos nacionais e internacionais utilizados para a mensuração do fenômeno Desenvolvimento, pode, possivelmente, evidenciar

¹⁹ Universidade de Campinas (UNICAMP), Universidade Federal da Bahia (UFBA), Escola Nacional de Ciências Estatísticas (ENCE), Universidade Federal de Pernambuco (UFPE), Universidade Federal do Ceará (UFC).

²⁰ Fundação João Pinheiro (FJP-MG), Fundação CIDE (RJ), Fundação Sistema Estadual de Análise de Dados (SEADE-SP), Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE-CE), Fundação de Economia e Estatística (FEE-RS), Instituto Brasileiro de Geografia e Estatística (IBGE-RJ), Instituto de Pesquisa Econômica Aplicada (IPEA-DF), Superintendência de Estudos Econômicos e Sociais da Bahia (SEI-BA) e Programa das Nações Unidas para o Desenvolvimento (PNUD, DF).

²¹ Por exemplo: abrangência territorial, instituição responsável, unidade da federação, cobertura (número de municípios), ano base do cálculo, periodicidade, fórmulas de cálculo etc.

²² Refere-se ao estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social e não apenas aos cientistas. A informetria pode incorporar, utilizar e ampliar os muitos estudos de avaliação da informação que estão fora dos limites, tanto da bibliometria como da cienciometria (MACIAS-CHAPULA, 1998).

²³ Braga (1996) explica que a aplicação das Leis de Zipf, Zipf/Booth, Ponto T de Goffman e similares produz uma listagem de palavras e suas respectivas frequências no texto. Segundo a autora, essa transformação do texto em seqüências de caracteres indica que é possível determinar o conteúdo semântico dos textos, por meio de tais procedimentos. Para limpeza e re-formatação da base de dados foi utilizado o Infortrans, software alemão de reformatação desenvolvido por um grupo de pesquisadores da IuK Rieth GmbH.

o “estado da arte” na construção de Índices e Sistemas de Indicadores de Desenvolvimento no Brasil.

4.1 BREVE HISTÓRICO DOS ÍNDICES E SISTEMAS DE INDICADORES DE DESENVOLVIMENTO

Nesta seção, faz-se uma breve discussão acerca de algumas questões que envolvem os conceitos e as formas de mensuração do desenvolvimento encontradas no inventário, estudados sem a pretensão de esgotar o tema.

Atualmente, diversos importantes trabalhos que tratam da mensuração do desenvolvimento o fazem considerando o monitoramento de casos de ‘mudança social acelerada’ e, paralelamente, de um conceito quase que coloquial de bem-estar. Grande parte desses Índices ou Sistemas de Indicadores de Desenvolvimento tem, por finalidade, não somente mostrar, de forma sucinta, a relação entre os fatos e as mudanças, mas também subsidiar o monitoramento e a avaliação dos processos de transformações de natureza econômica, política, social e ambiental.

Como foi apresentado no Capítulo 2, o termo Desenvolvimento e suas formas de mensuração são amplamente difundidos na sociedade mundial contemporânea. Diante da complexidade que envolve a mensuração do desenvolvimento, é de fundamental importância investir no estudo e aprimoramento dos métodos empregados na elaboração dessas formas de mensuração. A análise dos Índices e Sistemas de Indicadores de Desenvolvimento disponíveis proporciona uma abordagem integrada e que permite superar a simples superposição de percepções disciplinares – a abordagem interdisciplinar.

A discussão sobre Desenvolvimento encontra-se vinculada aos Índices e Sistemas de Indicadores Sociais, desde a década de 1960, quando se deram os primeiros estudos sobre tais indicadores nos Estados Unidos. A grande dificuldade encontrada naquela época, no entanto, era a insuficiência de dados e a falta de metodologias mais robustas de mensuração dos fenômenos sociais. O insucesso inicial daqueles estudos foi atribuído à falta de “estatísticas, séries estatísticas e todas as outras formas de evidências” (BAUER, 1996, p.1 citado por NOLL e ZAPF, 1994, p.1).

Esses primeiros sistemas de indicadores abordavam exclusivamente aspectos econômicos do desenvolvimento, negligenciando outros aspectos das transformações consideradas determinantes dos mais variados aspectos relacionados ao bem-estar do homem (social e político, além do econômico) e do ambiente (ecossistema e espaço).

Já no Brasil, os indicadores sociais ganharam importância nos anos 70, quando o Conselho de Desenvolvimento Social – CDS propôs a “construção de um sistema de indicadores sociais de produção periódica da informação, tentando consolidar e articular diversas metodologias, entre as quais aquela recomendada pela ONU” (SANTAGADA, 1993, p.250).

Nos anos 80 os diversos esforços nacionais e internacionais de desenvolvimento dessas métricas ainda tinham em comum a falta de solução dos problemas de ordem metodológica e teórica, o que contribuiu para um período de estagnação. Nesse momento, muitos dos países mais industrializados deslocaram a atenção das políticas sociais em prol de uma crescente orientação econômica de caráter mais conservador.

A queda da chamada “Cortina de Ferro”, o surgimento da União Européia (entre outros movimentos políticos dos anos 90), aliados à necessidade de monitorar as transformações sociais daí resultantes, fez ressurgir o interesse e as possibilidades de aplicação de novas métricas para subsidiar a tomada de decisão por parte de governos e organismos públicos. Noll e Zapf (1994, p.13) vêem este período como sendo da “revitalização da contabilidade e do monitoramento social” necessários para se chegar a modelos explicativos das relações entre componentes e elementos dentro de um sistema maior.

Para Beck (1997), os anos 90 são o início da modernização reflexiva²⁴. Os conceitos de bem-estar e qualidade de vida são reconsiderados e relacionados a novos conceitos de Desenvolvimento. É nessa década que surgem, nos Estados Unidos, os indicadores de bem-estar sintéticos, tais como o Human Development Index (HDI) e o Index of Sustainable Development (ISD), voltados principalmente para o entendimento do

²⁴ Ulrich Beck explica a “modernização reflexiva” como um fenômeno que, em virtude do seu inerente dinamismo, faz a sociedade moderna acabar com suas formações de classe, camadas sociais, ocupações, papéis dos sexos, família nuclear, agricultura, setores empresariais e também com os pré-requisitos e as formas contínuas do progresso técnico-econômico. Esse novo estágio, em que o progresso pode se transformar em autodestruição, em que um tipo de modernização destrói outro e o modifica, é o que Beck denomina de “etapa da modernização reflexiva”. “Em outras palavras, a modernização reflexiva também – e essencialmente – significa uma ‘reforma da racionalidade’ que faz justiça à ambivalência histórica a priori em uma modernidade que está abolindo suas próprias categorias de ordenação” (Giddens, A.; Beck, U.; Lash, S., 1997, p. 12).

grande público em geral e, portanto, para a gestão e manipulação da chamada ‘opinião pública’.

4.2 AS MÉTRICAS DO DESENVOLVIMENTO E SUA DINÂMICA

O debate conceitual e metodológico é essencial à produção das métricas de qualquer fenômeno, pois a concepção de diferentes modelos implica diferentes indicadores (MAXWELL, 1999). Por outro lado, para a mensuração do desenvolvimento é importante o uso de conceitos que possam dar conta de sua condição subjetiva, complexa e multidimensional, não prescindindo, portanto, do conhecimento social, econômico, político e ambiental. Com base nesses princípios conduz-se, a seguir, a análise das propostas de estruturação dos conceitos e metodologias para construção de Índices e Sistemas de Indicadores Sociais que possam mensurar conceitos intangíveis e relevantes de um sistema social.

Embora a abordagem proposta nesta pesquisa esteja focada numa reflexão sobre os aspectos teórico-conceituais que permeiam a construção dos indicadores, é importante destacar também que a utilização dos indicadores deve ser baseada em critérios técnicos e pragmáticos, tais como os defendidos por Tironi et al. (1991), Soares Jr. E Quintella (2002), Jannuzzi (2001) e Trzesniak (1998). Entre os principais critérios de utilização de indicadores observados por esses autores destacam-se:

- Relevância ou seletividade: o indicador deve retratar um aspecto importante, essencial, crítico do processo/sistema;
- Gradação de intensidade: o indicador deve variar suficientemente no espaço dos processos/sistema de interesse;
- Univocidade: o indicador deve retratar, com total clareza, um aspecto único e bem definido do processo/sistema;
- Padronização ou estabilidade: a geração do indicador deve basear-se em uma norma, um procedimento único, bem definido e estável no tempo;

- Rastreabilidade: os dados em que a obtenção do indicador é baseada, os cálculos efetuados e os nomes dos responsáveis pela apuração devem ser registrados e preservados;
- Desagregabilidade: é importante que os indicadores possuam desagregabilidade populacional e territorial. Deve ser possível construir indicadores referidos à população-alvo, a espaços geográficos reduzidos, a grupos sociodemográficos ou a grupos vulneráveis específicos;
- Validade: é importante que se disponha de medidas representativas do conceito abstrato que o indicador pretende operacionalizar;
- Simplicidade: devem ser de fácil obtenção, construção, manutenção, comunicação e entendimento pelo público em geral, interno ou externo.

Como já mencionado, existem diversos problemas práticos e conceituais relacionados aos processos e aos instrumentos de mensuração do desenvolvimento. Inicialmente optou-se por apontar algumas limitações de natureza mais objetiva. As mais importantes são:

- Disponibilidade de dados;
- Consistência e compatibilidade dos dados;
- Emprego correto de técnicas analíticas;
- Demanda intrínseca de recursos financeiros;
- Demanda intrínseca de recursos humanos;
- Demanda intrínseca de recursos temporais.

Além dos problemas mencionados, deve-se destacar ainda a importância dos vieses cultural e político que condicionam a perspectiva do decisor e, eventualmente, de seu público durante os respectivos processos de tomada de decisão e interpretação das métricas.

Diante de tais desafios, procurou-se analisar qualitativamente cada proposta inventariada, a partir de uma abordagem teórico-conceitual e metodológica. Esta primeira etapa tem o objetivo²⁵ de comparar os pressupostos conceituais dos mais importantes Índices ou Sistemas de Indicadores internacionais e nacionais e centra-se na análise de 23 entre as 397 mais importantes instrumentos internacionais, além dos 20 principais Índices ou Sistemas de

²⁵ Segundo objetivo específico, apresentado na introdução deste capítulo.

Indicadores brasileiros entre os 64 mais conhecidos do grande público, a partir do estudo sistemático do alinhamento destes indicadores com o referencial teórico abordado.

Para seleção dos mais conhecidos e representativos índices e sistemas de indicadores disponíveis na literatura nacional e internacional, foram considerados os seguintes critérios:

- Aderência dos indicadores ao conceito de Desenvolvimento, como bem-estar do homem (econômico, social e político) e do ambiente (ecossistema e espaço);
- Aceitação e reconhecimento internacional;
- Número de países onde sabidamente o índice foi aplicado;
- Momento histórico em que o índice foi tornado público.

Os trabalhos selecionados por esse procedimento são apresentados nas subseções 4.2.1 e 4.2.2 que se seguem²⁶.

4.2.1 Sistemas e Índices Internacionais de Desenvolvimento Selecionados

Foram selecionados, para análise inicial, 23 indicadores internacionalmente reconhecidos²⁷, listados no Quadro 2, a seguir, onde se destacam as dimensões de análise neles contidas.

²⁶ A listagem completa dos sistemas de indicadores analisados encontra-se no APÊNDICE

²⁷ Além das experiências analisadas, devem também ser citados os seguintes indicadores: World Value Survey 1990 a 1993; Painel Socioeconômico (Alemanha) desde 1984; índice de pobreza do Banco Mundial 1990; linha de indigência da Cepal (CEPAL, 1991).

Quadro 2 – Distribuição dos índices e sistemas de desenvolvimento mundiais por dimensão de análise, 2008

Descrição	Dimensão				Cobertura (Países)
	Ambiental	Econômico	Político	Social	
1. Combined Consumption Level Index (Bennett, 1951)	X	X		X	31
2. Human Resources Development Index (HRDI) (Harbison and Myers, 1964)				X	75
3. Real Index of Consumption (Beckerman and Bacon, 1966)		X		X	80
4. UNRISD General Index of Development (GID) (McGranahan et al., 1972)		X		X	58
5. Physical Quality of Life Index (PQLI) (Morris, 1979)				X	150
6. Composite Basic Needs Indices (Ram, 1982)		X		X	82
7. World Handbook of Political and Social Indicators (UN, 1982)		X	X	X	156
8. Index of Social Progress (ISP) (Estes, 1984)		X	X	X	107
9. World Standard Distance Scales (Ginsberg et al., 1986)		X		X	143
10. Human Suffering Index (HSI) (Camp and Speidel, as related in Hess, 1989)		X	X	X	mais de 60
11. Quality of Life Rankings (Slottje, 1991)		X	X	X	126
12. Combined Quality of Life Indices (CQLI) (Diener, 1995)	X	X	X	X	77
13. Index of Economic Freedom (Johnston and Sheehy, 1995)		X			mais de 100
14. Economic Freedom Indices (EFIs) (Gwartney et al., 1996)		X	X		103
15. Human Development Index (HDI) (UNDP, 1996)		X		X	174
16. Gender-related Development Index (GDI) (UNDP, 1996)		X		X	163
17. Gender Empowerment Measure (GEM) (UNDP, 1996)		X	X	X	100
18. Global Competitiveness Indices (World Economic Forum, 1996)		X	X	X	49
19. Human Poverty Index (HPI-1) for developing countries (UNDP, 1999)				X	92
20. Human Poverty Index (HPI-2) for developed countries (UNDP, 1999)		X		X	17
21. Ecological Footprint (Wackernagel et al. 2000)	X	X		X	52
22. Dashboard of sustainability (IISD, 2003)	X	X		X	mais de 200
23. Environmental Sustainability Index (YALE; COLUMBIA, 2005)	X	X	X	X	146

Fonte: Elaboração própria

4.2.2 Sistemas e Índices Nacionais de Desenvolvimento Selecionados

Com base nos mesmos critérios foram selecionados 20 indicadores nacionais, mostrados no Quadro 3 e analisados em função das dimensões neles contidas²⁸.

Quadro 3 – Distribuição dos índices e sistemas de desenvolvimento brasileiros por dimensão de análise

Descrição	Dimensão			
	Ambiental	Econômico	Político	Social
1. Linha da indigência da Cepal (CEPAL, 1991)		X		
2. Linha da Indigência do IBGE, IPEA, Cepal (Árias, 1991)		X		
3. Barômetro de Sustentabilidade (Prescott-Allen, 1995)	X	X		X
4. Indicador de Progresso Genuíno (GPI) (Cobb et al., 1995)	X	X		X
5. Índice de Qualidade de Vida Urbana (IQVU) (PMBH e PUC/MG, 1995)	X			X
6. Glossário da Pobreza e Desenvolvimento Humano (PNUD, 1997)		X		X
7. Indicadores de comunidade sustentável (Seattle, 1998)	X	X		X
8. Linha da Indigência do IPEA (Rocha, 2000)		X		
9. A Robust Poverty Profile for Brazil Using Multiple Data Sources (Ferreira et al., 2001)		X		X
10. Mapa do fim da fome (FGV, 2001)		X		
11. Índice de Desenvolvimento Econômico (IDE) (SEI, 2002)		X		X
12. Índice de Desenvolvimento Social (IDS) (SEI, 2002)				X
13. Indicadores do estado do Pará (Ribeiro, 2002)	X	X		X
14. Índice de Sustentabilidade Urbana (ISU) (Braga et al., 2002)	X	X	X	X
15. Atlas da Exclusão Social (Pochmann et al., 2004)		X		X
16. Geo Cidades (SMA/SP/MG, 2004)	X			
17. Índice de Desenvolvimento Socioeconômico (IDESE) (FEE, 2004)	X	X		X
18. Índice de Desenvolvimento Sustentável (IDS) (IBGE, 2004)	X	X		X
19. Mapeamento da Pobreza (Soares Jr; Quintella, 2005)		X		X
20. Índice DNA-Brasil (UNICAMP, 2005)	X	X	X	X

Fonte: Elaboração própria

²⁸ A listagem completa dos sistemas de indicadores nacionais analisados encontra-se no APÊNDICE A.

A análise das dimensões dos Índices e Sistemas de Indicadores de Desenvolvimento, selecionados nesta primeira etapa de caracterização e comparação, mostra limitações quanto a aspectos considerados relevantes para esta pesquisa. Merece destacar que, dos 43 indicadores mundiais e nacionais selecionados e listados nos quadros 2 e 3, apenas quatro indicadores contemplam todas as dimensões abordadas no referencial teórico adotado nesta pesquisa (ambiental; econômica; política e social). Ainda mais surpreendente é o fato de que para a maioria deles não se consegue referências explícitas quanto a suas abordagens teóricas e modelos empíricos. Ainda assim, a abordagem multidimensional e interdisciplinar, fundamentada nos trabalhos desses autores estudados, possibilitou a identificação de dimensões de análise comuns a todas as escolas. Essas dimensões comuns são: a Dimensão Social, a Dimensão Econômica, a Dimensão Política e a Dimensão Ambiental.

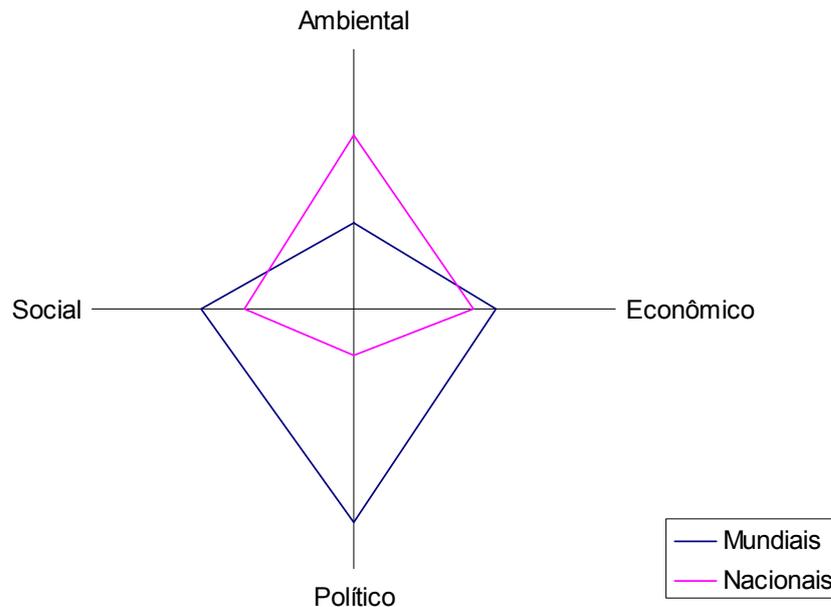
A análise dessas quatro dimensões permitiu caracterizar e comparar, com base nos critérios estabelecidos e citados anteriormente, os Índices e Sistemas de Indicadores Sociais mais conhecidos, assim como possibilitou a avaliação da aderência dos trabalhos estudados à representação do processo de mudança social acelerada em busca de um estado de bem-estar, indicando óbvias limitações para trinta e nove deles.

A Figura 4, que se segue, mostra um gráfico que permite identificar, de forma aproximada, a magnitude das diferenças de abordagem teórico-conceitual entre os trabalhos nacionais e mundiais analisados. A verificação das dimensões de análise revela diferenças significativas entre os índices e sistemas indicadores de desenvolvimento nacionais e mundiais. Constata-se que a maioria dos trabalhos analisados contempla apenas parcialmente as quatro dimensões (Econômica, Social, Política e Ambiental) preconizadas no referencial teórico sobre Desenvolvimento apresentado no Capítulo 2 desta tese. A maioria deles responde apenas às dimensões social e econômica. Entretanto, é possível identificar, entre os trabalhos nacionais, uma maior frequência de indicadores da dimensão ambiental do desenvolvimento, enquanto os sistemas mundiais abordam, com maior intensidade, a dimensão política, explicitando, nos dois casos, apenas algumas facetas do fenômeno que se propõem a mensurar.

Inspirado em reflexões sobre a Teoria do Desenvolvimento, portanto, esta tese busca oferecer uma abordagem diferenciada por meio do diálogo com escolas de pensamento distintas. Este exercício permite identificar as principais diferenças e similitudes nas abordagens sobre Desenvolvimento e reuni-las por corrente de pensamento. Constata-se,

também, que as escolas²⁹ apresentadas no Capítulo 2 desta tese, convergem ao aceitar o substantivo ‘bem-estar’ como o estado desejável de equilíbrio, tanto para o ambiente como para o homem.

Figura 4 – Dimensões dos Índices de Desenvolvimento Nacionais e Mundiais



Fonte: Elaboração própria

A avaliação de aderência dos 43 Índices ou Sistemas de Indicadores de Desenvolvimento às Dimensões (Econômico, Social, Político e Ambiental) do Desenvolvimento nesta pesquisa permitiu observar que também existem vários problemas práticos e conceituais relacionados aos processos e aos instrumentos de medida mais usualmente empregados no estudo do Desenvolvimento, conforme é discutido a seguir.

A próxima seção apresenta a análise dos Índices de Desenvolvimento Nacionais que contemplam a dimensão social do Desenvolvimento. Com o objetivo de abranger um maior número de propostas nacionais existentes, optou-se, na segunda etapa de seleção e análise dos Índices e Sistemas de Indicadores Sociais, por trabalhar com as informações de 46 Índices e Sistemas de Indicadores Nacionais, entre as 64 instrumentos armazenados na base de dados mencionada na abertura deste capítulo. Esta discussão está estruturada de maneira a

²⁹ Desenvolvimento como Crescimento Econômico e a escola denominada nesta tese de Via “alternativa”.

permitir caracterizar e comparar tais experiências, considerando a abrangência temática, que é muito diversificada; o processo de construção dos Índices; as fontes de dados utilizadas e o procedimento de cálculo; as escalas de conversão e os métodos de agregação dos indicadores.

4.3 PRINCIPAIS SISTEMAS DE INDICADORES E ÍNDICES DE DESENVOLVIMENTO SOCIAL NO BRASIL

A partir da década de 1990, quando o PNUD apresentou o Índice de Desenvolvimento Humano (IDH), a elaboração de índices e Sistemas de Indicadores apresenta tendência crescente no Brasil.

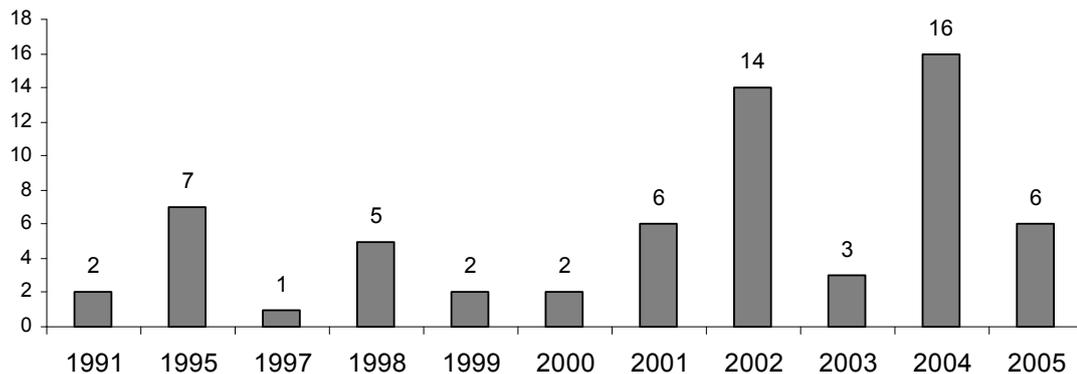
No Brasil, a abertura democrática impôs mudanças no sentido da descentralização de recursos governamentais. A Constituição de 1988 concedeu maior equilíbrio na distribuição destes entre as esferas de poder, com ganhos para estados e municípios. Nessa época, alguns estados e municípios começam a desenvolver políticas antes privativas do governo federal, “ocorrendo a descentralização, principalmente no âmbito do federalismo fiscal, mediante redistribuição de recursos às varias esferas de poder referente a algumas políticas públicas, caso das políticas de assistência social, educação e saúde, sem que exista, todavia, um verdadeiro programa nacional de descentralização” (ARRETCHE, 1996).

Segundo Mota (1994), a disponibilidade de indicadores e índices, como os que são estudados neste trabalho, passou a ser peça fundamental para a formulação de políticas estaduais compensatórias requeridas pelo processo de descentralização descrito. Deve-se, porém, lembrar que não foram eliminadas, na gestão pública, as práticas paternalistas, clientelistas e assistenciais, as quais visam a perpetuação no poder de grupos políticos que se utilizam da máquina pública em benefício próprio.

A partir de 1998, o PNUD disponibiliza o IDH-M, elaborado com o objetivo de dimensionar o Desenvolvimento Humano em todos os municípios brasileiros. Este índice foi o mais amplamente reconhecido, para avaliar, comparativamente, o conjunto de municípios no Brasil (FJP; IPEA; PNUD, 2000). A produção desses sistemas de indicadores no país se intensificou a partir da década de 2000, possivelmente motivado pela divulgação dos resultados do Censo Demográfico do IBGE, conforme apresenta o Gráfico 1, a seguir.

O crescimento das proposições de Sistemas de Indicadores de Desenvolvimento aí inauguradas levou a uma profusão de possibilidades, que justificam e tornam necessários trabalhos de pesquisa, como o que compõe esta tese.

Gráfico 1 – Índices e Sistemas de Indicadores Nacionais estudados, Brasil, 1991-2005



Fonte: Elaboração própria

Nesta segunda etapa de seleção³⁰ e análise foram estudados 46 desses sistemas - sete reúnem informações da totalidade dos municípios brasileiros e 39 reúnem sistemas que abrangem os municípios de suas respectivas Unidades da Federação, em parte ou no todo. Estas 46 formas de mensuração do desenvolvimento brasileiro enfocam 16 aspectos teóricos (termo teórico), estruturados a partir de 25 temas (construtos) e 816 indicadores (conceitos operacionais ou variáveis observáveis).

Entre os instrumentos de mensuração analisados nesta segunda etapa, o sistema de abrangência nacional que é composto do menor número de indicadores é o Índice de Desenvolvimento Humano³¹ (IDH) que agrega apenas três construtos (Longevidade, Renda e Educação) e quatro indicadores. Enquanto o Índice Potencial de Desenvolvimento dos Municípios³² (IPDM) é composto por sete construtos (Trabalho e rendimento, Saúde, Educação, Renda, Dinâmica Populacional, Organização Político-Institucional e Infra-estrutura) e por 21 indicadores.

³⁰ Esta etapa está relacionada com o terceiro objetivo específico deste Capítulo.

³¹ Criado pelo PNUD - Programa das Nações Unidas para o Desenvolvimento.

³² Criado pelo Consórcio Monitor/Bocinhas e Campos, no projeto Atualização dos Eixos Nacionais de Integração e Desenvolvimento do Ministério do Planejamento.

Sob o aspecto da distribuição espacial no Brasil, os 39 Sistemas estudados e os respectivos 816 indicadores que os constituem, estão distribuídos conforme a Tabela 1, a seguir.

Tabela.1 – Índices e Sistemas de Indicadores Nacionais de Desenvolvimento por região geográfica de elaboração – Brasil, 2010

Abrangência	Número de Sistemas	Número de Indicadores
Centro-oeste	2	44
Nordeste	6	101
Norte	6	80
Sudeste	18	379
Sul	7	128
Todos os Municípios	7	84
Total	46	816

Fonte: Elaboração Própria

Como pode ser observado, segundo a distribuição especial apresentada, as regiões Sul e Sudeste foram responsáveis pela elaboração da maioria (aproximadamente 61%) dos Sistemas de Indicadores estudados.

Segundo as categorias³³ de termo teórico identificadas, a maior parte dos trabalhos possui enfoque no Desenvolvimento Social (52,2%), sendo que entre os Sistemas com abrangência nacional (Brasil), o Desenvolvimento Social é focado em 71,4% dos trabalhos estudados. Embora exista uma grande diversidade de termos teóricos, observa-se a presença dos enfoques no Desenvolvimento Político, Desenvolvimento Local, Qualidade de Vida e Qualidade Socioambiental. Este último termo teórico se destaca com 10,9% dos sistemas estudados.

Assim como o termo teórico mais freqüente entre os trabalhos estudados é o Desenvolvimento Social, nota-se que os indicadores da temática social também têm uma maior participação entre os 25 construtos que compõem os 816 indicadores dos 46 sistemas de indicadores ou índices estudados.

Cinco construtos estão presentes de forma expressiva em quase todos os sistemas estudados: Educação (17,17%), Saúde (34,9%), Trabalho e Rendimento (36,5%), Segurança

³³ Foram identificadas 15 categorias de termo teórico: Desenvolvimento Social; Desenvolvimento Local/Municipal; Qualidade Socioambiental; Condições de Habitação; Distribuição de Recursos (Gestão Territorial); Qualidade Ambiental "stricto sensu"; Vulnerabilidade Social; Qualidade de Vida Urbana; Condições de Saneamento Básico; Gestão Pública; Pobreza; Sustentabilidade Fiscal; Qualidade de Vida; Condições de Educação; Desenvolvimento Político.

(5,2%) e Dinâmica Populacional (5,2%). Os citados construtos representam a abordagem tradicional do Desenvolvimento Social. Entretanto, outros construtos, diferentes dos tradicionais, também foram identificados em apenas dois trabalhos; entre os mais frequentes do grupo menos representado destacam-se o Meio Ambiente, Justiça, Cultura e Lazer.

Entre as fontes de dados utilizadas por esses sistemas, como seria de se esperar, o Instituto Brasileiro de Geografia e Estatística (IBGE) se destaca. O IBGE é citado como fonte de dados em 71,7% dos trabalhos estaduais e 85,7% dos trabalhos de abrangência nacional.

É possível também constatar a falta de reflexões teóricas explícitas em todos os documentos de divulgação encontrados, o que dificulta a compreensão dos construtos e conceitos que os Índices se propõem a medir. Também foi constatado que 30,4% dos sistemas não citam as fontes de dados utilizadas nos cálculos de seus indicadores ou índices. Com esse mesmo percentual (30,4%) estão presentes como fontes de dados vinculadas aos governos das Unidades da Federação.

A constatação da presença significativa de fontes de dados alternativas ao IBGE, nos sistemas de indicadores, evidencia o problema brasileiro da escassez de estatísticas públicas³⁴ oficiais amplas e confiáveis, que permitam fazer diagnósticos mais detalhados e conclusivos sobre a realidade social do país. Os problemas com a quantidade e a qualidade das informações sobre o Desenvolvimento Social vão desde falhas nos registros ou mesmo ausência de levantamento das informações, até limitações técnicas e tecnológicas de armazenagem, integração, tratamento e análise das informações disponíveis.

Assim como não estão disponíveis, na maioria dos documentos, as reflexões teóricas e as fontes de dados, também não foram observadas referências ao tipo de conversões de escala utilizado em 17,4% dos sistemas estudados. Entre os quatro procedimentos³⁵ de conversão de escala para padronização das variáveis observáveis, o mais frequentemente adotado foi o de escala de intervalo linear, presente em 52,2% dos sistemas. Os demais trabalhos (30,4%) empregam escalas do tipo escore padronizado (15,2%), escala de razão (6,5%) e escala de intervalo logarítmica (2,2%). Esse último tipo de escala de conversão é aplicado ao indicador de renda do IDH-M.

³⁴ "Estatística pública" ou "estatística oficial", refere-se à informação estatística produzida pelas agências estatísticas do governo – órgãos de recenseamento, departamentos de estatística e instituições semelhantes.

³⁵ O procedimento de cálculo das escalas de intervalo linear, escore padronizado, razão e intervalo logarítmica está disponível no Apêndice.

Por fim, os tipos de métodos e técnicas empregadas para agregação ou aglutinação dos indicadores em um único índice identificados nos 46 sistemas de indicadores estudados foram, em primeiro lugar, as médias aritméticas simples ou ponderada (71,8%); em segundo lugar está a análise multivariada, presente em 13% dos sistemas, seguida da média geométrica simples com 8,7% , conforme revela a Tabela 2 que se segue.

Tabela 2 - Métodos de Agregação utilizados nos Sistemas de Indicadores Nacionais - Brasil, 2009

Método de Agregação	Sistemas de Indicadores (%)
Média aritmética ponderada	41,30
Média aritmética simples	19,57
Técnicas de análise multivariada	13,04
Média aritmética simples e ponderada	10,87
Média geométrica simples	8,70
Outras formas	4,35
Sem informação	2,17
Total	100,00

Fonte: Elaboração própria

Diferentemente da situação de escassez de informações relacionadas com as reflexões teóricas e as fontes, entre os documentos estudados, no caso das informações sobre a agregação dos dados, apenas um Sistema não apresentou informações sobre a técnica de agregação empregada na obtenção dos Índices de Desenvolvimento divulgados³⁶. Dessa forma, foi possível constatar que o uso das médias aritmética ou geométrica, em 80,44% dos casos analisados, limita o poder de síntese da informação quando a medida de posição calculada a partir dos indicadores reduz a transparência necessária para representar as nuances do fenômeno multidimensional e complexo através de um número ou índice. Embora essas médias sejam capazes de dar uma idéia genérica sobre o fenômeno estudado, de forma mais rápida e objetiva, seu uso, nos casos em que os indicadores sociais apresentam assimetria e alta variabilidade interna, resulta em resultados inconsistentes entre si.

Empregadas em apenas 13,04% dos casos, as técnicas de análise multivariada, por sua vez, restringem-se à Análise de Regressão (AR). Este método, além de apresentar

³⁶ Este fato pode indicar uma possível preocupação dos responsáveis pela elaboração dos indicadores, em legitimar o trabalho através do uso da linguagem matemática que, infelizmente, não é suficiente para apresentar a abordagem teórico-conceitual do Índice ou Sistema de Indicador e parece ser pouco compreendida pela maioria dos usuários desses Sistemas.

dificuldades decorrentes da complexidade inerente aos testes³⁷ complementares para sua validação, não permite contemplar como ocorre a articulação entre os diversos indicadores agregados em torno de um índice.

Entre os métodos de aglutinação analisados não houve convergência em torno dos procedimentos de ponderação empregados na agregação dos indicadores. Segundo Mestrum (2002, p.37), a ponderação é a etapa arbitrária no processo de criação de índices sintéticos.

A diversidade de propostas é, por si só, razão para o consenso percebido, tanto na literatura, quanto na pesquisa de campo, de que é difícil mensurar o desenvolvimento; porém, como visto, existe, além da grande variedade, também uma grande discordância em relação à definição do conceito. Mais grave ainda é a significativa ausência de informações sobre as escolhas teórico-metodológicas intrínsecas a cada uma das métricas.

O conhecimento que pode potencialmente ser extraído desses Índices e Sistemas de Indicadores de Desenvolvimento tem o potencial de contribuir objetivamente no processo de organizar a agenda de ação do poder político; não obstante, também deve ser motivo de preocupação, uma vez que esconde as idiossincrasias próprias de cada grupo social responsável e interessado pela elaboração e disseminação dessas métricas.

A disseminação rápida das propostas de mensuração do desenvolvimento parece ser potencializada pela percepção generalizada de que números são, por suposição, imparciais, atribuindo-lhes aura de infalibilidade e universalidade. Contribui para essa difusão também a adoção de metodologias simplificadas, às vezes incompletas e com raríssimas referências sobre as abordagens teóricas nelas adotadas.

Todos estes fatores corroboram uma suposta e abstrata neutralidade ideológica a partir do estabelecimento de *rankings* de Desenvolvimento. Entretanto, Morgan (1968) e Babbie (1995) alertam que diferenças entre os indicadores devem refletir diferenças entre os modelos de mensuração. A análise desses modelos e seus respectivos métodos e técnicas empregados para agregação ou aglutinação revela que a maioria das equipes responsáveis por sua formulação recorre a métodos de seleção *ad hoc* baseados em experiência pessoal, em detrimento de técnicas reconhecidas cientificamente, a exemplo da popular Análise Fatorial Exploratória (AFé). Como é apresentada no Capítulo 6, a AFé é amplamente utilizada para identificar as variáveis latentes ou conceitos de um construto a ser investigado.

³⁷ Teste de Multicolinearidade; Teste de Normalidade de resíduos; Teste de Homocedasticidade; Teste de Correlação entre Resíduos.

Pode-se inferir que a maioria dos Índices publicados é resultado da aglutinação de indicadores ditados pela experiência cotidiana dos técnicos responsáveis por sua elaboração. A maioria dos Índices e Sistemas de Indicadores analisados não reflete a preocupação de ir além do senso comum e é freqüentemente expressa em termos do raciocínio prático para ação em detrimento de um raciocínio guiado por um esquema teórico, que serve de eixo para seleção dos construtos e respectivos conceitos e indicadores, além de assegurar a pertinência e a interpretação das informações reunidas.

Em síntese, as metodologias empregadas na elaboração dos Índices e Sistemas de Indicadores de Desenvolvimento Social têm capacidades distintas de apreender a complexidade do fenômeno. Cada uma delas reflete, de forma parcial, o caráter multidimensional e complexo do fenômeno, sendo que nenhuma delas apresenta-se plenamente satisfatória para dar conta dessa tarefa.

Os documentos dos 46 Índices e Sistemas de Indicadores nacionais revelam uma maior preocupação em descrever e ranquear os indicadores, sem explicitar ao menos uma tentativa de identificar as causas e conseqüências do Desenvolvimento Social no Brasil e demais localidades estudadas e, muito menos, apresentar a base teórica e conceitual de suas construções, o que é buscado na seção que se segue.

4.4 MODELO CONCEITUAL DO DESENVOLVIMENTO SOCIAL

As seções anteriores apresentaram o resultado do levantamento dos principais sistemas de indicadores e índices de desenvolvimento nacionais. A análise dos diversos aspectos teóricos e metodológicos destes sistemas evidenciou a existência de grande variabilidade de metodologias, demonstrando possivelmente uma necessidade de se desenvolver sistemas adequados à diversidade cultural brasileira. Entretanto, do ponto de vista teórico/conceitual, os documentos analisados não oferecem a possibilidade de identificar, com clareza, as relações existentes entre os indicadores associados ao conceito Desenvolvimento Social empregado. Apesar do enorme avanço, no Brasil, na elaboração desses sistemas, ainda existem muitas oportunidades de produzir elementos que subsidiem uma discussão mais embasada sobre o Desenvolvimento Social.

Conforme explicitado no Capítulo 3, a representação das estruturas que podem ser identificadas a partir das relações de influência que unem as variáveis sociais umas às outras, é uma das primeiras etapas do trabalho de mensuração de fenômenos sociais. Tal representação consiste em um diagrama que fornece uma visão pictórica das relações que conectam os diversos conceitos e construtos que compõem o fenômeno. A intenção é que uma representação dessa natureza ofereça uma visão conjunta do Desenvolvimento Social, de modo a ajudar a compreensão do seu caráter multidimensional.

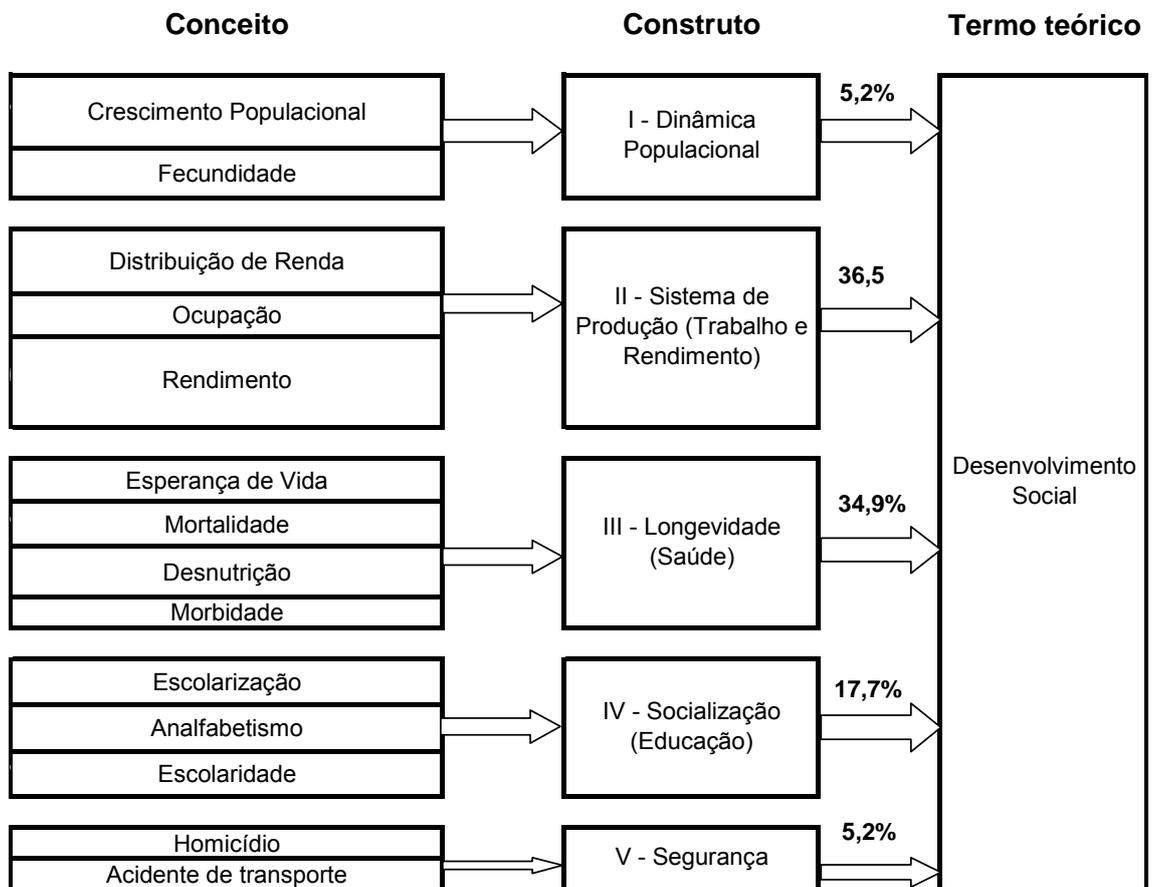
No caso dos Índices e Sistemas de Indicadores, essa estrutura latente pode ser revelada com a representação dos construtos e das relações por meio de diagrama de caminho, ou por um modelo matemático. Essa representação diagramática deve revelar tanto as relações existentes entre conceitos e os indicadores de Desenvolvimento Social, como as associações existentes entre os próprios construtos. Esse atributo faz com que o diagrama apresente de maneira inteligível a forma como se estruturam as relações representadas nos Índices ou Sistemas de Indicadores. Por oferecer tal explicitação, a representação em diagramas de caminho mostra-se adequada à investigação de fenômenos complexos e multidimensionais, já que comunica o caráter relacional do fenômeno social estudado.

Nesse contexto, o objetivo³⁸ desta seção é identificar o termo teórico (conceito ordenador de base) Desenvolvimento Social e os respectivos construtos e demais conceitos empregados nos documentos disponíveis dos principais Índices e Sistemas de Indicadores Sociais nacionais e representa-los no aqui denominado Modelo Conceitual.

Para suprir a estrutura hierárquica, escolhida para representar o diagrama de caminho que apresenta o modelo conceitual, necessário à representação aproximada da estrutura teórica do Desenvolvimento Social encontrado nos Índices e Sistemas de Indicadores brasileiros, foram selecionados os cinco construtos mais frequentes, entre o conjunto de indicadores estudados neste capítulo e os 15 principais conceitos associados a estes construtos e ao fenômeno social, conforme apresentado na Figura 5, a seguir:

³⁸ Terceiro objetivo específico, apresentado na introdução deste capítulo.

Figura 5 – Modelo Conceitual Hierárquico do Desenvolvimento Social disponível em 46 Índices e Sistemas de Indicadores nacionais



Abordar e analisar cada construto separadamente e por técnicas específicas permitiu representar graficamente (diagrama de caminho) o termo teórico Desenvolvimento Social empregado nos sistemas de indicadores analisados, conforme preconiza Ensslin et al. (2001). As relações entre os conceitos selecionados e os respectivos construtos estão representadas no modelo gráfico do tipo diagrama de caminho (*path-diagram*). Esse diagrama permite representar também a síntese dos modelos matemáticos empregados nos Sistemas de Indicadores e Índices de Desenvolvimento Social estudados. Optou-se pela representação em estrutura hierárquica, pois entre os trabalhos estudados não foi identificado o uso de Modelagem de Equações Estruturais³⁹ (MME). Em princípio, os modelos matemáticos identificados nos 46 sistemas de indicadores estudados são médias aritméticas ou geométricas e regressões múltiplas. Este tipo de modelagem matemática não exige o formato de

³⁹ A Modelagem de Equações Estruturais (MME) exige o emprego de modelos circulares.

representação circular, o qual, como foi dito no Capítulo 3, exige o emprego de equações estruturais no cálculo dos Índices.

De fato, um dos motivos que faz a representação por diagrama de caminho ser particularmente adequada aos fenômenos sociais é que muitas teorias e modelos elaborados nessa área do conhecimento baseiam-se em conceitos teóricos abstratos. Tais conceitos não podem ser diretamente observados, não sendo passíveis de mensuração. Suas inserções em modelagens estatísticas se dão através dos chamados “construtos” ou “variáveis latentes”, elaborados a partir da utilização de variáveis observáveis, que funcionam como indicadores daqueles conceitos trabalhados na seara teórica (BYRNE, 1994).

Como pode ser observado na Figura 4, entre os cinco construtos identificados e representados, foram selecionados e definidos, nas próximas subseções, os 15 conceitos utilizados para compor o referido diagrama de caminho.

4.4.1 Dinâmica Populacional: Crescimento Populacional e Fecundidade

O crescimento populacional é um conceito básico na análise da dinâmica populacional, e refere-se às mudanças no tempo que ocorrem ao tamanho do número de pessoas que residem no território.

No último século a população brasileira multiplicou por dez: em 1900 residiam, no Brasil, cerca de 17 milhões de pessoas; no ano 2000 quase 170 milhões. Desde o primeiro recenseamento (1872) ocorreram várias mudanças no padrão da evolução demográfica brasileira.

A fecundidade está associada à procriação humana, em termos do número efetivo de filhos em relação às mulheres em idade reprodutiva. Do ponto de vista da Dinâmica Populacional (I), a análise da fecundidade tenta medir em que grau e como vão ocorrendo os nascimentos. A importância deste conceito para o construto está no fato de que os nascimentos vão determinando, conjuntamente com a mortalidade e as migrações, o crescimento e a estrutura da população. Também, o número de filhos que as mulheres têm está estreitamente relacionado com aspectos, tais como a saúde materno-infantil, aspectos sociais ligados à constituição das famílias e a entrada da mulher no mercado de trabalho, que

tem contribuído com a redução significativa do número de nascimentos (natalidade) e por consequência da fecundidade.

O estudo da fecundidade permite avaliar tendências de uma das componentes mais importantes da dinâmica demográfica e fornece subsídios para a elaboração de projeções e estimativas populacionais, como também auxilia no planejamento de políticas públicas nas áreas de saúde, educação, trabalho e previdência social.

Segundo o IBGE (2008), a razão fundamental da queda das taxas de crescimento populacional no Brasil foi a diminuição da fecundidade (média de número de filhos por mulher em idade de procriar, entre 15 a 49 anos), que caiu de 6,3 filhos, em 1960, para 2,0 filhos, em 2006, o que significa que as famílias brasileiras estão diminuindo.

4.4.2 Sistema de Produção (Trabalho e Rendimento): Distribuição do Rendimento; Ocupação e Rendimento.

O primeiro conceito selecionado para compor o Construto II é distribuição do rendimento. Para Hoffman (2001), crescimento econômico e distribuição do rendimento são temas básicos da teoria econômica e das análises econômicas do desenvolvimento do Brasil, um país com desigualdade extremamente elevada. A análise da desigualdade ganhou especial notoriedade a partir de 1972, depois da publicação dos resultados do Censo Demográfico de 1970. Para o autor, parece consensual que é necessário reduzir essa desigualdade, mas toda medida prática nesse sentido gera polêmica. A distribuição da renda de forma equitativa é um dos principais desafios do desenvolvimento social.

Além da distribuição da renda, a ocupação também é um dos principais problemas que afetam tanto os países desenvolvidos quanto os países em desenvolvimento, além de ser um fator preponderante na determinação do nível de desenvolvimento social de uma nação.

A análise acerca do conceito de rendimento de uma sociedade deve contemplar a família, pois além de ser considerada unidade de produção e consumo, é, também, unidade de reprodução, agregação e socialização.

4.4.3 Longevidade (Saúde): Esperança De Vida Ao Nascer; Mortalidade; Desnutrição e Morbidade

O estudo da Longevidade (III) compreende quatro conceitos relacionados com as condições de saúde de uma população. A esperança de vida ao nascer está estreitamente relacionada às condições de vida e de saúde da população. A mortalidade é definida como a ação da morte sobre uma população e é um dos componentes centrais da dinâmica demográfica.

Por estar estreitamente relacionado ao rendimento familiar, ao nível da fecundidade, à educação das mães, à nutrição e ao acesso aos serviços de saneamento básico, a redução da mortalidade infantil é um dos importantes e universais objetivos do Desenvolvimento Social.

Szwarcwald e Castilho (1995), ao propor um procedimento para a estimativa da mortalidade infantil, no Brasil, na década de 80, baseado apenas na distribuição etária dos óbitos registrados, demonstram que a série temporal que descreve a evolução da mortalidade infantil no Brasil é marcada por nítidas interrupções na tendência de declínio, em que os aumentos foram atribuídos, sobretudo, às doenças relacionadas à desnutrição; o estudo revelou a persistência e a acentuação das desigualdades no processo de morte entre os menores de um ano.

A desnutrição geralmente está associada ao subdesenvolvimento. O direito à alimentação e à nutrição adequadas é um direito humano básico e sua promoção deve estar entre as prioridades de um país que busca o Desenvolvimento Social.

O quarto e último Conceito do Construto Longevidade (III) está relacionado com a morbidade da população provocada pela precariedade nos sistemas de abastecimento de água, esgotamento sanitário, coleta e destino final dos resíduos sólidos, drenagem urbana, bem como a higiene inadequada, que se constituem em risco para a saúde da população, sobretudo para as pessoas mais pobres dos países em desenvolvimento, que ainda ficam com sua dignidade afetada.

Szwarcwald et al. (1999) afirmam que vários trabalhos mostraram, de forma consistente, a existência de melhores condições de saúde em sociedades com distribuição de renda mais equilibrada. E também que algumas sociedades desfrutam de condições de saúde bem melhores do que as esperadas para os seus níveis de renda e vice-versa, sugerindo que

outros fatores relacionados à complexidade da estrutura social podem influenciar mais profundamente a situação de saúde do que os índices de desenvolvimento.

4.4.4 Socialização (Educação): Escolarização; Analfabetismo e Escolaridade

Assim como a atenção à saúde, a educação (Socialização – Construto IV) é uma das prioridades para um país. Ela é fundamental para promover o desenvolvimento social, capacitando os cidadãos para lidar com as questões que o envolvem, facilitando, assim, a aquisição de valores, habilidades e conhecimentos consistentes com a temática e necessários à implementação de estratégias local e nacional. Isto começa a partir da garantia do acesso universal à educação.

Embora, no Brasil, o ensino fundamental seja a escolaridade mínima obrigatória estabelecida na Constituição Federal de 1988 (Artigo 60, § 6º), a proporção da população adulta que é alfabetizada ainda é muito inferior aos países da Europa e Estados Unidos. A diminuição das desigualdades na educação conduz ao aumento do potencial humano, favorecendo a busca de um desenvolvimento social e redução da pobreza. A educação também estimula uma maior participação na vida política, desenvolve a consciência crítica e permite o discernimento, por parte dos cidadãos, de seus direitos e deveres para com a sociedade e o espaço que ocupam e no qual interagem, sendo agentes atuantes na organização e dinâmica do mesmo. Além disso, a inserção em um mercado de trabalho competitivo e exigente de habilidades intelectuais requer um maior nível de escolaridade e um ensino de qualidade.

Ainda que por si só a educação não assegure a produção e distribuição de riquezas, a justiça social e o fim das discriminações sociais, ela é, sem dúvida, parte indispensável do processo para tornar as sociedades mais prósperas, justas e igualitárias. A educação tanto pode legitimar um modelo perverso como pode aguçar a consciência dessas contradições contribuindo para sua superação no plano da realidade objetiva (SEVERINO, 2005).

O reconhecimento da associação entre educação e renda é consensual entre os pesquisadores da área, independentemente do termo teórico adotado. Ao estudar os determinantes da desigualdade de renda no Brasil, Ferreira (2000) conclui, com base nas

principais correntes de análise do tema no Brasil, que uma grande desigualdade educacional gera um alto nível de desigualdade de renda. Segundo o autor, a questão educacional no Brasil é caracterizada por um círculo vicioso, no qual uma grande heterogeneidade educacional gera uma grande desigualdade de riqueza, que se transforma em grandes diferenças de poder político, que por sua vez geram uma política educacional que perpetua a desigualdade educacional inicial.

4.4.5 Segurança: Homicídios e Morte no trânsito

Por fim, o Construto V do Desenvolvimento Social é a segurança, constituída de seus respectivos conceitos Homicídios e Morte no trânsito. O Homicídio tem sido provocado principalmente pela pressão social, consequência do modelo de desenvolvimento econômico vigente. Esse fenômeno social tem provocado o aumento da criminalidade nas grandes cidades e causado a redução da qualidade de vida dos cidadãos das regiões metropolitanas das grandes capitais. Diante da problemática da criminalidade, mensurar este fenômeno no tempo e no espaço passou a ser uma das preocupações entre os estudiosos do assunto. A Organização Mundial da Saúde (OMS) considera os acidentes de transporte como um dos maiores problemas de saúde pública no mundo, especialmente nos países em desenvolvimento, consequência da acelerada urbanização e motorização, não acompanhada, na mesma proporção, de infra-estrutura adequada.

A relação entre desocupação e mortalidade por violência vem sendo estudada nos últimos anos por muitos pesquisadores. Teixeira (2004), ao estudar as determinantes da violência no Brasil, observou que há uma associação negativa entre os indicadores, ou seja, quanto menor a taxa de ocupação maior a violência, e vice-versa. Porém, o desemprego, por si só, não é suficiente para a criminalidade. Igualmente, entre outras variáveis não menos relevantes, as desigualdades de renda possuem papel relevante para explicar os riscos de homicídios no Brasil.

De maneira a contribuir com o debate sobre as diversas formas de mensuração do desenvolvimento social, pressupõe-se, nesta pesquisa, através da aplicação do *Knowledge Discovery in Database (KDD)* em bases de dados públicas, ser possível uma maior aproximação do fenômeno, elegendo a análise de 15 conceitos e seus respectivos construtos

focados nas características de Dinâmica Populacional, Trabalho e Rendimento, Saúde, Educação e Segurança.

Embora condicionada às variáveis disponíveis nas bases de dados utilizadas, com o uso do método KDD é possível investigar o modelo conceitual de desenvolvimento social e como ele se relaciona com a realidade dos municípios brasileiros. A proposta apresentada para mensuração do desenvolvimento social através do uso do *KDD* nesta tese oferece a possibilidade de um olhar diferenciado para o fenômeno. A representação aqui proposta irá contribuir com o compartilhamento de idéias e pensamentos, bem como servir como padrão de comparação para futuros estudos, pois a utilização intensiva de técnicas de *Data Mining* para mensuração de fenômenos sociais ainda é pouco explorada na literatura. Essa configuração conforma o modelo de mensuração do Desenvolvimento Social no Brasil, a partir da qual são construídos e testados os modelos apresentados no Capítulo 7 deste trabalho.

O próximo capítulo apresenta, de maneira estruturada e resumida, o desenvolvimento da TI na área pública aplicada à tomada de decisão e, mais especificamente, discute suas formas de aplicação por meio do uso de informações sobre Desenvolvimento Social, disponibilizadas pelos governos por meio de bases de dados públicas.

Este capítulo apresenta e analisa as principais formas de mensuração encontradas nos índices e sistemas de indicadores de desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos. Na direção do objetivo proposto para o Capítulo, alguns objetivos específicos foram alcançados. Primeiro, realizar o levantamento, por meio de pesquisa bibliográfica e documental, dos mais conhecidos Índices e Sistemas de Indicadores de Desenvolvimento disponíveis na literatura nacional e internacional. A partir do levantamento, foram comparados os pressupostos conceituais dos 43 mais importantes Índices ou Sistemas de Indicadores internacionais e nacionais. Em seguida, foram selecionadas 46 dos mais conhecidos instrumentos nacionais que contemplam a dimensão Social do Desenvolvimento e analisados seus pressupostos teóricos, metodológicos e conceituais. Todos esses elementos foram utilizados para identificar e representar graficamente o termo teórico, os respectivos construtos e conceitos empregados nos instrumentos analisados.

5 KNOWLEDGE DISCOVERY IN DATABASES

In God we trust; all others must bring data. (W. Edwards Deming)

Como se discutiu nos capítulos anteriores, a utilização de Índices e Sistemas de Indicadores tem se constituído como importante instrumento na determinação de políticas públicas para o Desenvolvimento Social. Esses instrumentos de mensuração têm sido utilizados para analisar problemas, monitorar progressos e explicar mudanças sociais. Cada vez mais, a utilização de sistemas de indicadores tem se constituído importante elemento legitimador da agenda pública e social para o desenvolvimento.

A análise apresentada no Capítulo 4 mostrou que a construção de indicadores sociais no Brasil avançou significativamente no início deste século. Atualmente, os governos têm mais interesse em produzir e divulgar informações para mostrar os avanços decorrentes das políticas públicas adotadas. A adoção da Tecnologia da Informação e Comunicação (TIC) tem contribuído para disponibilizar aplicativos voltados a melhoria da prestação de serviços e informação, encorajando a participação dos cidadãos nos processos de tomada de decisão e tornando o governo mais transparente e efetivo, além de melhorar sua *accountability*.

No início deste século, o uso da TIC na esfera pública deixou de ter um papel restrito ao suporte administrativo, passando a ter, também, aplicações estratégicas nas tomadas de decisão, auxiliando, por exemplo, na implementação e avaliação de políticas governamentais. O processo de globalização, a internet e, no Brasil, a consolidação da democracia tornaram os cidadãos mais exigentes, ao mesmo tempo o mercado tornou-se mais competitivo e o cidadão passou a demandar mais do Poder Público em defesa de seus direitos. O aumento da procura por informações e a necessidade legal de maior transparência nas ações do gestor público culminaram em crescente disponibilização de informações por parte dos principais órgãos de governo na esfera federal, levando, gradativamente, as unidades da federação a, também, estruturarem e disponibilizarem mais informações à sociedade.

Entre os movimentos recentes da Tecnologia da Informação e Comunicação na esfera pública está o desenvolvimento de sistemas computadorizados que permitem análises e tomada de decisão a partir de bases de dados públicas disponibilizadas na Rede Mundial de Computadores.

Este capítulo tem o objetivo de apresentar, de maneira estruturada e resumida, o desenvolvimento da TI na área pública aplicada à tomada de decisão e, mais especificamente, discutir suas formas de aplicação através do uso de informações sobre Desenvolvimento, disponibilizadas pelos governos por meio de bases de dados públicas.

Inicialmente, são apresentados conceitos relacionados ao estudo da Gestão do Conhecimento e Sistemas de Informação procurando relacioná-los à sua aplicação na tomada de decisão. Com esta finalidade, são discutidos, ao longo do capítulo, temas como o Estudo e a Gestão do Conhecimento (GC); Sistemas de Informação (SI); Tecnologia da Informação (TI); Sistemas de Apoio à Decisão (SAD); Descoberta de Conhecimento em Bancos de Dados ou *Knowledge Discovery in Database (KDD)* e Mineração de Dados ou *Data Mining (DM)*.

5.1 GESTÃO DO CONHECIMENTO E SISTEMAS DE INFORMAÇÃO

Em geral, a gestão do conhecimento pode ser definida como o conjunto de processos para identificar o conhecimento que está presente nas pessoas e proporcionar condições adequadas para sua transferência, utilização e criação (LIEBOWITZ e BECKMAN, 1998; BECKMAN, 1999). Já para Davenport e Prusak (1998), a gestão do conhecimento é o conjunto de atividades relacionadas com a geração, codificação e transferência do conhecimento.

A discussão sobre o conhecimento, apesar de sua aparente modernidade, é, na realidade, milenar. Ainda que não se possa traçar um paralelo direto, há uma aparente relação da visão do primeiro grupo de autores com a linha do racionalismo de Platão, enquanto a conceituação de Davenport e Prusak encontraria maior respaldo no empirismo de Aristóteles.

Davenport e Prusak (1998) afirmam que a gestão do conhecimento (GC) deve ter os seguintes objetivos: 1) criar um repositório de conhecimento constituído por: conhecimento externo e conhecimento interno estruturado; 2) melhorar o acesso ao conhecimento; 3) desenvolver um ambiente e uma cultura organizacional propícios à criação, à transferência e ao uso do conhecimento e 4) tratar o conhecimento como um recurso mensurável.

A literatura apresenta diversas outras definições sobre Gestão do Conhecimento. Claramente, pode-se perceber, na atualidade, a existência de duas correntes principais: i) a do suporte tecnológico e ii) a do comportamento. Na corrente tecnológica parece haver um

predomínio de autores com formação na área de Tecnologia da Informação. Tais autores enfocam mais os conceitos de armazenamento, reaproveitamento e descoberta do conhecimento em detrimento de uma abordagem mais comportamental relacionada ao elemento humano adotada no segundo grupo. Assim sendo, parece, novamente, ser possível perceber maiores relações do primeiro grupo – o da TI – com o empirismo (de Aristóteles, na antiguidade, e Davenport e Prusak, na atualidade), enquanto, por outro lado, a corrente do comportamento encontraria maior suporte no racionalismo de Platão (na antiguidade) e em autores contemporâneos (LIEBOWITZ e BECKMAN, 1998; BECKMAN, 1999).

Polanyi (1958), talvez o mais importante autor da GC na atualidade, classifica o conhecimento humano em dois tipos: o conhecimento explícito, que pode ser articulado na linguagem formal, inclusive em afirmações gramaticais, expressões matemáticas, especificações e manuais, entre outros, e o conhecimento tácito, mais difícil de ser expresso na linguagem formal. Esta segunda corrente, identificada por Nonaka e Takeuchi (1997), claramente melhor se coaduna com a corrente racionalista do pensamento grego e, dentro desta, na abordagem do comportamento com os trabalhos, por exemplo, de Liebowitz e Beckman.

Com base na discussão apresentada, pode-se classificar o presente estudo na linha de pensamento do empirismo de Aristóteles e da tecnologia de Davenport e Prusak. Em um esforço para tornar ainda mais clara a inserção da presente pesquisa no vasto campo da GC, considerou-se o trabalho de O'Dell e Grayson (2000). Estes autores dividem a aplicação das ferramentas de Tecnologia na Gestão do Conhecimento em duas subclasses: i.1) Transmissão e Troca de Conhecimento e i.2) Análise de Dados e Suporte ao Desempenho.

Considerando-se as duas subclasses de O'Dell e Grayson, o presente trabalho tem foco no segundo grupo (i.2), mais especificamente em Exploração de Dados, Suporte à Decisão e Análise de Dados, que é considerado por esses autores como “O Território Inexplorado da Gestão do Conhecimento.” (O'DELL e GRAYSON, 2000, p.124).

5.2 SISTEMAS DE INFORMAÇÃO

De acordo com Laudon e Laudon (1994), o estudo de Sistemas de Informação (SI) constituiu um campo multidisciplinar. Este novo campo lida com questões e reflexões derivadas de disciplinas como Sociologia, Economia e Psicologia, no campo comportamental, e disciplinas como Ciências da Computação, Pesquisa Operacional e Ciências da Administração, nas abordagens técnicas.

O conceito de sistemas de informação (SI) tem evoluído substancialmente, fugindo de uma visão puramente técnica para uma visão social, mesmo que a palavra social ainda tenha um sentido vago na ciência da computação (IVANOV, 1998). É importante lembrar, também, que esta evolução e as orientações de pesquisa se diferenciam significativamente de um país para outro e de uma escola de pensamento para outra, não havendo, portanto, um paradigma universal de pesquisa em Informática Social.

Múltiplas perspectivas contribuem para a formação do conceito de Informática Social como área de estudo dos diferentes aspectos sociais das atividades computadorizadas nas organizações. Para Friedman e Kahn (1999), as preocupações éticas e sociais devem ser partes integrantes do desenvolvimento de sistemas de computadores. Portanto, se a tecnologia da informação tem um grande potencial para alterar nossas vidas, o desenvolvimento da informática social é uma oportunidade que não podemos simplesmente ignorar (SCHULER, 1994).

5.2.1 Sistema de apoio à decisão

O avanço tecnológico propiciou a redução dos custos e a difusão dos computadores. Conseqüentemente, houve um aumento da capacidade de coleta e armazenamento de dados não ocorrendo um aumento simultâneo e equivalente na capacidade de utilizar esses dados. Em meio a essa dinâmica, cresceu a demanda por diferentes sistemas de informação para apoiar a tomada de decisões, surgindo assim os chamados Sistemas de Suporte à Decisão (SSD), nesta tese denominado como Sistemas de Apoio à Decisão (SAD).

As definições de SSD e SAD podem ser reunidas em dois extremos conceituais: o de escopo mais amplo, onde os SADs “[...] são aqueles que contribuem de alguma forma para tomada de decisão”, e o de interpretação mais restrita, pelo qual “SADs são sistemas baseados em computador, interativos, que auxiliam gerentes na utilização de dados através de modelos para resolver problemas não estruturados” (SPRAGUE; WATSON, 1991, p.78).

A definição de SAD adotada neste trabalho é: sistemas que se utilizam de TI para tratar dados ou informações pouco estruturadas, de forma sistemática, visando transformá-las em conhecimento ou informações mais estruturadas destinados a apoiar a tomada de decisões.

Dhar e Stein (1997, apud Laudon e Laudon, 1999) reconhecem dois tipos básicos de SAD: o primeiro, chamado de SAD guiado por modelo, caracteriza os sistemas desenvolvidos de maneira isolada dos principais sistemas de informação da organização. Estes sistemas são baseados numa forte teoria ou modelo que se combina com uma boa interface, facilitando a execução pelo decisor através de simulações e outros tipos de análises.

Já o segundo tipo de SAD, o guiado por dados, é mais recente e voltado para a extração de informações úteis previamente desconhecidas independentemente da existência de um modelo prévio. Neste grupo, podem ser encontradas ferramentas como OLAP e *Data Mining*.

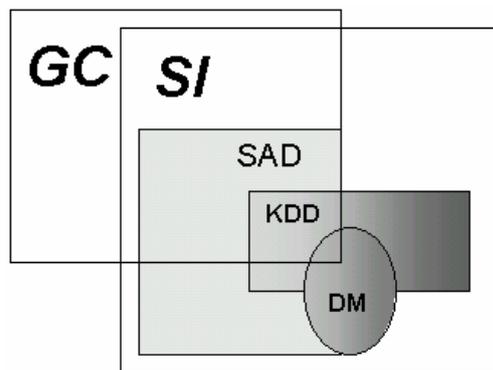
5.2.2 Características dos sistemas de apoio à decisão

Apesar de haver grande concordância entre as definições adotadas neste trabalho e os conceitos mais amplamente utilizados na literatura de Sistemas de Informação, propõe-se aqui uma abordagem distinta daquela feita por alguns importantes autores. Por exemplo: para Damiani (1998), os Sistemas de Informação podem ser divididos em três categorias básicas: os sistemas de apoio à gestão estratégica, os sistemas de apoio à gestão tática e os sistemas de apoio à decisão de nível operacional. Segundo este autor, a aplicação de SAD ocorre apenas no nível tático. Diferentemente do que preconiza Damiani (1998) e de acordo com a definição aqui adotada, entende-se que também o nível estratégico, e não apenas o tático, requer sistemas de apoio à decisão. Outra distinção entre a abordagem de Damiani (1998) e a aqui adotada é a clara dicotomia entre os três níveis de gestão assumida pelo referido autor. Tal dicotomia parece, hoje, um tanto quanto extemporânea, já que no paradigma da administração

estratégica (no qual se insere o presente trabalho) se pressupõe que mesmo a gestão operacional deve ser vinculada e sincrônica às grandes estratégias organizacionais.

Com base nas definições acima, optou-se nesta pesquisa por uma perspectiva sistêmica, a qual engloba todo o processo de descoberta de conhecimento útil em bases de dados. Visando oferecer uma melhor percepção dos principais conceitos relacionados a esta tese, procurou-se identificar a relação existente entre: Gestão do Conhecimento; Sistema de Informação; Sistema de Apoio à Decisão; Descoberta de Conhecimento em Base de Dados e *Data Mining*. Essas relações são ilustradas na Figura 6, a seguir.

Figura 6 – Gestão do Conhecimento, Sistemas de Informação, Sistemas de Apoio à Decisão, Knowledge Discovery in Databases, Data Mining e suas interfaces.



Fonte: Adaptado de Quintella e Soares Jr. (2003, p.10).

Conforme foi definido anteriormente, a corrente do suporte tecnológico na gestão do conhecimento pode utilizar os sistemas de informação que, por seu turno, possuem um tipo mais específico que é o SAD. Este tipo de Sistema incorpora ferramentas analíticas avançadas, possibilitando simulações e elaboração de cenários. Assim sendo, os SADs envolvem, mas não limitam, o processo de KDD, metodologia que, por sua vez, utiliza o *Data Mining* (DM) como uma de suas ferramentas ou técnicas. Observa-se, por fim, que o *Data Mining* pode ser utilizado em processos outros que não o KDD, não estando, portanto, por ele limitado, conforme mostra a Figura anterior.

As seções seguintes apresentam os aspectos teóricos da *Knowledge Discovery in Database – KDD*⁴⁰, método capaz de apreender a natureza relacional do desenvolvimento humano, e sua conexão com temas direta ou indiretamente relacionados com a identificação das principais propostas de etapas do processo *KDD* disponíveis na literatura.

⁴⁰ A sigla KDD é empregada neste trabalho por ser amplamente difundida na literatura.

5.3 ASPECTOS TEÓRICOS DO MÉTODO KDD

O crescimento rápido do volume, em tamanho e dimensionalidade, das bases de dados criou a necessidade e a oportunidade para extrair conhecimento destas. Neste contexto, surge no final da década de oitenta, um novo ramo da computação, conhecido como *KDD* – *Knowledge Discovery in Database*, com o objetivo principal de encontrar uma maneira automatizada de explorar essas bases de dados e reconhecer os padrões existentes através da modelagem de fenômenos do mundo real.

When the scale of data manipulation, exploration, and inference grows beyond human capacities, people look to computer technology to automate the bookkeeping. The problem of knowledge extraction from large databases involves many steps, ranging from data manipulation and retrieval to fundamental mathematical and statistical inference, search, and reasoning. Although the problem of extracting knowledge from data (or observations) is not new, automation in the context of large databases opens up many new unsolved problems. (FAYYAD et al., 1996, p.28).

O *KDD* refere-se às etapas que produzem conhecimentos a partir de dados relacionados e sua principal característica é a extração não-trivial de informações implicitamente contidas em uma base de dados. Essas informações são de difícil detecção por métodos tradicionais de análise e devem ser potencialmente úteis para tomada de decisão (FRAWLEY et al., 1992; FAYYAD et al., 1996). Enquanto os métodos tradicionais são capazes de tratar apenas as informações explícitas, o *KDD* é capaz de detectar informações implícitas armazenadas nos Bancos de Dados.

O processo de *KDD* é iterativo e, embora apresente uma definição semelhante também ao *DM*, deve ser composto de uma série de etapas sequenciais, podendo haver retorno a etapas anteriores, isto é, às descobertas realizadas (ou a falta delas). Este processo conduz, eventualmente, a novas hipóteses e descobertas. Neste caso, o usuário pode decidir pela retomada dos processos de mineração, ou uma nova seleção de atributos, por exemplo, para validar as hipóteses que surgiram ao longo do processo.

O produto esperado do *KDD* é uma informação relevante para ser utilizada pelos tomadores de decisão. Alguns autores, porém, defendem o ponto de vista de que o conhecimento descoberto não precisa necessariamente ser incorporado a um Sistema de Apoio à Decisão – SAD (EVANS; FISCHER, 1994).

A crescente demanda por prospecção e tratamento de dados fomentou a discordância entre os pesquisadores de diversas áreas e a conseqüente procura por definição, formalização e padronização de procedimentos. Estes temas foram discutidos durante os encontros científicos sobre *KDD*, realizados entre os anos de 1989 e 1994 e que culminaram na realização da primeira Conferência Internacional de Descoberta de Conhecimento e Mineração de Dados (*First International Conference on Knowledge Discovery and Data Mining*) no ano de 1995, na cidade de Montreal no Canadá.

São encontradas na literatura diversas nomenclaturas alternativas, muitas delas excêntricas, para o *KDD*. Este fato pode ser facilmente justificado pela multidisciplinaridade do tema e pelas sucessivas contribuições que o campo tem recebido, nos últimos anos, das mais diversas áreas do conhecimento. A tradução mais difundida para *KDD* na literatura é Descoberta de Conhecimento em Bases de Dados, também conhecido pela sigla DCBD, muitas vezes confundida com Mineração de Dados ou *Data Mining*, também conhecido como colheita de informação (*information harvesting*), extração de conhecimento de informação, processamento de padrões de dados, dragagem de dados (*data dredging*), aprendizado de máquina e arqueologia de dados, entre outros.

Sprague e Watson (1991) sugerem uma “dicotomia intuitiva” na classificação das atividades de processamento de informações, sendo um grupo de atividades caracterizado por possuírem uma estrutura definida e referir-se a problemas de natureza eminentemente algorítmica. Já o segundo grupo seria aquele que se caracteriza pelas atividades de acompanhamento, monitoramento, aconselhamento, resolução de problemas, análise e elaboração, portanto de natureza eminentemente heurística. De um modo geral, por conseguinte, o *KDD*, em sua visão mais ampla, contempla as atividades heurísticas e algorítmicas ou matemáticas.

O *KDD* tem sido utilizado, por exemplo, na área financeira, na qual, segundo Bose e Mahapatra (2001), é empregada principalmente para prever o futuro das finanças relacionadas aos sistemas bancários. Neste mercado, pode ser empregada para prever a falência tanto de clientes como do próprio banco; estimar a confiança que se pode ter em um usuário de cartão de crédito; detectar regras de negócios; aprovar e identificar fraudes em cartão de crédito; classificar clientes, etc. Embora os governos possuam bases de dados de grande volume e importância como, por exemplo, previdenciárias, demográficas, de contribuintes, e muitas outras, são raros os registros na literatura sobre a aplicação de *KDD* em bases de dados públicas brasileiras.

O *KDD* é de interesse comum a diversas áreas, e as primeiras contribuições científicas e técnicas foram apresentadas por pesquisadores de áreas como: aprendizado de máquinas; banco de dados inteligente; computação de alta *performance*; estatística; inteligência artificial; visualização de dados; reconhecimento de padrões e sistemas especialistas. Foram desenvolvidas aplicações de *KDD* também para astronomia, biologia, seguros, marketing, medicina e muitos outros campos.

5.4 FUNCIONAMENTO DO KDD

Normalmente, o processo *KDD* envolve três atores de diferentes habilidades: o usuário, o especialista do domínio e o analista de dados. O usuário é o demandante do trabalho e irá desfrutar dos resultados obtidos, pode ser uma pessoa, um departamento, uma empresa, uma comunidade, um país⁴¹. O especialista do domínio conhece o domínio que será estudado, normalmente um pesquisador ou profissional com larga experiência no tema. Em alguns casos, o usuário e o especialista são a mesma pessoa. Por fim, o analista de dados deve conhecer o funcionamento das ferramentas e dos algoritmos utilizados e, em função do seu nível de especialização, dificilmente conhece o domínio em questão.

Na abordagem de Inmon et al. (2001), o usuário também é visto como especialista do domínio e a relação entre Usuário e Analista de Dados deve-se processar segundo a dialética de Hegel. O analista de dados deve executar o processo de verificação (tese) ou anulação das hipóteses (antítese) criadas pelo usuário, que, por seu turno, irá entrar no ciclo de reformular a hipótese (síntese) para que seja testada novamente, seguindo uma trajetória em espiral até a finalização do processo. Antes de Hegel, os gregos, citados no início do Capítulo 2, já utilizavam esses conceitos explicitamente. Em outras palavras, pode-se dizer que a simples observação dos dados e a criação de uma hipótese caracteriza um ato indutivo, já o inverso – procurar determinar, com base nos dados, se uma hipótese é verdadeira – caracteriza um ato dedutivo.

Esta pesquisa tem como um de seus objetivos **identificar as principais etapas do processo *KDD***. Neste trabalho, seguindo as conclusões de Inmon et al. (2001) e Fayyad et al. (1996), é entendido que o sucesso de um processo *KDD* depende diretamente da participação

⁴¹ No caso de aplicar o processo *KDD* em bases de dados públicas, percebe-se o universo de usuários de maneira ampliada.

dos atores e da maneira com que é conduzida cada etapa do processo. Devido às características interdisciplinares do *KDD* e de sua dependência por tecnologia, ferramentas e técnicas de análise, optou-se inicialmente por apresentar, previamente, algumas definições consagradas na literatura.

5.4.1 Etapas da descoberta de conhecimento em bancos de dados

O termo *KDD* refere-se ao processo global de descoberta de conhecimento e, diante da profusão de conceitos, é esperado que não haja consenso em relação às etapas que devem ser observadas durante seu processo. Para chegar a uma solução apropriada para análise de dados de um problema, muitos passos precisam ser repetidos e refinados sucessivamente. Independentemente da abordagem, o *KDD* é composto por um conjunto de etapas que, de um modo geral, podem ser reunidas em três fases: preparação, análise e interpretação (ADRIAANS; ZANTIGE 1996; BRACHMAN, 1996; FAYYAD et al., 1996; HAN; KAMBER, 2000).

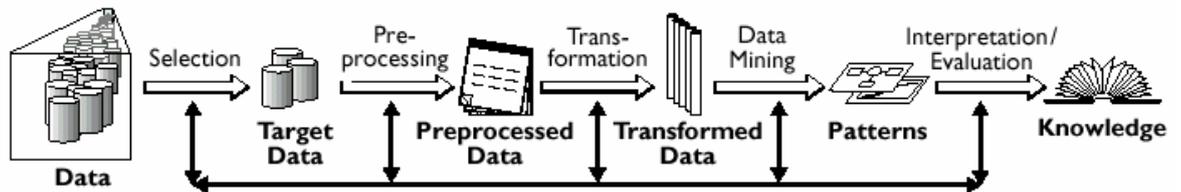
Todos os passos são críticos, sendo que o mais complexo deles é a Mineração de Dados ou *Data Mining*, inserido na fase de Análise e que tem como objetivo encontrar padrões nos dados armazenados. Esta fase é normalmente confundida com o próprio processo de *KDD* (HAN; KAMBER, 2000).

O primeiro registro sobre descrição de processos de *KDD* data de 1996 e pode ser atribuído ao artigo intitulado “The KDD Process for extracting useful knowledge from volumes of data” de autoria dos pesquisadores Usama Fayyad, Gregory Piatetsky-Shapiro e Padhraic Smyth. Os autores definem que o *KDD* deve ser interativo (ação recíproca) e iterativo (repetitivo), bem como descrevem o processo em nove etapas. A cada etapa, segundo eles, os atores devem incorporar sua experiência através da tomada de decisões, com o objetivo de obter melhores resultados. Resumidamente, as etapas (Figura 7) podem ser apresentadas como:

- 1 – desenvolvimento e entendimento do domínio da aplicação: evidencia a necessidade da participação de um especialista no domínio a ser investigado para formular as metas da aplicação;

- 2 – criação de um arquivo de trabalho: envolve a delimitação do universo, seleção dos registros e variáveis que fazem parte do escopo de trabalho;
- 3 – limpeza e pré-processamento dos dados: coleta da informação necessária para modelar, sendo definida a estratégia para tratar e mapear ruídos, subcamadas e campos perdidos ou desconhecidos. Nesta etapa, são definidas questões ligadas ao Sistema Gerenciador de Banco de Dados – SGBD;
- 4 – projeção e redução dos dados: busca por representações invariantes para os dados através de formas práticas; redução de dimensionalidade e métodos de transformação visando reduzir o número de variáveis que serão levadas em consideração;
- 5 – seleção da função (tarefa, modelo) do *DM*: escolha da função (sumarização, classificação, regressão e modularização) derivada do algoritmo de *Data Mining*;
- 6 – seleção do algoritmo para a função: definir de que maneira serão selecionados os métodos e parâmetros apropriados para procurar por modelos nos dados. Nesta etapa, o usuário pode estar interessado em compreender apenas o modelo ao invés das suas capacidades;
- 7 – execução da Mineração de Dados: consiste na busca por uma ou mais formas de representação para as funções de interesse, incluindo regras de classificação ou árvores, modelos de seqüência, dependência e análise linear, regressão e agrupamento;
- 8 – interpretação dos padrões minerados (encontrados): remoção de modelos redundantes e irrelevantes; interpretação e classificação como úteis daqueles compreendidos pelos usuários;
- 9 – consolidação do conhecimento descoberto: consiste da simples documentação e disseminação dos resultados até a incorporação deste conhecimento na *performance* do sistema, procurando resolver conflitos com o conhecimento anterior.

Figura 7 – Etapas do processo KDD



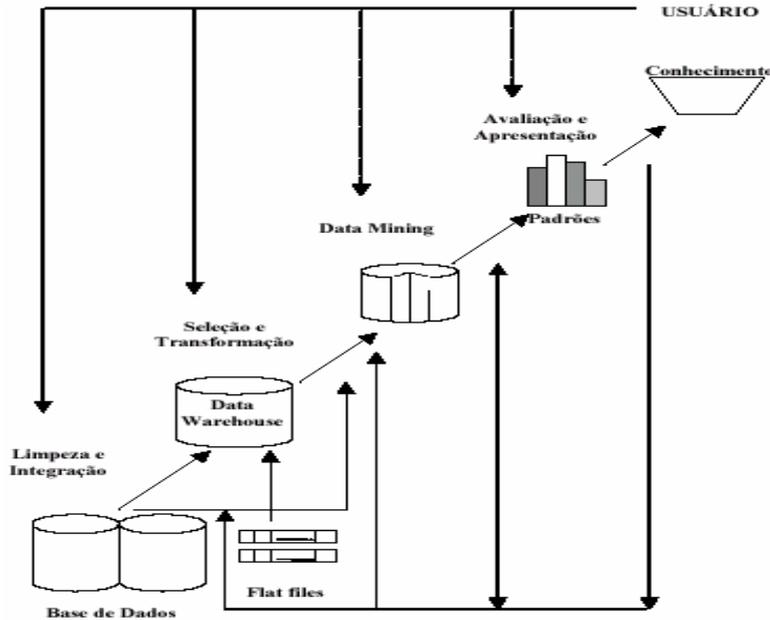
Fonte: Fayyad et al. (1996, p.29).

Fayyad et al. (1996, p.30), por outro lado, demonstram a preocupação de sistematizar as etapas do processo *KDD* já que, segundo eles: “A maioria dos trabalhos anteriores sobre *KDD*, dava ênfase principalmente a etapa de *Data Mining*. No entanto, os outros passos são igualmente, se não mais importantes para o sucesso da aplicação de *KDD* na prática”. A proposta de Fayyad et al. (1996) ampliou o escopo do *KDD*, antes deles o *Data Mining* limitava-se a resolver as questões inerentes às etapas 5 (Seleção da função), 6 (Seleção do algoritmo para a função), 7 (Execução da Mineração de Dados) e 8 (Interpretação dos padrões minerados).

Em um outro importante trabalho sobre o tema, Han e Kamber (2000) alertam para a importância de um *data warehouse* previamente concebido e para a necessidade de técnicas específicas na apresentação do conhecimento “descoberto”. Estes autores apresentam o processo *KDD* em sete etapas (Figura 8): 1 – limpeza: corrigir dados inconsistentes, tratar valores nulos e *outliers*⁴²; 2 – integração dos dados: compatibilização de formato das diversas bases de dados, como num *data warehouse*, devendo ser considerada a existência de conflitos de dados, a redundância e a integração do esquema; 3 – seleção dos dados: selecionar dados relevantes à tarefa de análise; 4 – transformação dos dados: traduzir para formas compatíveis para a etapa seguinte (mineração) através de agregações e sumarizações; 5 – *Data Mining*: a etapa de mineração consiste da extração de padrões através da aplicação de técnicas de associação, classificação, agrupamentos etc; 6 – avaliação dos padrões: tem o objetivo de verificar a serventia dos achados, avaliando a relevância para o estudo; 7 – apresentação do conhecimento: através de técnicas específicas (tabelas, gráficos, cubos, grafos, etc.) de representação e visualização, o conhecimento é apresentado.

Figura 8 – Etapas do Processo *KDD*

⁴² Termo amplamente empregado em *Data Mining* para caracterizar valores extremos ou atípicos à distribuição, valores que possuem características diferenciadas devendo, portanto, receber atenção especial.

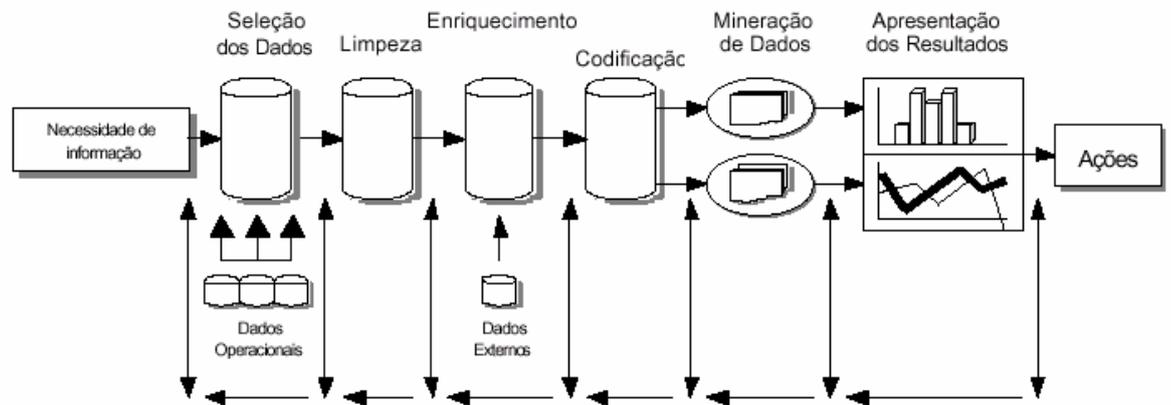


Fonte: Adaptado de Han e Kamber (2000, p.16).

Diferentemente das propostas mais robustas apresentadas até aqui, a abordagem de Adriaans e Zantige (1996) não suscita uma etapa de compreensão do domínio e dos objetivos do trabalho. Nesta abordagem, mais simplificada e claramente voltada para o meio empresarial, os autores baseiam-se na dinâmica organizacional, pela qual a obtenção de informações ocorre de maneira contínua e a especialização dos atores envolvidos no processo se dá apenas de forma natural.

Na abordagem de Adriaans e Zantige (1996), o *KDD* é percebido em um escopo mais reduzido, porque as questões a serem respondidas no processo são rotineiras. Por outro lado, para estes autores, o processo deve ser empregado continuamente na organização, sempre enfocando a necessidade de iteratividade, em que cada uma das suas seis fases é resultado da que a precede.

Figura 9 – Processo KDD



Fonte: Adaptado de Adriaans e Zantige (1996, p.38).

Como pode ser observado na Figura 9, a necessidade de informação resulta em uma seleção de dados, após uma análise prévia dos dados operacionais do banco de dados, sendo feita a seleção daqueles necessários ao início do processo. Adriaans e Zantige (1996) se diferenciam dos demais autores apresentados, por evidenciarem a necessidade de um dinamismo para o processo, pois em qualquer etapa os dados podem ser incluídos, alterados ou descartados. O dinamismo proposto pelos autores suscita a idéia de que o processo *KDD* pode ser executado inúmeras vezes durante um curto espaço de tempo.

Sobre essa proposta, deve-se ressaltar ainda que os autores limitam a aplicação do *KDD* ao banco de dados da organização, ou admitem um *Data Warehouse* plenamente implantado, ressaltando a necessidade de uma etapa de enriquecimento após a limpeza dos dados. Nas outras abordagens apresentadas, o enriquecimento é feito previamente, no momento da seleção dos dados que serão trabalhados.

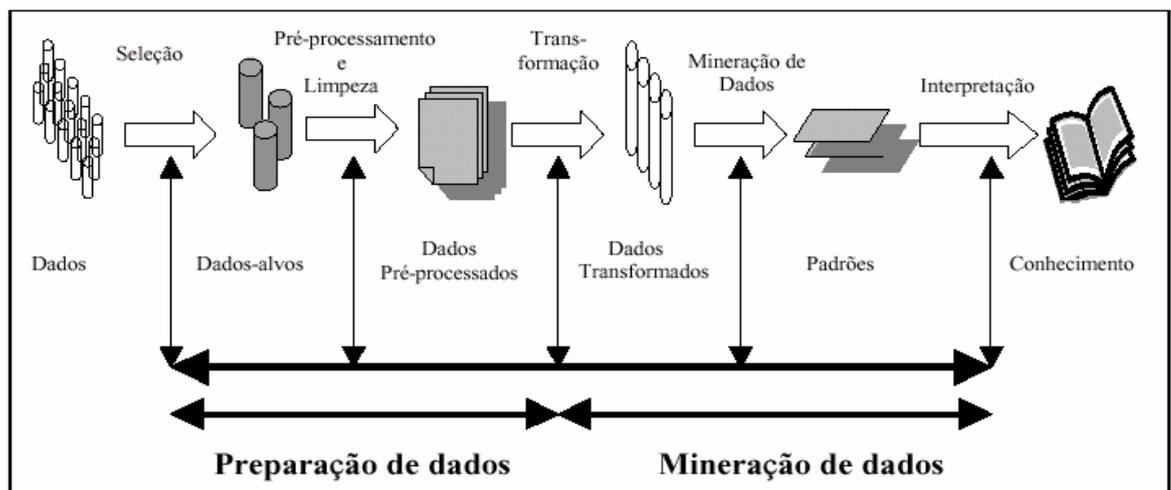
Somente depois de selecionados os dados é que Adriaans e Zantige (1996) propõem o início da etapa de codificação com o objetivo de formatar e recodificar os dados para atender às exigências dos algoritmos da etapa seguinte, o *Data Mining*. Na etapa de Mineração de Dados é que se extrai efetivamente o conhecimento propriamente dito através de uso intensivo de recursos computacionais. Os autores afirmam que 80% do conhecimento é extraído com uma “análise menos trivial” através de consultas *ad hoc* com o uso de ferramentas *SQL*, só então se devendo utilizar técnicas mais avançadas.

Comumente, os resultados obtidos da etapa de Mineração do *KDD* são apresentados em formato específico e pouco inteligível, principalmente no caso das abordagens mais complexas. Muitas vezes, é necessário ao Analista de Dados que opera a ferramenta, traduzir os códigos de resposta, obtidos através dos *softwares* empregados, e exibi-los de maneira a facilitar o entendimento do especialista do domínio e/ou do usuário demandante.

Ao se observar a abordagem típica de pesquisadores de TI, como Adriaans e Zantige (1996), pode-se constatar que, para operacionalizar o *KDD* em grandes bases de dados ou em bases de múltiplas origens (como é o caso da maior parte das Bases de Dados Públicas), é necessário efetuar pesados investimentos em TI.

Entre as abordagens revisadas na literatura, Amaral (2001) apresenta uma das mais simplificadas, que procura descrever e agrupar todo processo em duas fases principais (Figura 10). A primeira dessas fases envolve a Preparação e a segunda, a Mineração dos Dados. Durante a Preparação, é realizada a seleção dos dados relevantes para a solução do problema, sendo feita, em seguida, a limpeza através do pré-processamento, eliminando-se os dados incompletos, problemas de tipagem, repetição de tuplas⁴³ etc. Após a limpeza, segundo esses autores, deve ser feito o pré-processamento, quando os dados são dispostos em uma forma uniformizada para a fase seguinte de mineração de dados.

Figura 10 – Fases e Subfases do processo *KDD*



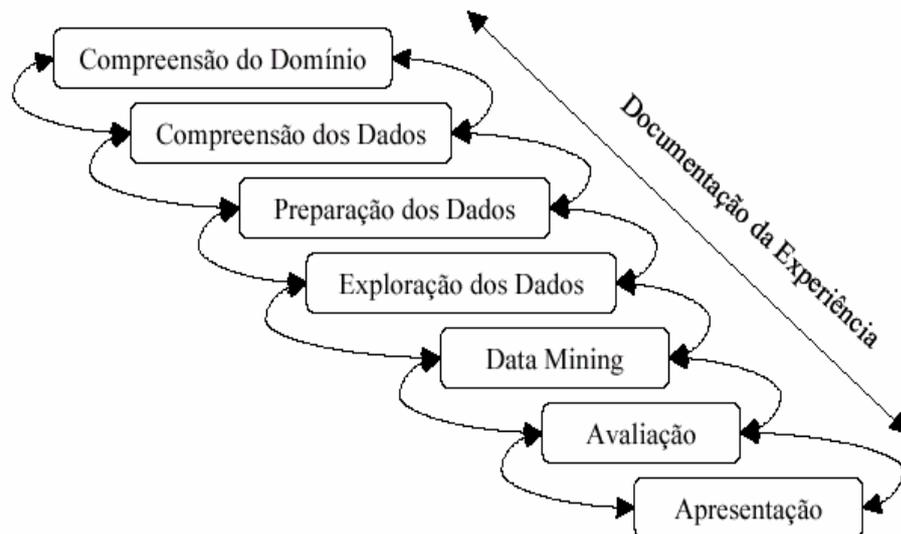
Fonte: Adaptado de Amaral (2001, p.42).

⁴³ O modelo relacional originalmente criado por Codd fez uso muito deliberado de certos termos, muitos deles pouco familiares nos círculos da Tecnologia da Informação. A necessidade de utilizar termos mais precisos, visando evitar termos que admitam diversos significados, levou o pesquisador a adotar “tupla” (rima com dupla) para identificar o equivalente a “registro” no jargão tradicional de T.I.

A escolha da técnica ou algoritmo a ser aplicado inicia a fase de preparação e a fase de mineração dos dados. Na busca por padrões nos dados e de acordo com os objetivos do trabalho, podem ser empregadas funções⁴⁴ isoladas ou combinadas de agrupamento, classificação, regressão, sumarização, visualização e até mesmo mineração de textos (*text mining*). Durante a busca, cada resultado encontrado é registrado através de relatórios de descobertas e, com o auxílio de técnicas de visualização, os analistas de mineração procuram interpretar as informações para, só então, obter o conhecimento.

Vale ainda ressaltar a contribuição de Reinartz (1999), que divide o processo também em sete etapas (Figura 11): 1 – compreensão do domínio; 2 – compreensão dos dados; 3 – preparação dos dados; 4 – exploração dos dados; 5 – *Data Mining*; 6 – avaliação e 7 – apresentação. Reinartz (1999) ressalta a necessidade de documentar a experiência adquirida durante todo o processo e sua iniciativa trouxe, definitivamente, a necessidade de documentação para as fases do processo, coisa que, na maior parte das propostas já apresentadas, era desprezada.

Figura 11– Esquema do processo de extração de conhecimento



Fonte: Adaptado de Reinartz (1999, p.25).

Outra importante contribuição à área é atribuída a Andrienko e Andrienko (1999), que vêem o tratamento de conjuntos de dados com atributos temporais e espaciais como um

⁴⁴ As funções utilizadas na Mineração de Dados são descritas a seguir, especificamente no tópico *Data Mining* ou Mineração de Dados.

caso particular de *KDD*. A visão dos autores se insere em um momento de crescente disponibilização de informações georreferenciadas.

Na literatura brasileira da área, o que se observa é uma quase total ausência de trabalhos com abordagem de *KDD*. Praticamente, toda a pesquisa nacional tem-se desenvolvido com enfoques de mineração de dados (os principais trabalhos brasileiros nesta área são revisados na seção 3.3, a seguir) e *business intelligence*. Como, talvez, o único trabalho nacional de *KDD* em literatura da área de administração há o artigo em que Quintella e Soares Jr. (2003, p.89), descrevem o *KDD* como “[...] o processo não-trivial para geração de conhecimento a partir da busca sistemática de padrões em grandes volumes de dados”, dividindo o processo em apenas três macroetapas:

- a) **Pré-processamento:** atividades que visam gerar uma representação conveniente para os algoritmos de mineração e atender aos objetivos da pesquisa a partir da base de dados. Inclui a seleção (automática e/ou manual de atributos relevantes), limpeza, amostragem, transformações de representação, discretização de atributos quantitativos, etc. Resulta na criação de uma base de dados específica para atender à Mineração de Dados;
- b) **Mineração de dados:** aplicação de algoritmos para descoberta de padrões (mais detalhadamente descrito a seguir);
- c) **Pós-processamento:** seleção e ordenação das descobertas relevantes, mapeamentos de representação de conhecimento, geração de relatórios e interpretação dos resultados encontrados.

Com base neste último trabalho e, principalmente, na simplicidade da abordagem de Amaral (2001), procedeu-se uma adaptação nas fases de Quintella e Soares Jr. (2003), subdividindo o processo *KDD* em apenas duas grandes fases conceituais: Prospecção e Mineração de dados.

As etapas operacionais identificadas para a fase de prospecção são, respectivamente: 1 – Identificação de Objetivos; 2 – Levantamento; 3 – Reunião; 4 – Seleção; 5 – Criação da base de dados; 6 – Consistência das bases de dados; 7 – Compatibilização das bases de dados. Já a fase de Mineração contempla as seguintes etapas operacionais: 1 –

Transformação dos Dados; 2 – Função e Algoritmos de Mineração; 3 – Avaliação dos resultados; 4 – Disseminação.

Após essa breve revisão de alguns dos principais trabalhos de *KDD* encontrados na literatura, apresenta-se, em seguida, a fase de Mineração de Dados. Mesmo não comungando a visão de Han e Kamber (2000) de que esta seria a mais importante etapa no processo *KDD*, reconhece-se aqui que é a fase onde ocorre, usualmente, grande parte das descobertas de conhecimento em bancos de dados.

5.5 MINERAÇÃO DE DADOS – DATA MINING

Como foi visto na seção anterior, o *Data Mining* ou Mineração de Dados (MD), neste trabalho é tratado como uma das etapas da Descoberta de Conhecimento em Bases de Dados – DCBD. Erroneamente, o termo “Mineração de Dados” tem sido utilizado com alguma frequência para identificar o processo *KDD*, porém, como já foi abordado, trata-se de uma etapa entre as demais.

Explicar *Data Mining* à luz da teoria existente não é uma tarefa muito fácil, é preciso lidar com um conceito originado na área de Sistemas de Informação e que recebeu, ao longo dos anos, significativas contribuições da Estatística, da Inteligência Artificial, da Engenharia, da Administração e, em especial, da Ciência da Informação. Para traduzir, de maneira clara e concisa, não só através de citações e definições, a genealogia do conceito, torna-se apropriado recorrer a uma analogia.

Segundo Spence (2001), a visualização é uma atividade interna, de cognição do ser humano, que conduz à formação de um modelo mental com base em um estímulo visual recebido através de uma imagem. Na perspectiva de Tenório (1998), é uma representação analógica que fornece, sobre um objeto, o máximo de informação, enquanto uma representação digital fornece o mínimo de informação – produtora de imagem. O autor ainda defende o ponto de vista de que uma analogia vale mil palavras. As Figuras 12 e 13 oferecem uma analogia para auxiliar no entendimento do conceito *Data Mining*.

Diante dos dados apresentados na Figura 12, é difícil identificar padrões, entretanto, com o uso das técnicas apropriadas, o pesquisador poderá deixar apenas de **olhar**

para poder passar a **enxergar** algo mais: “*Data! Data! Data! I cannot make bricks without clay!*” – Sherlock Holmes⁴⁵ (DOYLE, 1982, p.286).

Figura 12 – Base de dados antes da Descoberta de Conhecimento – DCBD

J	K	P	W	Q	O	R	S	B	C	X	X	S
L	A	A	D	C	J	A	I	R	B	G	T	O
M	S	P	A	P	L	M	N	H	G	S	R	E
N	S	G	T	C	D	E	R	F	V	O	G	T
O	T	P	A	N	H	Y	T	R	F	A	E	P
A	H	M	M	O	P	I	U	Y	R	R	W	V
X	S	F	I	J	J	K	H	G	F	E	E	W
F	B	G	N	M	K	J	I	U	Y	S	Z	X
G	I	L	I	U	Y	H	B	V	K	R	N	B
A	S	D	N	L	K	J	J	H	D	G	J	R
P	O	I	G	T	R	E	W	Q	D	P	U	Y
O	Z	C	V	B	N	M	O	I	U	Y	T	L
L	K	H	G	F	D	S	R	T	Y	U	I	O

Fonte: Elaboração própria, 2004.

A analogia com a prospecção mineral também pode ser utilizada com sucesso para poder explicar o *Data Mining* devido à similaridade com a exigência exaustiva na etapa de preparação para se obter sucesso, caso contrário, o fracasso torna-se um evento certo. Inmon et al. (2001. p.141) afirmam: “A mineração de Dados não é uma panacéia.[...] Frequentemente, os negócios olham para a mineração como uma solução mística para salvar um espaço perdido no mercado”.

Figura 13 – Base de dados depois da Descoberta de Conhecimento – DCBD

J	K	P	W	Q	O	R	S	B	C	X	X	S
L	A	A	D	C	J	A	I	R	B	G	T	O
M	S	P	A	P	L	M	N	H	G	S	R	E
N	S	G	T	C	D	E	R	F	V	O	G	T
O	T	P	A	N	H	Y	T	R	F	A	E	P
A	H	M	M	O	P	I	U	Y	R	R	W	V
X	S	F	I	J	J	K	H	G	F	E	E	W
F	B	G	N	M	K	J	I	U	Y	S	Z	X
G	I	L	I	U	Y	H	B	V	K	R	N	B
A	S	D	N	L	K	J	J	H	D	G	J	R
P	O	I	G	T	R	E	W	Q	D	P	U	Y
O	Z	C	V	B	N	M	O	I	U	Y	T	L
L	K	H	G	F	D	S	R	T	Y	U	I	O

Fonte: Elaboração própria, 2004.

⁴⁵ “Dados! Dados! Dados! Eu não posso fazer tijolos sem barro!” – Sherlock Holmes (DOYLE, 1982, p.286).

5.5.1 O Conceito de *Data Mining*

O conceito de *Data Mining* (ou mineração de dados), segundo Cabena et al. (1998, p.36), é a “[...] técnica de extrair informação, previamente desconhecida e de máxima abrangência a partir de bases de dados, para usá-la na tomada de decisão”.

Han e Kamber (2000, p.8), por sua vez, conceituam a técnica de forma mais detalhada e consonante com os objetivos e objetos desta pesquisa. Para os referidos pesquisadores, *Data Mining* é “[...] uma etapa na descoberta do conhecimento em bancos de dados que consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que normalmente não estão sendo visíveis”. Para isto, são utilizadas técnicas que envolvem métodos matemáticos, algoritmos e heurísticas que visam a descobrir padrões e regularidades entre os dados pesquisados.

Segundo Harrison (1998), o *Data Mining* contempla a exploração e a análise, por meios analíticos ou semi-analíticos, de grandes quantidades de dados para descobrir modelos e regras significativas. É particularmente útil para problemas de modelagem não-linear com grande número de variáveis.

Para Frawley, Piatetsky-Shapiro e Matheus (1992, p.214), a Mineração de Dados – MD também pode ser definida como um método para recuperar informação em dados e, quando empregada no *KDD*, pode focar o conhecimento. Conhecimento, informação e dados seriam, para eles, a mesma coisa, já que são comumente definidos um em função do outro, onde: “dado é a informação bruta” e “sem associatividade”; informação “é o dado trabalhado ou com associatividade dentro de um contexto” e o conhecimento “é a informação associada em múltiplos contextos”. Assim, dados são somente algo armazenado em algum lugar, sendo conhecimento e informação algo mais rico, defendem os autores.

Mesmo com a diferença de definições, pode-se perceber, nos diferentes autores, a percepção de que o objetivo da Mineração de Dados é produzir uma particular quantidade de padrões em um subconjunto.

Diante das definições apresentadas, a seguir são discutidas as diferentes abordagens e como elas estão relacionadas.

5.5.2 Diferentes abordagens

Impulsionado pelo aumento do poder computacional, principalmente de armazenamento e processamento, o *DM* tem sido um assunto amplamente discutido por diversas áreas nos últimos anos. Dados armazenados e nunca antes analisados hoje são vasculhados em seus mínimos detalhes por técnicas de “mineração”. O poder do *DM* está diretamente ligado à capacidade de alcançar informação, antes invisível, armazenada nas Bases de Dados. Nenhum dos outros métodos tradicionais de análise de dados existente é tão parecido com a forma humana de pensar. Não há necessidade de se explicitar o que procurar, ou seja, é um modo de encontrar a informação totalmente desconhecida.

Ao estudar os Sistemas de Informação com uma visão sociotécnica, Inmon et al. (2001, p.141) afirmam que, para se ter uma MD eficaz, é necessária uma estrutura organizacional que valorize a pesquisa e que entenda o valor derivado das informações. Desta forma, se os dados são inconsistentes, normalmente gerados por processos organizacionais falhos, o resultado da MD agregará pouco ou nenhum valor a estes. Uma aplicação imprudente de métodos de *Data Mining* pode ser uma atividade perigosa que conduz à percepção de padrões incorretos ou sem sentido (AGRAWAL, 1996).

Para Inomn, Terdeman e Imhoff (2001), a mineração de dados não é um fenômeno novo⁴⁶, sua origem remonta ao ano de 1965, quando os computadores passaram a ser utilizados para análise de dados. Os autores citam como precursores os pacotes computacionais Biomed (origem nas ciências físicas) e o *Statistical Package for the Social Sciences – SPSS* (origem nas ciências sociais) que continham em suas rotinas originais funções estatísticas, que são o núcleo das atuais ferramentas de análise de alto nível. Para Inmon (1997), a mudança real nas ferramentas está na facilidade de uso.

Berry e Linoff (1997) defendem a idéia de que a MD é apropriada para analisar grandes volumes de dados sob diferentes perspectivas, visando descobrir informações úteis que normalmente não são imediatamente visíveis. Para Niederman (1997), o conhecimento gerado a partir de grandes bases de dados deve ser novo e relevante. Moxon (1996) complementa ao dizer que a MD consegue encontrar padrões que não são encontrados por sistemas ditos tradicionais, mas são capazes de gerar hipóteses, aprender e apoiar a realização

⁴⁶ É importante lembrar que o parâmetro de tempo em Tecnologia da Informação é especialmente diferente das demais áreas. Por exemplo: o computador pessoal foi inventado na década de 80 e a Internet foi difundida, no Brasil, na segunda metade da década de 90.

de descobertas a partir de dados. A mineração de dados, por sua vez, permite aos usuários explorar e inferir informação útil a partir dos dados brutos, descobrindo relacionamentos dispersos nos bancos de dados.

O termo *Data Mining* teve conotações negativas na área da Estatística na década de 60, quando técnicas de análise de dados baseadas em computador foram inicialmente introduzidas. A preocupação surgiu decorrente do fato de que pesquisas minuciosas em qualquer conjunto de dados podem identificar padrões que parecem ser estatisticamente significantes, mas, de fato, não o são (FAYYAD et al., 1996). O *Data Mining*, entretanto, produz resultados eficazes desde que usado corretamente.

O KDD é apresentado, por grande parte dos autores, como sendo um processo hipotético-indutivo, sendo raros os que fazem referência ao processo hipotético-dedutivo. Diante do grande volume de dados, do poder e da facilidade de operação das ferramentas de modelagem, torna-se natural que o usuário sintá-se tentado a encontrar padrões sem se preocupar com uma construção de um modelo de análise baseado em uma reflexão teórica prévia. Conforme se discute a seguir, esta é uma atitude temerária.

Associado ou não a um processo KDD, atualmente o termo *Data Mining* é amplamente utilizado por diversas áreas, ainda assim alguns autores utilizam termos diferentes para definir o reconhecimento de relacionamento entre os dados. Mineração de Dados, como foi visto anteriormente, surge no ambiente de mercado, diante da necessidade de analisar os repositórios de dados.

Corroborando a abordagem utilizada neste trabalho, profissionais de Inteligência Artificial tratam o *Data Mining* como apenas uma das etapas da Descoberta de Conhecimento em Base de Dados. Por outro lado, nos trabalhos da área de Estatística e Tecnologia da Informação, DCBD confunde-se com *Data Mining*, como é o caso de Inmon, Terneman e Imhoff (2001), entre outros. Entretanto, fica clara a preocupação das duas correntes em ampliar o escopo do processo visando alertar sobre as conseqüências de um uso desconexo dos conceitos e ferramentas.

5.5.3 Objetivos da Mineração de Dados

Segundo Fayaad (1996) e Menezes et al. (1998), os objetivos da mineração de dados podem ser agrupados em predição e descrição. A predição busca solucionar o problema de inferências de valores futuros ou desconhecidos a partir de valores previamente conhecidos, e a descrição visa descobrir padrões relevantes em todo conjunto de dados minerado. Os padrões descritivos (agrupamento, regras de associação e padrões seqüenciais), por serem mais fáceis de interpretar, são mais importantes que os preditivos (regressão e classificação).

Já Campos (2000) enumera três objetivos usuais da mineração de dados, que são os seguintes:

- explanatório: quando se deseja explicar algum evento ou medida observada;
- confirmatório: ao se objetivar confirmar uma hipótese;
- exploratório: ao analisar os dados buscando relacionamentos novos e não previstos.

Em síntese, as operações típicas do processo de *Data Mining* envolvem grandes volumes de dados que, inicialmente, devem ser classificados em alguma classe de um conjunto predefinido. Em cada uma destas classes, é necessária a **limpeza dos dados** através da remoção daqueles que representam “ruídos”. Para tanto, deve-se efetuar a **avaliação estatística** destes dados antes e após a “limpeza” da base de dados. A etapa de **agrupamento** (identificação de um conjunto finito de grupamentos para descrever os dados) e a busca das **regras de associação** (modelagem de dependências) freqüentemente requerem o **reconhecimento de padrões e relacionamentos** complexos entre grupos de dados e, em consequência, a interação com usuários e conhecimentos prévios ao processos em análise.

Para Diniz e Louzada Neto (2000, p.28), em geral, os algoritmos de mineração de dados consistem em uma mistura de três componentes: a) o modelo; b) o critério de preferência e c) o algoritmo de busca.

- a) O **modelo** pode ser dividido em: função do modelo e representação do modelo, em que as funções são basicamente: classificação e regressão; dependência; associação; análise de seqüência; *cluster* e sumarização. São abordadas em detalhe, a seguir, no item 5.5.4 As representações dos modelos mais comuns são os lineares, os não-lineares e as árvores de decisão. A interpretação do

modelo está intimamente ligada à complexidade que este possui para explicar melhor os dados, porém são mais difíceis de interpretar e explicar.

- b) O **critério de preferência** indica se um modelo e seus parâmetros atendem aos critérios do DCBD, sendo também conhecido, na Estatística, como seleção de modelos. Os critérios podem ser classificados como implícitos e explícitos. Os implícitos refletem o vício subjetivo do analista em relação aos modelos que são inicialmente considerados. Os explícitos, quantitativos, ficam embutidos nos algoritmos de busca, assim como o critério da máxima verossimilhança para estimar parâmetros.
- c) No **algoritmo de busca**, um problema freqüente de otimização é encontrar os parâmetros mais adequados, ou seja, encontrar o máximo global de uma função não-linear no espaço paramétrico. Estes algoritmos visam o ajuste do modelo em um espaço de modelos e a estimação de parâmetros, dado um modelo selecionado.

Como anteriormente já feito neste capítulo, usar-se, aqui, mais uma analogia. Antes de iniciá-la, porém, deve-se observar que o próprio nome “Mineração de dados” originou-se da analogia deste processo com aquele utilizado por geólogos na exploração das riquezas minerais.

Assim como na Exploração Mineral⁴⁷, a etapa de Mineração de Dados depende da escolha do método mais apropriado. Vários métodos diferentes de prospecção podem ser utilizados para atingir o mesmo objetivo. Para o geólogo pesquisador Maranhão (1983, p.244): “Não existe *a priori* nenhum meio de definir o melhor método⁴⁸, e por isso, às vezes vários processos são utilizados numa mesma área, para que eles possam ser checados mutuamente”. Situação semelhante existe com o *Data Mining*, como podem confirmar os pesquisadores de Tecnologia da Informação Inmon et al. (2001, p.145): “A chave para o sucesso na mineração de dados é escolher uma técnica que se encaixe ao problema”.

⁴⁷ Explorar é um termo utilizado na Economia e na Mineração para definir o processo de tirar proveito econômico de determinada área, sobretudo quanto aos recursos naturais.

⁴⁸ O autor faz referência à escolha dos métodos de Pesquisa Mineral.

5.5.4 Funções da Mineração de Dados

Independente da metodologia de mineração escolhida, basicamente, todas se baseiam nas mesmas técnicas. Embora de natureza distinta, podem ser empregadas em um mesmo fenômeno com o objetivo de extrair conhecimento relevante.

Atualmente, as técnicas empregadas em *DM* podem ser divididas em dois grandes grupos: heurísticas e matemáticas. Entre as heurísticas, as Redes Neurais Artificiais (RNA) mais se destacam, seguidas da Inteligência Artificial Simbolista. Dentro do grupo da matemática, destacam-se a Análise Estatística e a Modelagem Matemática.

Essas técnicas são empregadas nas funções do *Data Mining* e, entre as mais utilizadas, estão classificação, regressão, particionamento, sumarização, dependência do modelo, análise de vínculo e seqüência de análise (FAYAAD et al., 1996).

Para Turban, Rainer e Potter (2001), o *DM* identifica fatos e sugere conclusões com base no exame dos dados para detectar anomalias ou padrões. Pode ser apresentado em cinco funções: classificação; agrupamento; associação; seqüenciamento e previsão.

A componente modelo, apresentada pelos principais autores anteriormente no item Etapas do processo *KDD*, como pode ser observado, difere basicamente na nomenclatura. Por este motivo, optou-se por um reagrupamento das funções consideradas mais apropriadas para este estudo. O principal objetivo deste procedimento foi padronizar os termos que serão utilizados no decorrer deste trabalho, principalmente no capítulo que trata das estratégias metodológicas. Além disso, a uniformização permitiu focar a abordagem nas técnicas empregadas neste trabalho durante o processo de descoberta de conhecimento em bases de dados públicas.

Basicamente, os algoritmos de *DM* são comumente associados às funções de a) classificação; b) regressão; c) associação e modelos de dependência e análise de seqüência; d) *clustering* e e) sumarização. Estas funções são descritas a seguir.

a) Classificação

Para Carvalho (2001), a classificação é uma das funções mais utilizadas no *DM*, simplesmente porque é uma das tarefas cognitivas humanas mais realizadas no auxílio à compreensão do ambiente em que vivemos. A classificação pressupõe características que

definem grupos específicos e associa ou classifica um item em uma ou várias classes predefinidas. (FAYYAD et al., 1996).

Com base em um conjunto inicial de registros, é gerado um conjunto de regras ou uma árvore de decisão, resultando em um modelo composto de padrões que será utilizado para diferenciar as classes e classificar automaticamente os demais registros.

Técnicas como Redes Neurais Artificiais, Estatística e Algoritmos Genéticos dedicam-se ao estudo de classificação de dados. A função classificação utiliza, basicamente, métodos como Árvore de Decisão, Regra de Decisão e Análise Discriminante. A regressão logística será apresentada como função Regressão.

A Árvore de Decisão pode ser considerada um aperfeiçoamento das técnicas utilizadas no Aprendizado de Máquinas (*machine learning*). Tem seu funcionamento baseado em testes automáticos que servem para selecionar os itens que atendem a uma classificação previamente estabelecida, além de auxiliar na identificação de novas classificações. A representação do método tem forma de árvore, em que as ramificações representam as questões que determinaram a predição da classe. Os principais algoritmos utilizados são: CART – *Classification and Regression Tree*; Índice de Gini; C4.5 que é uma extensão do conhecido ID3 e suas evoluções (ID4 e ID6); Regra de classificação Bayesiana; MPI – *Mean Posterior Improvement*, entre outros.

Os algoritmos clássicos empregados na função classificação baseiam-se em árvores de decisão, e entre os mais empregados estão: o *Iterative Dichotomizer 3rd* – ID3 (QUINLAN, 1983; NEVILLE, 1999); o *Automatic Interaction Detection* (MORGAN; SONQUIST, 1963; NEVILLE, 1999) e o *Chi Square Automatic Interaction Detection* (KASS, 1975).

A modelagem através do método Regras de Decisão utiliza a família de algoritmos CN2 e AQ, fortemente influenciados pelos métodos da Engenharia Elétrica. Já o sistema de geração de regras RAMP utiliza a metodologia R-Mini para gerar regras “minimais” completas e consistentes.

Entre os selecionados, o último deles é um método derivado da técnica estatística, denominado de Análise Discriminante, recomendada para identificar as variáveis (explicativas) que melhor discriminam grupos previamente identificados (variáveis explicadas). Utiliza o algoritmo da Função Discriminante de Fischer para dois grupos ou vários grupos.

b) Regressão

A função regressão é similar à função classificação, diferencia-se desta apenas porque a regressão procura predizer um valor real ao invés de um atributo nominal ou uma categoria.

Os primeiros ensaios sobre o método Regressão⁴⁹ datam do início do século XIX, originalmente chamado de reversão⁵⁰, sendo um dos métodos estatísticos mais antigos. É usado para estimações de parâmetros e ajustar dados. Busca associar um item a uma ou mais variáveis de predição de valores reais, faz estimativas de valores futuros, com base nos padrões identificados no conjunto de dados.

Com a popularização do computador, os cientistas sociais passaram a utilizar as técnicas de regressão (INMON et al., 2001). Atualmente, as ferramentas de análise de regressão são encontradas nos diversos níveis de plataformas de computação, até mesmo no popular MS-Excel. Em sua forma mais simplificada, a função Regressão aparece como o Método Estatístico de Regressão Linear Simples: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

Existem outros modelos de regressão mais complexos, envolvendo maior número de variáveis explicativas e relacionamentos não-lineares, e, entre eles, destacam-se a Regressão Múltipla – RLM, Probit e a Regressão Não-linear.

A Regressão Múltipla é o método de análise mais apropriado quando o problema de pesquisa envolve mais de duas variáveis. Outra técnica largamente empregada para este objetivo são as Redes Neurais Artificiais – RNA.

Na análise de regressão clássica, há uma única variável dependente e múltiplas variáveis de predição (independentes). Quando se procura mensurar a probabilidade de ocorrência dos resultados entre uma variável resposta (explicada) do tipo dicotômica⁵¹ e as variáveis explicativas são categóricas ou contínuas, é utilizada a Regressão Logística ou modelo logístico.

No jargão estatístico, os modelos de predição com classificação e com regressão são chamados, respectivamente, Árvore de Classificação e Árvore de Regressão. Para as

⁴⁹ O método estatístico regressão pode ser utilizado na função Regressão do *Data Mining*.

⁵⁰ Francis Galton descobriu o fenômeno estatístico de regressão e Charles Darwin usou reversão em um contexto biológico no trabalho *A Origem das Espécies*.

⁵¹ São variáveis qualitativas que são representadas apenas por duas categorias.

Árvores de Regressão, devem ser considerados os aspectos relacionados ao ajuste do modelo e sua verificação, bem como a seleção de variáveis explicativas que farão parte do modelo.

c) Associação, Modelos de Dependência e Análise de Seqüência

A função associação identifica relações significativas existentes entre os eventos ocorridos em determinada ocasião (relações entre campos de um banco de dados). Os modelos de dependência procuram descrever dependências significantes entre variáveis (AGRAWAL, 1996). Estes últimos se apresentam em dois níveis: o estrutural e o quantitativo. Nos modelos de dependência estruturais, são especificadas as variáveis localmente dependentes umas das outras, enquanto, nos modelos de dependência quantitativos, são utilizadas escalas numéricas para determinar as forças dessas dependências entre as variáveis.

Cabena et al. (1998) descrevem a função associação como o processo de interconexão de objetos, na tentativa de expor características e tendências. Tipicamente, procuram determinar quais fatos ocorrem simultaneamente com probabilidade razoável ou que itens de uma base de dados estão presentes com uma certa *chance*. Esta função pode ser usada para entender e explorar padrões de comportamento dos dados. Os principais métodos são Regras de Associação e Característica Seqüencial.

Uma Regra de Associação possui como grande vantagem sua simplicidade. Expõe, explicitamente, quais itens implicam a presença de outros itens. Assim “se A, então B”, onde A é chamado de corpo da regra e B de cabeça da regra. Cabe ao analista decidir sobre a validade e a relevância da regra com base em dois parâmetros: o fator de suporte que indica a taxa de ocorrência da regra em um conjunto de dados de transações, e o fator de confiança que é utilizado para selecionar as regras mais fortes.

Diferentemente da técnica anterior, a Característica Seqüencial exige um grande número de registros para assegurar a representatividade dos resultados. E, da mesma forma, procura determinar a freqüência de combinação de cada transação que pode ser produzida nas seqüências de registros.

Por fim, a análise de seqüência procura identificar desvios e tendências no tempo. Tem comportamento semelhante à associação, diferenciando-se apenas pelo fato de que a relação existe durante um dado período de tempo.

d) *Cluster* ou Agrupamento

Diferentemente da função classificação, em que os grupos são predefinidos, os *clusters* são definidos através de agrupamentos dos dados baseados em medidas de semelhança ou modelos de densidade de probabilidade, isto é, para a função *cluster*, não é necessário qualquer informação *a priori* sobre a composição do grupo ou *cluster* para qualquer de seus objetos. Os grupos são sugeridos pelos dados, e não definidos *a priori*. A fase de *cluster* ou agrupamento é também chamada de análise de classificação, taxonomia numérica, ou análise Q (MALHOTRA, 2001).

A função *cluster* freqüentemente está presente nas primeiras fases da mineração de dados, com o intuito de reunir os registros em grupos com características em comum para serem utilizados nas fases seguintes. Procura identificar, baseada em modelos probabilísticos ou em medidas de similaridade, grupos (*clusters*) que compartilham de uma característica específica. Para Quintella e Soares Jr. (2003), o *cluster* procura identificar um conjunto finito de grupamentos para descrever os dados.

O objetivo principal da função agrupamento é classificar, com base em um conjunto de variáveis considerado, os indivíduos pertencentes a uma população em subconjuntos (*clusters*) relativamente homogêneos. Este método só pode ser utilizado, obviamente, quando os indivíduos forem agrupáveis. Os grupos são formados por indivíduos semelhantes e esta semelhança pode ser avaliada através de medidas de similaridade ou dissimilaridade, também chamadas de distâncias.

Os principais algoritmos utilizados na modelagem tipo *cluster*, bem como em outros modelos citados anteriormente neste capítulo, já são antigos na Estatística e só foram disseminados devido à complexidade dos cálculos, após a difusão dos computadores. Os resultados do agrupamento podem ser utilizados para produzir um sumário da base de dados, bem como fornecer subpopulações (grupos) para aplicação posterior de outras técnicas específicas.

Entre os vários algoritmos usados na função *cluster*, merecem destaque aqueles baseados na teoria de conjuntos nebulosos, particularmente apropriados para este fim: o *fuzzy c-means* (BEZDEK, 1981; KLIR; YUAN, 1995), o *extended fuzzy c-means* (KAYMAK; SETNES, 2000; SETNES; KAYMAK, 2001) e o algoritmo de agrupamento participativo (SILVA, 2003).

e) Sumarização

A função sumarização é responsável pela organização e resumo dos dados, permitindo a visualização de forma a perceber as informações subjacentes. É utilizada em uma fase preliminar aos demais modelos ou funções. Visa, principalmente, orientar e motivar análises posteriores mais complexas. Pode ser relacionada à técnica da Estatística, mais especificamente, à Análise Exploratória de Dados ou Estatística Descritiva.

Principalmente no *DM*, a sumarização utiliza funções complementares de Caracterização e Visualização para observar a presença de alguma característica estrutural ou de interesse nos dados. A Visualização, como foi visto, é um poderoso recurso de análise de dados, sendo muitas vezes suficiente para obter as respostas necessárias. Já a Caracterização permite a generalização de qualidades relevantes dos dados através de análises quantitativas que propiciam descrições compactas.

O KDD potencializa a análise crítica, através da ‘Exploração de Dados’, dos principais indicadores de desenvolvimento e suas relações, disponíveis na literatura, por meio de suas funções, técnicas e algoritmos aplicados a grandes bases de dados públicas.

O próximo capítulo apresenta uma breve discussão sobre o método de pesquisa adotado na aplicação do *Knowledge Discovery in Databases* e tem o objetivo de fornecer o detalhamento dos passos e caminhos que foram percorridos para se chegar a modelagem do desenvolvimento através do KDD. O capítulo seguinte visa, também, apresentar todas as especificações técnicas materiais empregadas; delimitar a população de estudo e instrumentos de pesquisa utilizados; como os dados foram tratados, minerados e, por fim, os procedimentos empregados na elaboração do modelo hipotético do Desenvolvimento Social que será, no Capítulo 7, confrontado com os dados disponíveis na jazida de dados, resultando na identificação do modelo empírico validado comum às ferramentas brasileiras de mensuração do desenvolvimento social.

6 O PROTOCOLO KNOWLEDGE DISCOVERY IN DATABASE E OS MODELOS HIPOTÉTICOS

É a teoria que decide o que podemos observar. (Albert Einstein)

A busca pela mensuração do desenvolvimento social na literatura não é recente, pelo contrário, há muitos registros de diversas tentativas, desde a utilização de um único indicador (conceito operacional), como no caso da “renda”, até modelos sofisticados que recorrem a uma grande gama de indicadores, como é apresentado no Capítulo 4 desta tese. Ainda assim, a mensuração do Desenvolvimento Social é um tema bastante controverso, que envolve um amplo leque de possibilidades na tentativa de aproximação do fenômeno estudado.

Este capítulo tem o objetivo de descrever os passos do KDD empregados na elaboração do modelo hipotético do Desenvolvimento Social que é testado no Capítulo 7.

6.1 SÍNTESE DOS PROCEDIMENTOS

Para melhor compreensão da estratégia metodológica do *Knowledge Discovery in Database (KDD)*, buscou-se fornecer neste item a descrição dos passos e caminhos que foram adotados na aferição dos modelos hipotético e empírico do Desenvolvimento Social.

A metodologia aqui adotada variou em suas diferentes etapas. A estratégia de pesquisa, no entanto, seguiu aproximadamente duas etapas do KDD: Prospecção e Mineração, ambas resultado de interpretação e síntese da literatura revisada nos capítulos anteriores. Atendendo a um dos objetivos específicos estabelecidos, concluída a revisão teórica e a identificação⁵² das relações (diagrama de caminho) entre os indicadores utilizados nas ferramentas de mensuração do desenvolvimento mais reconhecidas nacional e internacionalmente, é apresentada, a seguir, a proposta de etapas para o KDD utilizadas em bases de dados públicas para confrontar⁵³ o modelo hipotético do desenvolvimento social

⁵² Terceiro objetivo específico desta pesquisa, conforme apresentado no Capítulo 1

⁵³ Quinto objetivo específico desta pesquisa, conforme apresentado no Capítulo 1.

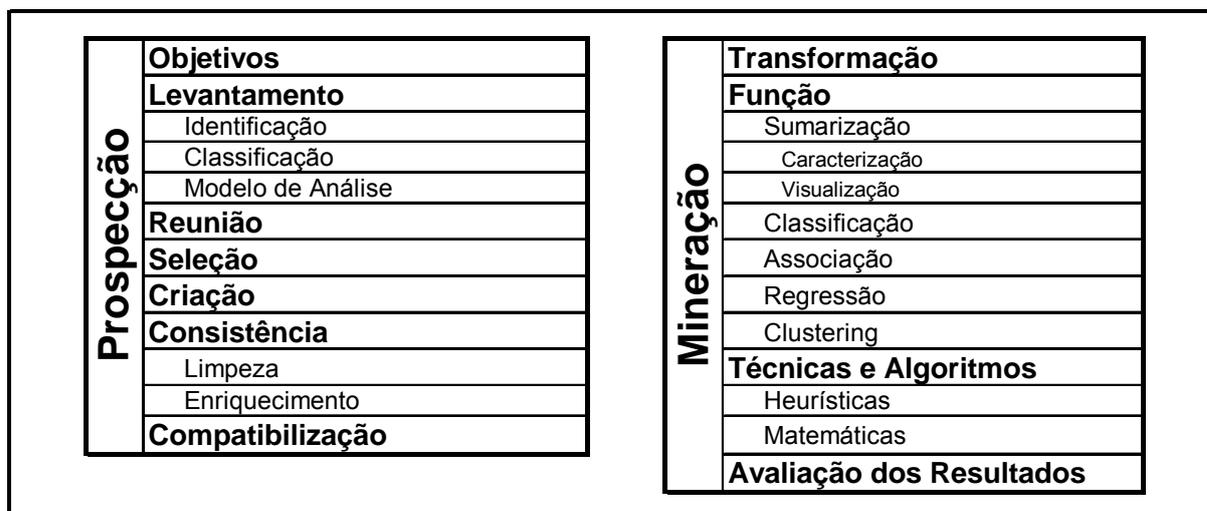
brasileiro sobre os dados empíricos referentes aos indicadores dos 5.561 municípios brasileiros.

A Figura 14 descreve as duas fases principais, ‘Prospecção’ e ‘Mineração’ de dados, empregadas durante a aplicação do KDD nesta tese.

Conforme pode ser visto na figura, as etapas identificadas para a fase de ‘Prospecção’ (I) foram respectivamente: i) Identificação de ‘Objetivos’; ii) ‘Levantamento’ (identificação e classificação das fontes existentes e definição do ‘Modelo’ de análise); iii) ‘Reunião’; iv) ‘Seleção’ e v) ‘Criação’ das bases de dados; vi) ‘Consistência’ (limpeza ou eliminação de ruído e enriquecimento) das bases de dados e vii) ‘Compatibilização’ das bases de dados.

A fase de ‘Mineração’ (II) compreende as etapas: i) ‘Transformação’ dos dados; ii) ‘Função’ iii) ‘Técnicas e Algoritmos’ e iv) ‘Avaliação’ dos resultados.

Figura 14 – Fases e etapas em um processo de *Knowledge Discovery in Database – KDD*.



i) Etapas de Prospecção:

- 1 – Identificar e analisar as bases de dados públicas disponíveis no País e adequadas ao recorte proposto nesta pesquisa.

- 2 – Estruturar uma metabase de dados⁵⁴, com o objetivo de permitir o cruzamento das informações existentes nas diferentes bases sobre seus registros e variáveis;
- 3 – Especificar um modelo hipotético entre os indicadores de desenvolvimento social.

ii) Etapa de Mineração:

- 4 – A partir da metabase de dados (“Jazida de Dados”), foram utilizadas funções de mineração de dados com o objetivo de testar modelo hipotético formulado na etapa de prospecção;
- 5 – Após o teste do modelo hipotético reformulado na etapa anterior, procedeu-se a tarefa de determinar qual dos modelos hipotéticos alternativos sobre as relações de influência para o fenômeno tem maior aderência aos dados empíricos selecionados para o Brasil e, portanto representa o modelo empírico comum aos índices e sistemas de indicadores de desenvolvimento social.

6.2 CRITÉRIOS PARA ADOÇÃO DO KDD

Entre outros importantes autores, Fayyad et al. (1996) definem alguns critérios de viabilidade de ordem prática e técnica, para execução de um processo KDD. Para avaliar o enfoque administrativo, são examinados quatro critérios práticos: a) Impacto Potencial; b) Existência de Processos Alternativos; c) Suporte Organizacional e d) Existência de Problemas Legais. Em relação à viabilidade do processo, são observados outros três critérios de natureza técnica: e) Quantidade de Dados e Atributos Relevantes; f) Qualidade dos Dados e g) Conhecimento acerca do Domínio.

Ao analisar os quatro primeiros critérios de ordem prática podem ser feitas as seguintes observações:

⁵⁴ A base de dados estruturada para o *KDD*, a partir das bases de dados originais, foi chamada nesta tese de “Jazida de Dados”. Este conceito é explorado no capítulo 5 desta tese.

- a) Impacto Potencial da Aplicação: a base de dados utilizada nesta tese (IBGE, Ministério da Justiça e Ministério da Saúde, 2000) revela o retrato do País no fim do milênio e pode ser considerada a principal fonte⁵⁵ de informações para a análise sobre a realidade nacional, servindo para fundamentar o exercício da cidadania, os investimentos públicos e privados, entre outros aspectos. Com isso, o resultado da pesquisa poderá ser utilizado, entre outras coisas, para a formulação de políticas públicas. A princípio, poderá também servir de referência nos estudos das instituições de planejamento, pesquisa e estatística, das Secretarias de Planejamento dos governos estaduais, bem como às Secretarias de Planejamento dos Municípios Brasileiros. A proposta apresentada nesta tese poderá também servir de referência para aplicações nos demais Índices e Sistemas de Indicadores brasileiros, além de servir de subsídio para diversas outras pesquisas que discutam o tema.
- b) Existência de Métodos Alternativos: embora não fosse o KDD a única alternativa, no momento da execução desta pesquisa não foi encontrado outro mecanismo economicamente viável para aferição do modelo hipotético do Desenvolvimento Social no Brasil a partir de grandes massas de dados. O volume de dados tratado foi significativo e a análise realizada permitiu o melhor nível de abrangência e detalhe possível, através da utilização de bases de dados pública.
- c) Suporte Organizacional: o desenvolvimento desta pesquisa contou com um financiamento de quatro anos (2005 a 2008) da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Parte de seu desenvolvimento ocorreu no âmbito das atividades profissionais do próprio pesquisador na Superintendência de Estudos Econômicos e Sociais da Bahia – SEI, o que possibilitou o acesso a equipamentos, material e corpo técnico com elevado nível de especialidade do domínio estudado. Além disso, esta tese é produto do resultado das atividades acadêmicas realizadas no grupo de pesquisa sobre Sistemas de Apoio à Decisão Estratégica - Uma Comparação de Características, Potenciais e Resultados (2003-2005), Descoberta de

⁵⁵ Censo Demográfico (IBGE), Sistema Nacional de Estatísticas de Segurança Pública e Justiça Criminal (SINESPJC/Ministério da Justiça) e Banco de dados do Sistema Único de Saúde (DATASUS/Ministério da Saúde).

Conhecimento em Bancos de Dados Públicos (2005-2008), Modelagem e Construção de Um Sistema de Apoio à Decisão (2006-2010) e Geração de Indicadores de Desenvolvimento Social por Meio da Descoberta de Conhecimento em Bancos de Dados (2007-2010), pertencentes ao Núcleo de Política e Administração de Ciência e Tecnologia (NACIT) da Universidade Federal da Bahia – UFBA, coordenado pelo Professor Rogério Quintella.

- d) Problemas Legais: embora as questões éticas e os problemas legais sejam de grande importância no ambiente organizacional de natureza privada, na abordagem utilizada nesta tese, KDD em Bases de Dados Públicas, este critério não é dos mais relevantes. Esta afirmação justifica-se, principalmente, pela gratuidade e pelo caráter público de grande parte das Bases de Dados disponíveis. Observa-se, também, que o IBGE, o Ministério da Justiça (MJ) e o Ministério da Saúde (MS) só disponibilizam dados que obedecem ao princípio do “sigilo estatístico”, o que garante a confidencialidade da informação primária. Os dados dos cidadãos são disponibilizados apenas de maneira agregada, impossibilitando a individualização do informante.

Avaliados e atendidos os quatro critérios de ordem prática, foram avaliados outros três critérios de ordem técnica de viabilidade do processo, são eles:

- e) Quantidade de Dados e Atributos Relevantes: a população da pesquisa foi de 168.370.893 habitantes⁵⁶ distribuídos nos 44.795.101 domicílios particulares permanentes existentes nos 5.561 municípios brasileiros. Diante do propósito de apresentar apenas o resultado empírico da aplicação do KDD em bases de dados públicas e do nível de agregação de necessário para obtenção dos indicadores selecionados, os microdados correspondentes aos questionários das pesquisas foram agrupados por nível territorial municipal.
- f) Qualidade dos Dados: a principal vantagem em optar por trabalhar com dados públicos disponibilizados por instituto oficial de estatística é à consistência e confiabilidade, além do baixo custo destes dados. Este fato contribuiu sobremaneira para a execução das fases iniciais da prospecção no KDD. É

⁵⁶ Os dados demográficos apresentados nesta tese têm como referência o Censo 2000. Segundo o IBGE (2000), a população brasileira em 2000 era de 169.799.170 habitantes.

notório que os dados oficiais do IBGE, entre as bases de dados públicas disponíveis, demonstram um nível de qualidade elevadíssimo. O Ministério da Saúde, por meio do DATASUS e o Ministério da Justiça, por meio do Sistema Nacional de Estatísticas de Segurança Pública e Justiça Criminal, disponibilizam dados com qualidade suficiente para o propósito desta pesquisa.

- g) Conhecimento acerca do domínio: foram empreendidos estudos específicos para aprofundamento do tema Desenvolvimento, conforme apresentado no Capítulo 2 desta tese. Além disso, fatores como o conhecimento adquirido ao longo da realização de trabalhos de natureza técnica sobre a realidade social do Brasil, de algumas unidades da federação e de alguns municípios brasileiros por parte do autor desta tese, a participação no desenvolvimento prévio de pesquisas sobre Sistemas de Indicadores e Índices de Desenvolvimento nas Principais Cidades do Brasil (Soares Jr et al., 2005, 2006, 2007, 2008, 2009), juntamente com o apoio de especialistas do domínio (Desenvolvimento), representados por pesquisadores de Universidades⁵⁷, por Instituições de Planejamento, Pesquisa e Estatística⁵⁸ e, Ministérios⁵⁹, contribuíram sobremaneira para a consecução dos trabalhos.

Após a exposição de todos os requisitos propostos na literatura acerca da utilização do *KDD*, considerou-se o processo adequado para investigar o Desenvolvimento Social no Brasil. Desta forma, efetuou-se sua execução conforme descrito a seguir.

⁵⁷ Universidades de Campinas (UNICAMP), Universidade Federal da Bahia (UFBA), Escola Nacional de Ciências Estatísticas (ENCE), Universidade Federal de Pernambuco (UFPE), Universidade de Brasília (UNB), Universidade Federal do Ceará (UFC)

⁵⁸ Fundação João Pinheiro (FJP-MG), Fundação CIDE (RJ), Fundação Sistema Estadual de Análise de Dados (SEADE-SP), Instituto Municipal de Urbanismo Pereira Passos (IPP-RJ), Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE-CE), Fundação de Economia e Estatística (FEE-RS), Instituto Brasileiro de Geografia e Estatística (IBGE-RJ), Instituto de Pesquisa Econômica Aplicada (IPEA-DF), Superintendência de Estudos Econômicos e Sociais da Bahia (SEI-BA) e Programa das Nações Unidas para o Desenvolvimento (PNUD, DF)

⁵⁹ Ministério da Justiça, Ministério da Saúde e Ministério das Cidades.

6.3 ETAPAS DO KDD EMPREGADAS

Como foi visto no Capítulo 5 desta tese, a bibliografia sugere diversas abordagens para o *KDD*, algumas delas com um encadeamento linear e sucessivo das fases, fato este que, nesta pesquisa, não foi verificado. As fases aconteceram muitas vezes de maneira simultânea, ou, em alguns casos, fora da “ordem” proposta pelos principais autores. Durante a consecução de cada etapa, muitas vezes foi necessário retornar à etapa anterior, pois o conhecimento extraído na etapa corrente conduzia a uma melhor definição da etapa anterior. Dessa forma, para este estudo, optou-se por uma abordagem híbrida gerada com base nas semelhanças entre as propostas estudadas.

Assim, o processo ocorreu de forma iterativa, pois cada etapa possuía interseções com as demais. Ficou claro que se trata de um processo interativo, em que, a cada etapa, as informações geradas são analisadas e enriquecidas com outras informações para obter-se resultados melhores.

Visando contribuir para o entendimento do protocolo de métodos *KDD* voltado para as Bases de Dados Públicas, buscou-se descrever⁶⁰ neste tópico as duas fases principais, Prospecção e Mineração de Dados, empregadas durante a realização desta pesquisa.

I – Prospecção

As etapas identificadas para a fase de prospecção foram, respectivamente:

- 1) **Identificação de Objetivos** – definição dos objetivos do estudo e delimitação das perguntas a serem respondidas;
- 2) **Levantamento** – seleção e classificação das fontes existentes; redefinição do modelo de hipotético, Construtos que foram avaliadas, Conceitos e Indicadores utilizados;
- 3) **Reunião, seleção e criação** das bases de dados;
- 4) **Consistência** das bases de dados;
- 5) **Compatibilização** das bases de dados.

⁶⁰ Alguns itens apresentados já foram respondidos anteriormente e por este motivo serão feitas apenas referências sem necessitar de um maior detalhamento.

II – Mineração

- 1) **Transformação dos Dados;**
- 2) **Função e Algoritmos de Mineração;**
- 3) **Avaliação dos resultados;**
- 4) **Disseminação dos resultados.**

Comungando o pensamento de Reinartz (1999), foi feita a documentação de todo o processo, porém, com o objetivo de tornar a leitura mais acessível, optou-se por apresentar, nesta tese, apenas uma síntese dos principais procedimentos adotados em cada etapa.

6.3.1 Prospecção

Em um processo de *KDD*, esta etapa destina-se a delimitação das perguntas, definição dos objetivos, organização da equipe de trabalho e planejamento de todas as atividades a serem executadas.

6.3.1.1 Objetivos

No contexto desta tese, o emprego do *KDD* teve o objetivo de atender os objetivos específicos estabelecidos no item 6.1.1.

Diante do desafio de representar o desenvolvimento social utilizando exclusivamente bases de dados públicas, optou-se em fazer uma abordagem voltada para viabilização da aplicação do *KDD* na tarefa de modelar o fenômeno social para o Brasil. Em outras palavras, este capítulo descreve as fases de um processo de *KDD*, método mais adequado do que os métodos quantitativos tradicionais, para lidar com bases de dados de grande volume e com a abordagem multidimensional do Desenvolvimento, uma vez que é capaz de refletir, na esfera das investigações empíricas, a complexidade do fenômeno, em afinidade com as formulações teóricas mais recentes. O Objetivo principal do *KDD* foi

interpretado da seguinte maneira: confrontar o modelo hipotético e os dados empíricos reunidos na jazida de dados, resultando em uma proposta integrada para a operacionalização do conceito de KDD no estudo dos Índices e Sistemas de Indicadores de Desenvolvimento Social no Brasil.

Como foi visto no Capítulo 5, a equipe de um processo *KDD* deve ser composta de três entidades: o usuário, o especialista no domínio (desenvolvimento social) e o analista de dados. Para esta tese, todas as três entidades foram representadas pelo próprio autor, sendo que a entidade especialista do domínio recebeu valiosas contribuições externas⁶¹. Embora a literatura revisada não faça muitas referências sobre a necessidade do analista de dados aprofundar seus conhecimentos no domínio (desenvolvimento social) a ser estudado, no caso desta pesquisa, mostrou-se crucial este aprofundamento, ou seja, no *KDD* quanto mais o analista de dados conhecer sobre o domínio estudado (e vice versa), mais conhecimento poderá ser agregado a cada interação com as bases de dados e melhores serão os resultados.

A seguir, são apresentadas as demais fases propostas para o *KDD* em bases de dados públicas, denominadas nesta tese de Prospecção e Mineração de Dados.

6.3.1.2 Levantamento

Após a comprovação da viabilidade para execução do *KDD*, foi feito um levantamento das principais fontes de informação públicas visando identificar bases de dados com capacidade para suprir a demanda estabelecida.

a) Modelo Hipotético Proposto

Buscou-se, neste tópico, medir, através de um conjunto relações entre os indicadores selecionados, um fenômeno de caráter multidimensional⁶² – o desenvolvimento social. Conforme apresentado no Capítulo 4, procurou-se, através das limitadas informações disponíveis nos documentos e bibliografia disponíveis, apresentar o modelo hipotético que proporcione uma maior aproximação da realidade do desenvolvimento social medido pelos índices ou sistemas de indicadores nacionais.

⁶¹ Pesquisadores de Universidades, Técnicos e Pesquisadores de Instituições de Planejamento, Pesquisa e Estatística e Técnicos dos Ministérios.

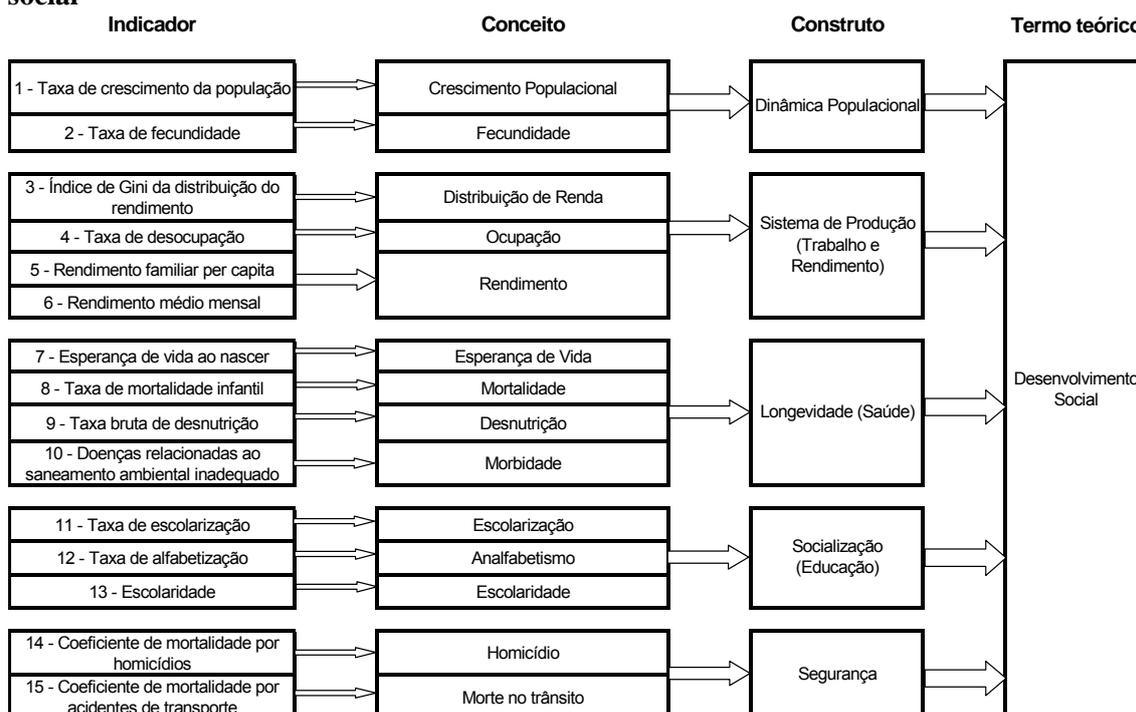
⁶² Amplamente discutido nos Capítulos 2, 3 e 4 desta tese.

De maneira a contribuir com o debate sobre as diversas formas de mensuração do desenvolvimento social, acreditou-se nesta pesquisa, através da aplicação do *KDD* em bases de dados públicas, ser possível uma maior aproximação do fenômeno, elegendo a análise em cinco dimensões focadas nas características de Dinâmica Populacional, Trabalho e Rendimento, Saúde, Educação e Segurança.

Embora condicionada às variáveis disponíveis nas bases de dados utilizadas, a proposta apresentada para mensuração do desenvolvimento social através do uso do *KDD* nesta tese ofereceu a possibilidade de um olhar diferenciado para o fenômeno. A representação aqui proposta irá contribuir com o compartilhamento de idéias e pensamentos, bem como servir como padrão de comparação para futuros estudos, pois a utilização intensiva de técnicas de *Data Mining* para mensuração de fenômenos sociais ainda é pouco explorada na literatura.

A partir dos Conceitos levantados com o estudo das principais propostas disponíveis na literatura foi elaborado (Figura 15) o modelo gráfico hipotético das relações entre os indicadores para esta pesquisa que posteriormente, no Capítulo 7 é feita à aferição do ajuste do modelo hipotético, avaliado através de técnicas específicas de *Data Mining* aplicadas à jazida de dados.

Figura 15 – Modelo gráfico hipotético das relações entre os indicadores de desenvolvimento social



Apresentado o modelo hipotético proposto para mensurar o Desenvolvimento Social, a seguir são aduzidos os indicadores propostos.

b) Indicadores Seleccionados

Para suprir o modelo hipotético para representar o Desenvolvimento Social no Brasil, foram seleccionados, a partir dos trabalhos de mensuração identificados na literatura revisada no Capítulo 4, todos os 15 indicadores de situação⁶³ de bem estar relativos às características das pessoas residentes nos 5.483 municípios⁶⁴ brasileiros.

Em sintonia com o que é preconizado na literatura⁶⁵, os indicadores de bem estar foram seleccionados a partir das variáveis existentes nos bases de dados públicas do Instituto Brasileiro de Geografia e Estatística (IBGE), Ministério da Saúde (MS), Ministério da Educação (ME) e Ministério da Justiça (MJ).

Levantar os indicadores de Desenvolvimento Social para esta pesquisa constituiu tarefa bem difícil. As informações estatísticas necessárias quase sempre não se encontram disponíveis, e aquelas disponíveis são vistas com restrições. Entretanto, conforme apresentado no Capítulo 4, a escolha dos indicadores, para compor o conjunto equilibrado de medidas, considerou alguns critérios pragmáticos defendidos por Tironi et al. (1991), Soares Jr. e Quintella (2002), Jannuzzi (2001), Trzesniak (1998). Entre os principais, destacam-se: relevância, gradação de intensidade, univocidade, padronização, rastreabilidade, estabilidade, representatividade e simplicidade. Além do mais, os indicadores encontram-se em um mesmo nível de desagregação espacial, permitindo comparações intersetoriais de maneira confiável.

Atendidos os critérios desejáveis para os indicadores procedeu-se à sua seleção, com base nas possibilidades oferecidas pela “Jazida de Dados” construída e nas definições apresentadas no decorrer desta tese. Para operacionalização do modelo hipotético proposto foram seleccionados os indicadores correspondentes às seguintes dimensões (construtos) de análise: Dinâmica Populacional, Trabalho e Rendimento, Saúde, Educação e Segurança (Figura 14).

⁶³ Como foi definido no Capítulo 3 desta tese.

⁶⁴ Foram incluídos na jazida de dados apenas os municípios com informação em ao menos um dos 15 indicadores seleccionados. Nesta etapa de Levantamento do KDD, foram excluídos 98 municípios dos 5.581 existentes.

⁶⁵ Entre os mais citados estão: BERQUÓ E CAVENAGHI, 2005; IBGE, 2008; ANVISA, 2007; SZWARCOWALD E CASTILHO, 1995; SEVERINO, 2005; FERREIRA, 2000; CASTEL, 1998; FERREIRA, 2000; GUIMARÃES, 2002; POCHMANN, 2001; FRANÇA et al., 2001; entre outros autores citados no Capítulo 4.

I) Dinâmica Populacional

1. **Taxa de crescimento da População**, calculada a partir da taxa média geométrica de crescimento anual da população utiliza as variáveis referentes à população residente em dois distintos marcos temporais. O Crescimento populacional é um conceito básico na análise da dinâmica populacional, refere-se às mudanças no tempo, do tamanho do número de pessoas que residem no território. É calculado através da expressão: $i = \sqrt[n]{\frac{P(t+n)}{P(t)}} - 1$, na qual P(t+n) e P(t) são as populações correspondentes a duas datas sucessivas (t e t+n), e n é o intervalo de tempo entre essas datas, medido em ano e fração de ano. A taxa *i* é o resultado desta equação, expressa em percentual.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, a partir de informações oriundas do Censo Demográfico.

2. **Taxa de fecundidade total** representa o número médio de filhos nascidos vivos que uma mulher teria ao fim do seu período reprodutivo, estando sujeita a uma determinada lei de fecundidade. Os níveis de fecundidade de um país fazem parte de um elenco de indicadores que orientam os formuladores de políticas públicas, dadas as implicações que taxas abaixo do nível de reposição (número médio de filhos por mulher igual a 2.1) têm na redução do volume da população e no seu envelhecimento. (BERQUÓ E CAVENAGHI, 2005). Segundo o IBGE (2000), as informações utilizadas para obtenção da taxa de fecundidade⁶⁶ total são os filhos tidos nascidos vivos nos 12 meses anteriores à data de referência da pesquisa e o total de mulheres segundo os grupos quinquenais de idade dentro do período fértil (15 a 49 anos). O quociente entre os filhos tidos nascidos vivos nos últimos 12 meses provenientes das mulheres em um determinado grupo de idade e o respectivo efetivo de mulheres neste grupo fornece a intensidade com que estas mulheres tiveram seus filhos. No caso de grupos quinquenais, a taxa de fecundidade total é o produto da amplitude do intervalo (quinquenal) e a soma destas intensidades dentro do período reprodutivo.

⁶⁶ A fonte natural desta informação seria o registro de nascimentos ocorridos durante um determinado ano civil, contudo, em função do não-registro e do registro tardio de nascimentos que ocorrem com maior ou menor intensidade em algumas Unidades da Federação, utiliza-se o procedimento citado anteriormente.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas do Censo Demográfico e da Pesquisa Nacional por Amostra de Domicílios – PNAD.

II) Trabalho e Rendimento

3. O **Índice de Gini da distribuição do rendimento** é um dos indicadores mais utilizados para expressar o grau de concentração na distribuição do rendimento da população e avaliar a distribuição dos rendimentos pela população. O índice de Gini é expresso por um valor que varia de 0 (zero), situação de perfeita igualdade, a 1 (um), situação de desigualdade máxima. Em situações concretas, é muito difícil que o índice atinja estes valores extremos. Um índice em torno de 0,5 já é considerado representativo de fortes desigualdades.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, relativas à população de 10 anos ou mais de idade e seus rendimentos mensais de todas as fontes, oriundas da PNAD e do Censo Demográfico.

4. A **Taxa de desocupação** é o indicador que reflete a incapacidade na geração de emprego e rendimento e pode ser obtida através do cálculo da percentagem das pessoas desocupadas na semana de referência em relação às pessoas economicamente ativas nessa semana. A população economicamente ativa (PEA) é o conjunto de pessoas em idade de trabalhar, de ambos os sexos, que constituem a mão de obra disponível para a produção de bens e serviços. Dito por outras palavras, a PEA compreende as pessoas que trabalham (ocupadas) e as que procuram ativamente um trabalho (desocupadas), incluindo aquelas que o fazem pela primeira vez. Segundo recomendações internacionais, a PEA é considerada como a população que participa na atividade econômica e que tenha 15 anos de idade e mais.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas da Pesquisa Mensal de Emprego – PME e do Censo Demográfico.

5. **Rendimento familiar per capita** é o indicador que corresponde a soma total da renda bruta no mês de todos aqueles que compõem a família, dividida pelo

número de seus integrantes. O cálculo do rendimento mensal familiar per capita, considera a soma dos rendimentos mensais de todas as fontes dos componentes da família, exclusive os das pessoas cuja condição na família é pensionista, empregado doméstico ou parente do empregado doméstico, dividido pelo número de componentes da família. As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas da Pesquisa Nacional por Amostra de Domicílios – PNAD e do Censo Demográfico.

6. **Rendimento médio mensal.** Este indicador expressa a média dos rendimentos mensais da população de 10 anos ou mais de idade, com rendimento na semana de referência. Reflete o nível médio do rendimento da população e revela a maneira como se dá a apropriação das riquezas produzidas por um país. Permite avaliar o grau de equidade existente na apropriação do rendimento de homens e de e mulheres, e entre pessoas brancas e pessoas pardas ou pretas.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas da Pesquisa Nacional por Amostra de Domicílios – PNAD e do Censo Demográfico.

III) Logevidade (Saúde)

7. **Esperança de vida ao nascer** representa o número médio de anos de vida esperados para um recém-nascido, mantido o padrão de mortalidade existente na população residente, em determinado espaço geográfico, no ano considerado. Expressa o número médio de anos que se esperaria que um recém-nascido vivesse, representa uma medida sintética da mortalidade, não estando afetada pelos efeitos da estrutura etária da população, como acontece com a taxa bruta de mortalidade. É calculada a partir de tábuas de vida elaboradas para cada área geográfica, toma-se o número correspondente a uma geração inicial de nascimentos (10) e determina-se o tempo cumulativo vivido por essa mesma geração (T0) até a idade limite. A esperança de vida ao nascer é o quociente da divisão de T0 por 10. As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE e pelo Fundo de População das Nações Unidas – UNFPA.

8. **Taxa de mortalidade infantil** representa o número de óbitos de menores de um ano de idade, por mil nascidos vivos, na população residente em determinado espaço geográfico, no ano considerado. É um dos indicadores mais empregados para medir níveis de saúde e de desenvolvimento social de uma região (ANVISA, 2007), mede o risco de um nascido vivo morrer no seu primeiro ano de vida.

Estima o risco de morte dos nascidos vivos durante o seu primeiro ano de vida, reflete, de maneira geral, as condições de desenvolvimento socioeconômico e infra-estrutura ambiental, bem como o acesso e a qualidade dos recursos disponíveis para atenção à saúde materna e da população infantil e expressa um conjunto de causas de morte cuja composição é diferenciada entre os subgrupos de idade. Costuma-se classificar o valor da taxa como alto (50 por mil ou mais), médio (49) e baixo (menos de 20), parâmetros esses que necessitam revisão periódica, em função de mudanças no perfil epidemiológico. Valores abaixo de 10 por mil são encontrados em vários países, mas deve-se considerar que taxas reduzidas podem estar encobrindo más condições de vida em segmentos sociais específicos.

A fonte natural destas informações seria as estatísticas vitais do Registro Civil, nascimentos e óbitos, contudo, em função do sub-registro destes eventos, foram utilizadas informações provenientes dos censos demográficos e pesquisas por amostragem do IBGE.

9. **Taxa Bruta de Desnutrição:** O indicador **Prevalência de desnutrição total** representa a proporção de crianças menores de 5 anos de idade com manifestações de desnutrição aguda, crônica, ou total, através dos índices peso-para-altura (P/A), altura-para-idade (A/I) e peso-para-idade (P/I), respectivamente. O cálculo se baseia na pesagem e determinação da relação entre a massa corporal e a idade para as crianças menores de 5 anos de idade. Esta relação é conhecida como índice P/I, peso para a idade. Considera-se como desnutrição total o valor de peso corporal abaixo de menos 2 desvios-padrão da mediana do peso para a idade, comparado ao valor de referência. O indicador expressa, em percentual, a proporção de crianças desnutridas em relação ao total de crianças sujeitas à pesagem.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, em convênio com o Instituto Nacional de Alimentação e Nutrição – INAN e em colaboração com o Instituto de Pesquisa Econômica Aplicada – IPEA.

10. **Taxa de incidência de doenças** relacionadas ao saneamento ambiental inadequado (DRSAI) representa as internações por doenças relacionadas ao saneamento ambiental inadequado. A Organização Mundial de Saúde (OMS) classifica como DRSAI as doenças de transmissão feco-oral, transmitidas por inseto vetor ou através do contato com a água, doenças relacionadas com a higiene, geo-helmintos e teníases. O indicador é a razão entre o número de internações hospitalares por DRSAI por 100 mil habitantes.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo Ministério da Saúde, Departamento de Informática do SUS – DATASUS, oriundas dos registros de Autorização de Internação Hospitalar – AIH, e pelo IBGE.

IV) Socialização (Educação)

11. **Taxa de Escolarização:** a taxa de escolarização mostra o acesso, abrangendo desde o pré-escolar até o curso superior. A Taxa de Escolarização representa a proporção da população que freqüentam a escola, por faixa etária adequadas à educação infantil (5 e 6 anos), ao ensino fundamental (7 a 14 anos), ao ensino médio (15 a 17 anos), ao superior incompleto (18 e 19 anos) e ao superior completo (20 a 24 anos). Entretanto, é importante ressaltar que no Brasil a distorção idade-série é elevada.

As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas do Censo Demográfico e da Pesquisa Nacional por Amostra de Domicílios – PNAD.

12. **Taxa de Alfabetização:** representa a proporção da população adulta que é alfabetizada. Para a construção deste indicador, são utilizadas as seguintes variáveis: população adulta alfabetizada, considerando as pessoas de 15 anos ou mais de idade que sabem ler e escrever pelo menos um bilhete simples no idioma que conhecem; e total da população nesta faixa de idade, por sexo e por cor ou raça. As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas do Censo Demográfico e da Pesquisa Nacional por Amostra de Domicílios – PNAD.

13. **Taxa de Escolaridade:** este indicador apresenta a média de anos de estudo da população de 25 anos ou mais de idade. Idealmente, as pessoas com 25 anos

ou mais de essas pessoas deveriam ter no mínimo o ensino médio completo. As informações utilizadas para a elaboração deste indicador foram produzidas pelo IBGE, oriundas do Censo Demográfico e da Pesquisa Nacional por Amostra de Domicílios – PNAD.

V) Segurança

14. O Coeficiente de **mortalidade por homicídios**: representa as mortes por causas violentas. As variáveis utilizadas são: número de óbitos por homicídios e população residente. O indicador é a relação entre a mortalidade por homicídios e a população, expressa em homicídios anuais por 100 000 habitantes. Os registros primários de homicídios utilizados para a elaboração deste indicador foram obtidos junto às Secretarias Estaduais de Saúde, por meio das declarações de óbito, e processados pelo Ministério da Saúde, Secretaria de Vigilância em Saúde, Divisão de Análise em Situação de Saúde, disponíveis no Sistema de Informações sobre Mortalidade – SIM.
15. Coeficiente de **mortalidade por acidentes de transporte** é calculado a partir do número de óbitos por acidentes de transporte e da população. O indicador é a relação entre mortalidade por acidentes de transporte e a população considerada, expressa em óbitos por 100 000 habitantes. Os registros primários de mortes por acidentes de transporte, utilizados para a elaboração deste indicador, foram obtidos junto às Secretarias Estaduais de Saúde, por meio das declarações de óbito, e processados pelo Ministério da Saúde, Secretaria de Vigilância em Saúde, Divisão de Análise em Situação de Saúde, disponíveis no Sistema de Informações sobre Mortalidade (SIM).

As informações utilizadas para a elaboração dos indicadores 14 e 15 foram produzidas Ministério da Saúde e estão disponíveis no Sistema de Informações sobre Mortalidade (SIM). Apenas durante a etapa de Levantamento, mais especificamente na identificação e classificação das fontes existentes para definição do modelo do hipotético foi possível constatar que o Ministério da Justiça (MJ) não disponibiliza⁶⁷ as taxas de homicídio

⁶⁷ O Ministério da Justiça disponibiliza somente dados por estado e grandes municípios (acima de 100 mil habitantes). O levantamento de informações sobre homicídios por município implicaria consultar as Secretarias de Segurança Pública de cada Unidade da Federação.

por município, por esse motivo os dados utilizados para o cálculo dos indicadores foram obtidos do Sistema de Informações sobre Mortalidade (SIM).

Em função dos indicadores aqui apresentados e suas respectivas variáveis manifestas, é possível observar que as bases de dados públicas possuem uma característica particular, relacionada ao sigilo estatístico que não permite a individualização da informação e, por esse motivo, os dados são sempre disponibilizados de forma agregada, inviabilizando cruzamentos e análises mais detalhadas.

Estabelecidos os objetivos, feita a seleção e classificação das fontes, definido o modelo hipotético a ser utilizado com seus respectivos construtos, conceitos e indicadores que os compõem, deu-se início à criação da base de dados efetiva (“Jazida de Dados – JD”) que serviria ao processo *KDD* propriamente dito. Como foi visto no Capítulo 5 desta tese, alguns autores (HAN; KAMBER, 2000; ADRIAANS; ZANTIGE, 1996; AMARAL, 2001) denominam de *KDD* apenas as etapas posteriores à formação da “Jazida de Dados”.

A partir deste ponto, os dados das fontes selecionadas foram trabalhados com o objetivo de estruturar uma base de dados específica para atender à etapa da Mineração de Dados. Em uma abordagem mais voltada para o ambiente de negócios, Inmon et al. (2001, p.21) relacionam algumas características desejáveis para esta base de dados específica (“Jazida de Dados”) que pretende atender à Mineração de Dados propriamente dita. Entre as principais características apontadas por Inmon et al. (2001) e observadas nesta tese, estão a integração, a estrutura flexível, a amplitude robusta, a facilidade de acesso e a documentação.

A integração está relacionada à definição, à estrutura e aos significados consistentes dos dados. A estrutura flexível deve permitir diferentes moldes adaptáveis para cada Função do *Data Mining*: a amplitude robusta está relacionada ao escopo do estudo; a facilidade de acesso diz respeito à disponibilidade para consultar a base de mineração e, por fim, a documentação da estrutura, às fontes e aos dados existentes.

Para Inmon et al. (2001, p.21), além de possuir as características supracitadas, os dados devem ser granulares, pois “[...] não há outra maneira de cumprir todas as exigências, a não ser que os dados estejam no nível fino de granulosidade”. Esta estrutura exclusivamente dedicada à Mineração, o autor define como *Exploration Warehouse*.

As etapas subseqüentes de Reunião, Seleção e Criação e Consistência e Compatibilização destinam-se exclusivamente à estruturação do que aqui foi chamado de “Jazida de Dados – JD”. Na abordagem proposta nesta tese, uma “Jazida de Dados” também

pode ser representada por um *Data Warehouse* ou até por um componente de armazenagem *near-line*⁶⁸.

6.3.1.3 Reunião, Seleção e Criação

O emprego do *KDD* pressupõe que serão trabalhadas base de dados já existentes, provenientes de uma agregação de várias bases de dados (dados secundários) e do resultado de uma ou mais *surveys* (dados primários) ou de ambos.

Antes da etapa de Reunião, normalmente os registros e as variáveis de interesse para o estudo estão dispersos em vários arquivos e em diferentes formatos. No caso desta pesquisa, foram utilizados dados do Instituto Brasileiro de Geografia e Estatística (IBGE), do Ministério da Justiça (MJ) e do Ministério da Saúde (MS).

O IBGE forneceu o acesso à maioria dos dados necessários para atender os requisitos dos construtos I. Dinâmica Populacional; II. Sistema de Produção; III. Longevidade e IV. Socialização. Esses dados foram obtidos do Banco Multidimensional de Estatísticas⁶⁹ (BME) e do Arquivo Agregado de Setores Censitários⁷⁰ foram armazenados em matrizes sob a forma de planilhas, permitindo o agrupamento por unidades da federação, município e setor censitário.

Os dados para suprir a demanda dos Indicadores do Construto V. Segurança, foram obtidos do Sistema Nacional de Estatísticas de Segurança Pública e Justiça Criminal (SINESPJC) do Ministério da Justiça. O SINESPJC reúne as informações de segurança pública e justiça criminal produzidos pelos Estados da Federação. Sua gestão é de responsabilidade da Coordenação Geral de Pesquisa e Análise da Informação e conta com o

⁶⁸ A armazenagem *near-line* complementa a armazenagem em disco, alojando dados raramente acessados na armazenagem óptica ou magnética.

⁶⁹ O BME é uma base de dados formada por microdados originados nas pesquisas estatísticas efetuadas pelo IBGE e pela metainformação associada a sua produção. Os microdados correspondem aos dados existentes nos questionários das pesquisas. Cada questionário vai constituir um ou mais registros de informações, sendo, portanto, a informação mais desagregada possível para a pesquisa estatística. Para facilitar o manuseio dos microdados são, também, acrescentadas algumas outras variáveis derivadas (IBGE, 2009).

⁷⁰ O arquivo Agregado de Setores Censitários 2000 disponibilizado pelo IBGE é gerado a partir dos resultados do universo (microdados) do Censo Demográfico (IBGE, 2003). Os dados estão armazenados em matrizes sob a forma de planilhas Excel 97 e encontram-se armazenado em 5 (cinco) CD-roms, abrangendo cada uma das grandes regiões geográficas brasileiras: Norte, Nordeste, Sudeste, Sul e Centro-Oeste. São 21 planilhas de dados para cada Unidade da Federação, contendo mais de 3.200 variáveis e 215.811 setores censitários.

trabalho de gestores estaduais que tem a responsabilidade de enviar e validar os dados sistematicamente.

O Banco de dados do Sistema Único de Saúde (DATASUS)⁷¹, do Ministério da Saúde, forneceu informações para o Construto III. Longevidade (Saúde), mais especificamente os dados para os indicadores do Conceito morbidade.

A partir de cada uma das formas de acesso foram selecionados os dados referentes a todos os 5.483 municípios brasileiros para o ano referência 2000 e, posteriormente, reunidos e organizados em um arquivo, no formato Texto (TXT). Este procedimento de manter apenas um arquivo com todos os dados que foram trabalhados, favoreceu significativamente a etapa de seleção, tanto em relação à performance quanto à praticidade das operações subsequentes.

Para evitar falsas interpretações, em todos os casos, é de extrema importância conhecer profundamente a metodologia empregada na coleta e sistematização dos dados e seus respectivos significados. Antes de dar início aos trabalhos com as Bases de Dados do Censo Demográfico, SINESPJC e DATAUS, foi necessário realizar um treinamento específico, junto às fontes produtoras dos dados (IBGE, Ministério da Justiça e Ministério da Saúde respectivamente), para compreender o significado dos dados selecionados. No caso de outras bases de dados públicas, muitas vezes, a possibilidade de treinamento não existe, além de os dados desejados não estão disponíveis, não existem ou não atendem exatamente ao que se busca.

O Exploration Warehouse ou a “Jazida de Dados” utilizada nesta pesquisa é resultado da reunião de informações coletadas (dados primários) por meio da aplicação de dois tipos de questionários, básico e da amostra, ambos utilizados no levantamento do Censo Demográfico do IBGE. O questionário básico contempla as características básicas dos moradores e seus respectivos domicílios, sendo aplicado em todas as unidades domiciliares existentes no momento da coleta, exceto naquelas selecionadas para a amostra. Os elementos da amostra foram investigados através da aplicação do questionário específico da amostra, mais abrangente, contemplando, além das informações do questionário básico, informações sociais, econômicas e demográficas dos moradores.

A população da pesquisa foi constituída de todos os setores comuns ou não especiais e seus respectivos domicílios e moradores residentes nos 5.561 municípios

⁷¹ As informações disponibilizadas no DATASUS são produto de ação integrada com principais sistemas de informação de base nacional utilizados - Ministério da Saúde, Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Pesquisa Econômica Aplicada (IPEA) e Ministério da Previdência Social.

brasileiros em 1º de agosto de 2000. Estão excluídos desta população os setores classificados como Especiais do tipo: aglomerado subnormal; quartéis, bases militares etc.; alojamento, acampamentos etc.; embarcações, barcos, navios etc.; aldeia indígena; penitenciárias, colônias penais, presídios, cadeias etc. e asilos, orfanatos, conventos, hospitais etc.

O Censo Demográfico de 2000 (IBGE, 2000) define como moradora “a pessoa que tinha o domicílio como local de residência habitual e que, na data de referência, estava presente ou ausente por período que não tenha sido superior a 12 meses em relação àquela data, por um dos seguintes motivos: viagens: a passeio, a serviço, a negócio, de estudos, etc.; internação em estabelecimento de ensino ou hospedagem em outro domicílio, visando a facilitar a frequência à escola durante o ano letivo; detenção sem sentença definitiva declarada; internação temporária em hospital ou estabelecimento similar; e embarque a serviço (marítimos)”.

Visando uma melhor aproximação do fenômeno, optou-se por trabalhar apenas com os domicílios particulares permanentes, ou seja, aqueles que “o relacionamento entre seus ocupantes era ditado por laços de parentesco, de dependência doméstica ou por normas de convivência. Construído para servir exclusivamente à habitação e, na data de referência, tinha a finalidade de servir de moradia a uma ou mais pessoas” (IBGE, 2000). Foram excluídos da base de dados os domicílios particulares Improvisados⁷² e Coletivos⁷³.

A partir do sub set de dados oriundos da fase de reunião, seleção e criação do KDD, deu-se início à fase de consistência com todos os domicílios particulares permanentes do tipo Casa, Apartamento e Cômodos compostos de um ou mais aposentos localizados em uma casa de cômodos, cortiço, “cabeça-de-porco” etc.

Os dados foram selecionados para compor o sub set independentemente da condição de ocupação do terreno: se Próprio; Cedido ou Outra condição. E do domicílio, que é classificado como Próprio (quitado, em aquisição); Alugado; Cedido por empregador ou Cedido de outra forma ou Outra condição diferente das citadas anteriormente (oficina, loja; estabelecimento agropecuário arrendado; e, também, o domicílio ocupado por invasão).

⁷² O domicílio particular improvisado foi aquele localizado em “unidade não-residencial que não tinha dependências destinadas exclusivamente à moradia, mas que, na data de referência, estava ocupado por morador”. São enquadrados, nesta definição, as lojas, fábricas, os prédios em construção, vagões de trem, carroças, tendas, barracas, grutas etc.

⁷³ O Domicílio coletivo é caracterizado “quando a relação entre as pessoas que nele habitavam era restrita a normas de subordinação administrativa” (IBGE, 2000). Ficam incluídos, nesta definição, os hotéis, pensões, presídios, cadeias, penitenciárias, quartéis, postos militares, asilos, orfanatos, conventos, hospitais e clínicas (com internação), alojamento de trabalhadores, motéis, camping etc.

Por fim, de posse dos registros e variáveis de cada arquivo, totalizando as informações de 5.483 municípios brasileiros, foi dada seqüência através da fase de consistência dos dados.

6.3.1.4 Consistência

As fases de Limpeza e consistência ou Eliminação de ruído e Enriquecimento foram feitas previamente pelas fontes (IBGE, MJ e MS) e apenas foi necessária uma pequena verificação para tratar os registros com valor Branco (ausência de valor) e Letra “X”, para os casos onde o dado numérico foi omitido para evitar a identificação do informante. Este caso se aplica para os setores que possuem de 1 a 4 domicílios particulares permanentes. Para o valor Branco, foi atribuído valor zero, e os setores que apresentaram a letra “X”, não foram utilizados no cálculo dos indicadores.

Por fim, a base de dados resultantes da fase de coleta e pré-processamento foi composta apenas de 87 variáveis. Foram trabalhados 44.795.101 registros referentes aos domicílios, pessoas residentes e responsáveis, totalizando 168.370.893 habitantes.

6.3.1.5 Compatibilização

Na abordagem proposta nesta tese, o produto resultante da fase de compatibilização foi denominado como a “Jazida de Dados” – JD⁷⁴. Esta classificação está baseada na analogia com a prospecção mineral, proposta no Capítulo 5, na qual o produto da fase de compatibilização pode ser classificado como uma ocorrência mineral, mais especificamente uma jazida, ou seja, o depósito com características relevantes que compensem sua exploração. É importante ressaltar que, na Geologia e Mineração, os conceitos de Jazida e ocorrência mineral não são estáveis no espaço e no tempo, visto que a mudança de alguns fatores básicos (econômicos, tecnológicos, etc.) podem transformar uma ocorrência em jazida e vice-versa.

⁷⁴ A Jazida de Dados será o próprio Universo de estudo ou população de estudo, matériaprima (insumo) para as fases seguintes.

Como toda analogia deixa brechas, não foi diferente neste caso: na Geologia, a etapa de Mineração só tem início se a jazida for economicamente viável, no caso do *KDD*, só é possível determinar se o resultado será economicamente viável após a Mineração.

Conforme defende Reinartz (1999) todo processo deve ser devidamente documentado e, ao término da fase de prospecção, deve ser disponibilizado um manual de procedimentos para futuras atualizações dos dados. Normalmente, isto é feito através de um arquivo que possa, preferencialmente, ser manipulado e atualizado de forma automatizada. Este arquivo deve conter as definições dos elementos de dados e outras características como padrões de propriedades, relacionamentos, utilização e demais documentações inerentes ao processo.

Após a fase de **Compatibilização**, a “Jazida de Dados” já está no formato adequado para dar início à “**Transformação**” dos dados selecionados em Informação, agregando valor do conhecimento sistematizado durante a etapa de **Levantamento**, mais especificamente, durante a definição do modelo e seleção dos indicadores.

6.3.2 Mineração

No decorrer da Mineração de Dados, foram executadas as seguintes tarefas: de Transformação dos Dados; Escolha da Função de Mineração; Algoritmo de Busca; Avaliação dos resultados e Disseminação. Todas estas tarefas tiveram uma única fonte de dados cujo suporte foi a “Jazida de Dados”.

Para diversos autores estudados, a etapa de transformação não faz parte da etapa de *Data Mining*, entretanto estes mesmos autores afirmam que, na etapa de transformação, os dados devem estar no formato exigido pelos algoritmos de *DM* e deve-se procurar reduzir o número de variáveis que serão mineradas. Como o trabalho com o *DM* exige o uso de diversas técnicas e muitas delas necessitam de que os dados estejam dispostos em formatos diferenciados, optou-se por manter a fase de transformação dentro da etapa de Mineração de Dados.

6.3.2.1 Transformação

Durante esta etapa, os dados sofreram uma transformação sintática e semântica. A transformação sintática não altera o significado dos dados, visa apenas atender aos requisitos das ferramentas de Mineração que serão utilizadas nas etapas subseqüentes. Normalmente, recorre a conversões, recodificações, conversões de unidades, padronizações das variáveis disponíveis na “Jazida de Dados”.

Já a transformação semântica visa basicamente atender, através do cálculo de indicadores, ao modelo hipotético previamente definido. Podem ser feitas categorizações, transformações de representação, discretização⁷⁵ de atributos quantitativos, entre outras transformações.

Em outras palavras, pode-se dizer que a subfase de **Transformação** promove o enriquecimento e adequação da “Jazida de Dados” com o objetivo de atender aos algoritmos selecionados e suas respectivas **Funções** empregadas na fase de Mineração, tarefa esta que certamente se repete diversas vezes.

6.3.2.2 Funções e Algoritmos

Após todo o trabalho de **Prospecção** e de posse da “Jazida de Dados”, deu-se início à escolha da **Função** ou conjunto de Funções (Sumarização, Associação, Regressão, *Cluster*, etc.), empregada na Mineração dos Dados. Esta escolha estava condicionada ao que foi estabelecido nos objetivos do estudo e na delimitação das perguntas a serem respondidas, ambos definidos na subfase de **Identificação de Objetivos**.

A seleção da **Função** determina a maneira como é feita a busca por reconhecimento de padrões e relacionamentos complexos no processo *KDD*. Esta seleção é apresentada, por grande parte dos autores, como o primeiro passo da fase da Mineração e pode ser considerada uma tarefa simples se examinada na ótica do analista de dados, porém, no caso do especialista do domínio e do usuário, estes dificilmente detêm conhecimento necessário acerca das vantagens e limitações de cada solução que pode ser formulada. Para

⁷⁵ A discretização consiste na observação de uma variável ou sinal contínuo no espaço/tempo em posições/instantes discretos.

Diniz e Louzada Neto (2000), o sucesso de uma tarefa de Mineração de Dados está diretamente ligada à experiência e à intuição do analista.

Embora as **Funções** devam ser selecionadas separadamente, nos casos onde mais de uma Função é escolhida, deve-se estudar a seqüência que será utilizada para o seu encadeamento. Este fato é extremamente relevante, visto que nem todas as Funções são aplicáveis a cada e/ou toda a tarefa. Em certos casos, somente algumas Funções são apropriadas e sua concatenação nem sempre é possível.

Nesta tese, visando buscar uma maior aproximação da representação do fenômeno desenvolvimento social, foram utilizadas as seguintes funções de Mineração de dados: Sumarização e Associação.

Resumidamente, a Mineração envolveu as seguintes etapas:

- a) Análise preliminar dos dados por meio da **Sumarização**;
- b) Cálculo da matriz de correlações para os indicadores (variáveis) selecionados a partir da função **Associação** e a redução de dimensionalidade através da criação de um novo vetor de variáveis de menor dimensão, obtido através da Análise de Componentes Principais;

Definido o conjunto de Funções e seu respectivo encadeamento, partiu-se, então, para a busca do algoritmo ou grupo de algoritmos mais apropriados para cada Função. Como foi visto no Capítulo 5, algoritmos derivados de Redes Neurais, Algoritmos Genéticos, Lógica Fuzzy, Árvore de Decisão e, principalmente, a Análise Estatística têm sido amplamente utilizados na Mineração de Dados. Nesta tese, como apresentado a seguir, optou-se por trabalhar com algoritmos derivados da Análise Estatística.

Cada tipo de técnica possui uma significativa gama de algoritmos, o que possibilita testar mais de uma técnica em uma mesma Função ou até mais de um algoritmo dentro de uma mesma técnica. Para atender aos requisitos destes algoritmos, a “Jazida de Dados” deverá ser transformada.

De posse da “Jazida de Dados” contendo todas as variáveis e registros para o estudo, descreve-se detalhadamente, a seguir, cada uma dessas etapas.

a) Sumarização

A função **Sumarização** foi aplicada durante a Mineração dos Dados em dois momentos: no primeiro momento, buscou-se um estudo preliminar dos indicadores selecionados e armazenados na “Jazida de Dados”, bem como a orientação quanto à escolha das técnicas para as funções de Mineração subseqüentes. Em um segundo momento, após as funções **Associação (B)**, **Regressão (C)** e **Cluster (D)**, foi possível elaborar uma síntese das relações hipotéticas do modelo de desenvolvimento social extraídas durante o processo.

Nos dois momentos, inicial e final, onde foi aplicada a função complementar A.1) **Caracterização**, optou-se por Análise Exploratória de Dados e Estatística Descritiva por mera questão de disponibilidade de *software*. Cabe ressaltar que também poderiam ser utilizadas as técnicas de *SQL* tradicional ou *OLAP*, entre outras. Através dessas técnicas mais simples da estatística, foi possível organizar e sintetizar os dados. Foram calculadas medidas de posição e dispersão, as medidas de posição utilizadas foram: máximo, mínimo e média aritmética, mediana e separatrizes. Já as medidas de dispersão foram desvio-padrão e coeficiente de variação de Pearson.

A função complementar A.2) **Visualização** teve maior participação no segundo momento. Este importante recurso de análise contribuiu de forma significativa para a interpretação dos resultados, foram construídos gráficos e tabelas como base nos resultados obtidos durante e após a conclusão da etapa de **Cluster (D)**.

Através dos diagramas de caminho (path-diagram) é possível representar as estruturas hierárquicas, gerados durante a função de Visualização, foi possível atender ao objetivo de mostrar a forma como se estruturam as relações que conformam o Desenvolvimento Social mensurado por meio dos Índices e Sistemas de Indicadores estudados.

b) Associação

Um problema recorrente em *Data Mining* está relacionado ao grande número de variáveis que são inicialmente selecionadas para análise. Diante desta situação, é necessário recorrer a métodos de redução de dimensionalidade⁷⁶ do universo a ser estudado.

⁷⁶ Como é comum em *KDD*, reduzir dimensionalidade significa uma redução do número de variáveis ou indicadores utilizados no estudo.

No decorrer da etapa de **Associação**, como é freqüente no *KDD*, foi necessário reduzir o número de variáveis, sendo escolhido o método de redução de dimensionalidades Análise de Componentes Principais - ACP. A ACP foi empregada com o objetivo de identificar o menor número de fatores comuns através da análise das inter-relações entre os 15 indicadores de Desenvolvimento Social inicialmente selecionados.

Existem diferentes formas de reduzir o número de variáveis a serem mineradas e estas formas podem ser agrupadas em dois tipos. O primeiro, mais utilizado, apela para o especialista do domínio que, através da heurística, seleciona o conjunto de variáveis que farão parte da base de dados que será minerada. O segundo tipo está relacionado aos métodos estatísticos conhecidos como técnicas de seleção de variáveis (DRAPER; SMITH, 1998 apud DINIZ; LOUZADA NETO, 2000).

As técnicas de seleção de variáveis podem ser subdivididas em duas categorias: as que levam em conta apenas a variância comum nos dados e selecionam apenas algumas variáveis do conjunto original através de regressões, eliminação *backward*, regressão *stepwise*, entre outras; e as que levam em conta a variância total dos dados e mantêm todas as variáveis originais e pode ser realizada através da ACP - Análise de Componentes Principais⁷⁷. A ACP de Harold Hotelling permite a criação de um novo vetor de variáveis de menor dimensão, cujos componentes são combinações lineares das variáveis ortogonais.

As técnicas baseadas em ACP e Análise Fatorial (AF) são amplamente utilizadas em diversas áreas, principalmente em análises de fenômenos complexos e multidimensionais como é o caso do Desenvolvimento Social (ROCHA, 1998), além de serem amplamente utilizadas para testar modelos hipotéticos de múltiplos indicadores (HAIR et al., 1992; MARTÍNEZ, 2001).

A Análise de Componentes Principais, aplicada aos indicadores selecionados, permite identificar, entre as 15 variáveis originais (indicadores) iniciais para os 5.483 municípios brasileiros, o subconjunto de indicadores que estão correlacionados com os cinco construtos associados ao Desenvolvimento Social, contribuindo, desta forma, para a confirmação dos Construtos selecionados – Dinâmica Populacional, Trabalho e Rendimento, Saúde, Educação e Segurança –, conforme o modelo hipotético inicialmente proposto.

⁷⁷ Uma grande contribuição na estatística para selecionar as variáveis. Foi anunciada pela primeira vez em 1933 em artigo publicado no *Journal of Educational Psychology*, intitulado "Analysis of a Complex of Statistical Variables with Principal Components", de autoria de Harold Hotelling.

A análise fatorial pode ser explicada como uma técnica que consiste em resumir a informação contida em uma base de dados com X variáveis⁷⁸. Para isso, identifica-se um número de fatores F , sendo que o número de fatores deverá ser sempre inferior ao número de variáveis originais ($F < X$). Os fatores representam as variáveis originais com uma perda mínima e mensurável de informação. Estes fatores não são conhecidos *a priori*, sendo estimados como combinações lineares das variáveis observáveis. A expressão geral para o j -ésimo fator (F_j) pode ser escrita da seguinte maneira:

$$F_j = \sum_{i=1}^p w_{ij} x_i = w_{1j} x_1 + w_{2j} x_2 + \dots + w_{pj} x_p, \text{ onde } w_{pj} \text{ são os coeficientes fatoriais,}$$

x_i são as variáveis observáveis e p é o número de variáveis.

O objetivo de empregar a AF foi identificar um número de fatores (Cinco Construtos) inferior ao número de variáveis originais (15 Indicadores), de maneira que se pudesse descrever o fenômeno Desenvolvimento Social de forma mais simplificada, em que os fatores identificados, aqui chamados de Construtos, são vistos como dimensões do fenômeno que não podem ser observados diretamente.

A aplicação da Análise Fatorial não pode ser feita indiscriminadamente, pois a técnica exige alguns pressupostos que devem ser verificados antes da sua aplicação. Para verificar a adequação da técnica, inicialmente foi testada a hipótese de que a matriz de correlação é uma matriz identidade, consistindo na transformação qui-quadrada do determinante da matriz de correlação. Para este fim, foi utilizado o teste de esfericidade de Bartlett, em outras palavras, um teste estatístico para contrastar a hipótese nula de que os $p-m$ autovalores restantes são iguais (sendo p o número original de variáveis e m o número de fatores ou componentes retidos). Cada autovalor foi excluído de maneira seqüencial até que não se pudesse rejeitar a hipótese nula através de um teste qui-quadrado.

Em seguida, foram comparados os valores dos coeficientes de correlação observados com os valores dos coeficientes de correlação parcial, e o resultado da comparação foi calculado através da medida de Kaiser-Meyer-Olkin (KMO), obtida através da fórmula:

⁷⁸ Poderia ser utilizada a palavra Indicadores.

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} a_{ij}^2}, \text{ onde } r_{ij} \text{ e } a_{ij} \text{ são, respectivamente, o coeficiente de}$$

correlação simples e o coeficiente de correlação parcial entre as variáveis x_i e x_j . De um modo geral, o *KMO* é considerado muito bom se está entre 0,8 e 0,9 e, quanto mais próximo da unidade ele estiver, tanto mais será apropriada a utilização da Análise Fatorial (JOHNSON; WICHERN, 1998).

Após a confirmação da adequabilidade da AF e selecionados os indicadores para análise, foi efetuada a exclusão dos indicadores em que a comunalidade⁷⁹ explicada fosse menor que 0,50 (HAIR et al., 1992). Este procedimento permitiu manter apenas os indicadores muito significantes para explicar o fenômeno Desenvolvimento Social.

Diante dos indicadores selecionados na fase anterior, determinou-se o número de Construtos (fatores) através da análise dos autovalores⁸⁰. Nos casos em que o autovalor foi superior a 1,0 (HAIR et al., 1992) ficou mantido o Construto proposto no modelo hipotético já apresentado (Figura 14); nos demais casos, o Construto foi excluído. Desta forma, permaneceram apenas os Construtos que explicaram mais do que um Conceito representado por seus indicadores (Conceitos operacionais), poderia explicar isoladamente.

Por fim, para identificar os indicadores com cargas fatoriais altas no Construto (componentes de cada fator), recorreu-se à matriz dos fatores, rotacionada pelo método Varimax⁸¹ (HAIR et al., 1992).

A busca por um Construto (fator) comum aos indicadores significa que os Conceitos operacionais (indicador) que entram na composição de um determinado Construto devem estar fortemente correlacionados entre si e fracamente com os Conceitos Operacionais que fazem parte da composição de outro Construto (JOHNSON; WICHERN, 1998).

Segundo Ensslin et al. (2001), para cada critério superior (Construto) deve haver ao menos dois critérios inferiores (Conceitos operacionais) para que se justifique a necessidade de uma estrutura hierárquica. Os critérios (Conceitos operacionais) de nível hierárquico inferior devem ser mutuamente exclusivos e devem fornecer uma descrição exaustiva do critério de nível superior (Construto). Para cada Construto deve haver ao menos

⁷⁹ Comunalidade pode ser explicada como sendo a proporção da variância total de cada variável que é explicada pelo conjunto de fatores comuns.

⁸⁰ Valores próprios da matriz de correlação, em inglês *eigenvalue*.

⁸¹ A rotação ortogonal permite que os coeficientes de correlação entre os indicadores e as dimensões fiquem o mais próximo de zero, 1 ou -1, contribuindo para a interpretação do modelo.

dois Conceitos operacionais para que se justifique a representação por meio de uma estrutura hierárquica.

A interpretação dos Construtos extraídos pela função da Mineração de Dados foi feita a partir da análise dos indicadores com maior carga⁸² no fator. Este procedimento foi utilizado porque, quando a análise fatorial faz parte de uma matriz de correlação, as cargas fatoriais são os próprios coeficientes de correlação entre os indicadores (Conceitos Operacionais) e os Construtos (Fatores). Estes coeficientes representam as próprias cargas fatoriais, que, por sua vez, indicam quanto um indicador está relacionado a um Construto (Fator).

Conforme recomenda Matinez (2001), durante a aplicação da AF, foram observadas duas condições básicas: parcimônia e interpretabilidade. Para o autor, mediante o princípio da parcimônia, os fenômenos devem ser explicados com um menor número possível de elementos, e não teria sentido um número de fatores maior que o número de variáveis. Já para o princípio da interpretabilidade, a estrutura da base de dados resultante deve ser susceptível a uma interpretação plausível. Uma boa solução fatorial deve ser, portanto, simples e interpretável.

6.3.2.3 Avaliação dos resultados

Embora a avaliação dos resultados seja feita de maneira iterativa durante todo o processo, optou-se por delimitar uma fase exclusivamente para este fim. Esta fase visa avaliar e validar todo processo e basicamente pode ser avaliada sob quatro aspectos: a) adequação do modelo hipotético aos dados empíricos selecionados; b) conveniência das Funções de Mineração selecionadas; c) adequação dos algoritmos escolhidos para processá-las e, principalmente, d) os achados e conhecimento extraído de todo o processo. Enquanto, para a fase de prospecção, o produto final é a “Jazida de Dados”, na fase de Mineração, o total dos resultados esperados pode ser definido como a soma da adequação do modelo hipotético proposto (a); das funções selecionadas (b); algoritmos empregados (c) e conhecimento extraído (d), ou seja, de maneira simplificada, pode ser representada como $Resultados = função(a,b,c,d)$.

⁸² As cargas foram analisadas em valores absolutos.

Em função do nível de detalhe que a avaliação dos resultados em um processo *KDD* exige, nesta tese, este tema é tratado no próximo capítulo, onde estão analisados os resultados da avaliação da aplicação do *KDD*, suas etapas nas bases de dados públicas trabalhadas com o objetivo de confrontar os dados empíricos sobre a realidade das condições de Desenvolvimento Social no Brasil com um modelo hipotético que traduz as postulações teóricas identificadas nos Índices e Sistemas de Indicadores de estudados, a fim de detectar particularidades locais do fenômeno e estabelecer e representar conjecturas explicativas por meio de um modelo empírico do Desenvolvimento Social. São apresentadas as principais funções do *KDD* utilizadas, suas respectivas verificações de adequação para uso, o teste do modelo hipotético proposto e, por fim, o modelo empírico do Desenvolvimento Social estudado.

6.3.2.4 Disseminação dos resultados

Normalmente, a aplicação do protocolo de métodos *KDD* ocorre inserida em um contexto de Sistemas de Informação (SI), mais especificamente Sistemas de Apoio à Decisão (SAD). Estes sistemas já possuem mecanismos definidos e implantados de disseminação da informação, que possibilitam fornecer *feedback* e aplicar os conhecimentos extraídos nos processos decisórios. No caso de o *KDD* ser utilizado fora do contexto da Gestão do Conhecimento ou de um Sistema de Informação, a fase de Disseminação dos resultados pode ser considerada como fator crítico de sucesso e deve ser tratada com a devida importância.

Para Fayyad et al. (1996), esta etapa consiste da simples documentação e disseminação dos resultados até a incorporação deste conhecimento na *performance* do sistema, procurando resolver conflitos com o conhecimento anterior. Os resultados podem ser disseminados em diferentes meios, tais como o próprio sistema de informação em que o *KDD* está inserido, caso contrário pode-se recorrer a material impresso, reuniões, apresentações, correio eletrônico ou mesmo pela *Internet/Intranet*.

A partir das propostas de *KDD* discutidas anteriormente, optou-se por formatar as fases de acordo com o estudo em questão. Como cada trabalho deve ter sua metodologia adequada a suas necessidades, a tentativa de generalização dos autores, citados anteriormente, não dispensa uma tentativa de sistematizar os conceitos de uma área recente, que cresce com a velocidade do crescimento das inovações da Tecnologia da Informação.

Por fim, neste capítulo foi elaborado o modelo gráfico hipotético das relações entre os indicadores para esta pesquisa, a partir dos conceitos levantados com o estudo das principais propostas de Índices e Sistemas de Indicadores de desenvolvimento Social no Brasil, disponíveis na literatura.

No próximo Capítulo 7, são apresentados os resultados obtidos com a aplicação do *KDD* em bases de dados públicas com o objetivo de confrontar o modelo hipotético e os dados empíricos reunidos na jazida de dados, resultando em uma proposta integrada para a operacionalização do protocolo de métodos *KDD* no estudo do Desenvolvimento Social.

7 MODELOS EMPÍRICOS DO DESENVOLVIMENTO SOCIAL

An appropriate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. John W. Turkey (1948).

Este capítulo apresenta uma análise da estrutura de relações entre os indicadores associados ao desenvolvimento social. São evidenciados os principais resultados da aplicação do *Knowledge Discovery in Databases* (KDD) nas bases de dados públicas trabalhadas, bem como as respectivas verificações de adequação para seu uso. O objetivo deste capítulo é estabelecer um modelo empírico, com base nos dados disponibilizados pelo IBGE e Ministério da Saúde para 5.483 municípios brasileiros, a fim de detectar e estabelecer conjecturas explicativas sobre este fenômeno no Brasil.

Para consecução do objetivo proposto para este capítulo, apresentam-se inicialmente os resultados da aplicação das funções sumarização e associação do KDD. Em seguida, são confrontados, o modelo hipotético⁸³ e os dados disponíveis na jazida de dados, resultando na identificação do modelo empírico validado, comum às ferramentas brasileiras de mensuração do desenvolvimento social, por meio da aplicação do método KDD em bases de dados públicas.

No capítulo anterior, após a consolidação de todo material tratado na fase de Prospecção do *KDD*, os dados foram submetidos à fase de transformação, quando foram calculados os indicadores de Desenvolvimento Social selecionados para este estudo de aproximação. O resultado da aplicação dessas fases foi uma “Jazida de Dados” com 15 indicadores e 5.483 registros, que foi utilizada em seguida, na etapa de mineração com o suporte computacional dos pacotes⁸⁴ *SPADWIN*, *R*, *WizRule*, *MS-Excel*, *Weka3* e *NetDraw*.

Essa “Jazida de Dados”, contendo os dados obtidos para todos os indicadores selecionados para representar o Desenvolvimento Social, foi confrontada com o modelo hipotético, por meio da aplicação do método *Knowledge Discovery in Databases* (KDD) em bases de dados públicas, resultando em um modelo empírico para termo teórico, comum às ferramentas brasileiras de mensuração do desenvolvimento social.

⁸³ Apresentado no capítulo 6 desta tese.

7.1 SUMARIZAÇÃO DA BASE DE DADOS

A primeira função da fase de Mineração empregada foi a da Sumarização, o que resultou nos dados apresentados na Tabela 3. Como pode ser observado na referida tabela, os indicadores que apresentam maior dispersão relativa são: Doenças relacionadas ao saneamento ambiental inadequado; Coeficientes de mortalidade por homicídio e Prevalência de desnutrição total. O coeficiente de variação de Pearson demonstra a grande dispersão relativa existente em indicadores relacionados com os construtos III. Longevidade (Saúde) e V. Segurança e oferece indícios relacionados com problemas de consistência nos dados disponíveis no Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde.

Tabela 3 – Resultado da sumarização dos 15 indicadores do Desenvolvimento Social para os 5.483 municípios brasileiros, Brasil – 2000.

INDICADOR DE DESENVOLVIMENTO SOCIAL	MÍNIMO	MÁXIMO	MEDIANA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO (%)
1-Taxa de Crescimento da População	0,48	4,44	1,08	0,24	21,38
2-Taxa de fecundidade	1,56	7,79	2,67	0,74	25,85
3-Índice de Gini da distribuição do rendimento	0,36	0,80	0,56	0,06	10,33
4-Taxa de desocupação	0,06	0,83	0,52	0,09	18,44
5-Rendimento familiar per capita	31,93	954,65	159,57	96,30	56,21
6-Rendimento médio mensal	99,02	2.119,96	365,56	202,24	51,39
7-Esperança de vida ao nascer	54,90	78,18	68,28	4,85	7,15
8-Taxa de mortalidade infantil	5,38	98,12	29,38	18,38	54,12
9-Prevalência de desnutrição total	0,00	119,34	0,00	3,74	636,48
10-Doenças relacionadas ao saneamento ambiental inadequado	0,00	3.487,66	0,00	83,46	2.327,98
11-Taxa de escolarização	0,00	30,74	1,64	3,66	121,36
12-Taxa de alfabetização	39,34	99,09	82,09	12,42	15,86
13-Escolaridade	0,81	9,65	4,08	1,28	31,75
14-Coeficientes de mortalidade por homicídios	0,00	1.406,74	1,13	25,47	981,47
15-Coeficiente de mortalidade por acidentes de transporte	0,00	27,66	1,32	1,69	98,11

Fonte: IBGE / DATASUS, 2000

Segundo os dados investigados, no ano 2000 ainda existiam no Brasil 3.049 municípios sem informação sobre Prevalência de desnutrição total (9). Entre esses municípios (sem informação sobre desnutrição) estão algumas capitais, como Manaus – AM, Belém – PA, João Pessoa – PB e Maceió – AL. Outro indicador que apresentou problemas de subnotificação foi o de Doenças relacionadas ao saneamento ambiental inadequado (10). Totalizaram 1.731 municípios sem informação e, entre esses, 21 municípios grandes, com população acima de 100.000 habitantes.

⁸⁴ SpadWin – CISIA-CERESTA, 1998, Saint-Mandé, France. R – R Foundation for Statistical Computing. Wizrule – WizSoft Incorporated. Weka3 – University of Waikato. NetDraw – Harvard: Analytic Technologies.

Após esta breve descrição dos indicadores de Desenvolvimento Social selecionados para compor o modelo empírico, segue-se a análise dos resultados da Função Associação.

7.2 RESULTADO DA ASSOCIAÇÃO – MATRIZ DE CORRELAÇÕES

Com o objetivo de apurar a intensidade da associação entre os indicadores, conforme descrito no Capítulo 6, o primeiro algoritmo empregado na etapa de **Associação** foi o coeficiente de correlação de Spearman⁸⁵. Os resultados da aplicação podem ser observados através da matriz de correlações⁸⁶ da Tabela 4 a seguir:

Tabela 4 – Resultado da associação dos 15 indicadores selecionados para compor o Modelo Empírico do Desenvolvimento Social para os 5.483 municípios brasileiros, Brasil – 2000.

INDICADOR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-Taxa de Crescimento da População	1,00														
2-Taxa de fecundidade	0,07	1,00													
3-Índice de Gini da distribuição do rendimento	0,04	0,40	1,00												
4-Taxa de desocupação	0,02	0,34	0,32	1,00											
5-Rendimento familiar per capita	0,16	-0,70	-0,27	-0,49	1,00										
6-Rendimento médio mensal	0,20	-0,61	-0,22	-0,48	0,96	1,00									
7-Esperança de vida ao nascer	0,05	-0,62	-0,38	-0,46	0,80	0,76	1,00								
8-Taxa de mortalidade infantil	-0,04	0,64	0,41	0,49	-0,84	-0,80	-0,95	1,00							
9-Prevalência de desnutrição total	0,07	-0,02	0,06	-0,01	0,05	0,07	-0,02	-0,01	1,00						
10-Doenças relacionadas ao saneamento ambiental inadequado	0,23	0,14	0,28	0,23	-0,08	-0,04	-0,20	0,21	0,11	1,00					
11-Taxa de escolarização	0,05	-0,64	-0,21	-0,37	0,76	0,72	0,62	-0,66	0,10	-0,01	1,00				
12-Taxa de alfabetização	0,15	-0,65	-0,36	-0,47	0,90	0,88	0,80	-0,84	0,04	-0,17	0,70	1,00			
13-Escolaridade	0,19	-0,69	-0,28	-0,36	0,91	0,88	0,75	-0,79	0,07	-0,03	0,77	0,92	1,00		
14-Coefficientes de mortalidade por homicídios	0,17	0,02	0,15	0,07	0,13	0,16	-0,01	-0,02	0,10	0,25	0,10	0,06	0,14	1,00	
15-Coefficiente de mortalidade por acidentes de transporte	0,21	-0,28	-0,05	-0,18	0,48	0,49	0,34	-0,38	0,10	0,06	0,41	0,39	0,45	0,25	1,00

Fonte: IBGE / DATASUS, 2000

Algumas observações merecem maior atenção na análise da Tabela 7.2 e, entre as principais, destacam-se quatro: a primeira observação diz respeito à forte associação ($R_{xy}=0,91$), por município, da Taxa de Escolaridade (13) com Rendimento familiar per capita (5). Esta associação demonstra que, nos municípios onde a média de anos de estudo da

⁸⁵ O coeficiente de correlação de Spearman é utilizado em variáveis com escala ordinal ou quando se deseja medir a correlação entre séries com forte assimetria ou dispersão.

⁸⁶ Na matriz de correlações apresentada, foram selecionadas apenas as correlações mais expressivas para a análise.

população é maior, a proporção de famílias com renda bruta elevada (Rendimento familiar per capita)(,) também é maior.

A segunda observação feita se refere à forte associação entre Rendimento médio mensal (6) e as Taxas de Alfabetização (12) e de Escolaridade (13). Esses resultados corroboram o reconhecimento da relação entre os indicadores pertencentes aos construtos II. Sistema de Produção (Trabalho e Rendimento) e IV. Socialização (Educação). Fato que confirma outros estudos empíricos e a teoria revisada (Capítulos 2 e 4).

A terceira observação é voltada para o fato de que existe uma forte associação entre os indicadores Taxa de Alfabetização (12) e de Escolaridade (13), dos construtos IV. Socialização (Educação) e os indicadores: Esperança de vida ao nascer (7) e Taxa de mortalidade infantil (8), do construto III. Longevidade (Saúde). Em outras palavras, isto quer dizer que a Educação dos moradores dos municípios sugere uma associação direta com o aumento do número médio de anos de vida, esperados para um recém-nascido e uma redução do número de óbitos de menores de um ano de idade, de uma população de 168.370.893 habitantes, distribuídos nos 44.795.101 domicílios particulares permanentes, existentes nos municípios analisados.

A quarta e última observação que merece maior atenção diz respeito à associação negativa entre o indicador que representa o número médio de filhos nascidos vivos que uma mulher teria ao fim do seu período reprodutivo (2.Taxa de fecundidade total), pertencente ao construto I. Dinâmica Populacional, e os indicadores dos construtos II. Sistema de Produção (Trabalho e Rendimento) e IV. Socialização (Educação). Ou seja, segundo o coeficiente de correlação de Spearman , calculado a partir dos dados de todos os municípios brasileiros em 2000, quanto maior for o nível da Socialização e do Sistema Produtivo do município, menor será o número médio de filhos de suas habitantes.

É importante alertar para o fato de que o nível de agregação dos dados não permite afirmações mais contundentes sobre a associação entre os Indicadores e construtos, isto porque não estão sendo analisados os domicílios e as pessoas individualmente e sim, o valor total agregado por município (unidade de observação). De qualquer forma, os resultados auxiliam na proposição de questões importantes acerca das condições de Desenvolvimento Social no Brasil.

Outro ponto que chama a atenção, ao se observar a Tabela 7.2 é a baixa associação existente (coeficientes de correlação próximos a Zero) entre os indicadores que

integram o construto I. Dinâmica Populacional (1. Taxa de Crescimento da População e 2. Taxa de fecundidade) e os indicadores que integram o construto V. Segurança (14. Coeficientes de mortalidade por homicídios e 15. Coeficiente de mortalidade por acidentes de transporte). A fraca associação existente entre os indicadores 14 e 15 é um forte indício da falta de consistência interna dos indicadores de cada Conceito pertencente aos construtos I. Dinâmica Populacional e V. Segurança.

Assim como os indicadores Prevalência de desnutrição total (9) e Doenças relacionadas ao saneamento ambiental inadequado (10), não apresentam correlação significativa com os indicadores do construto III. Longevidade (Saúde) ou com os demais indicadores dos construtos I, II, IV ou V.

7.3 A ANÁLISE DO MODELO EMPÍRICO DO DESENVOLVIMENTO SOCIAL BRASILEIRO

Os procedimentos adotados nesta seção objetivam exclusivamente verificar a aceitabilidade do ajuste do modelo hipotético, porém os procedimentos adotados não permitem concluir que a aferição realizada no modelo hipotético, resulte em validação empírica. A partir do emprego do KDD, apenas é possível afirmar que o modelo hipotético⁸⁷ testado tem maior aproximação com os dados empíricos utilizados para representar a realidade do desenvolvimento Social no Brasil.

Para Hox e Bechger (1998), quando um modelo hipotético é corroborado pelos dados empíricos, não se pode afirmar que o modelo empírico seja verdadeiro, porém, pode-se dizer que o modelo hipotético não foi falseado. Tal resultado não implica na inexistência de outros modelos hipotéticos concorrentes que possam ser avaliados como verdadeiros. Esse argumento dos autores guarda uma forte relação com a perspectiva falsificacionista de produção do conhecimento de Popper (1972), para quem "o método da ciência é o método de conjecturas audazes e engenhosas seguidas de tentativas rigorosas de falseá-las".

⁸⁷ O modelo hipotético foi elaborado a partir dos indicadores selecionados entre os Índices e Sistemas de Indicadores de Desenvolvimento Social apresentado no Capítulo 6 desta tese.

Segundo Wallis (1976, p.94) uma estrutura é considerada representada por um determinado modelo e um conjunto de dados selecionados, apenas nos casos em que existe uma única estrutura admissível para os respectivos modelos e dados avaliados.

O emprego do KDD nesta pesquisa não tem apenas a pretensão de comprovar a existência de uma única estrutura empírica para o modelo hipotético identificado, e sim, permitir também uma reflexão sobre as possibilidades de utilização desse poderoso método em estudos futuros que possam verificar a plausibilidade de tais conjecturas.

7.3.1 Avaliação da consistência interna dos construtos do modelo empírico do desenvolvimento social

Para verificar a consistência interna dos construtos, como foi apresentado no Capítulo 6, em pesquisas da área de Administração, é comum empregar a Análise Fatorial (AF) dos indicadores propostos para cada um dos construtos identificados no modelo empírico sugerido inicialmente. Essa análise permite verificar os indicadores que estão fortemente associados entre si e representam um construto único.

Além da Análise Fatorial, a consistência interna dos Indicadores de cada construto do modelo empírico também pode ser aferida por meio do popular coeficiente Alfa de Cronbach. São considerados aceitos os construtos que, em pesquisas exploratórias, apresentam coeficientes de Cronbach com valores entre 50% e 70% S (HAIR et al. 1992). Porém, como salienta Pasquali (2003), quando o número de itens é pequeno este dado deve ser relativizado, visto que, nesse caso, o próprio item em análise afeta substancialmente o escore total a seu favor.

Um construto ou modelo empírico é classificado, *grosso modo*, como tendo fidedignidade⁸⁸ apropriada quando o Alfa de Cronbach é, pelo menos, 0,70 (NUNNALLY, 1978). Contudo, em alguns cenários de investigação das ciências sociais, um Coeficiente Alfa

⁸⁸ O termo “reliability” foi traduzido como “fidedignidade”. Em outros textos este termo tem sido traduzido por “precisão”, “fidelidade” ou “fiabilidade”. Indica o grau de consistência interna dos indicadores, que representam o conceito de cada construto associado, sendo determinada pelo valor do “alfa de Cronbach” (CRONBACH, 1984). Uma fidedignidade alta significa dizer que é mais fácil fazer uma distinção entre os diversos estágios ou níveis de implementação de determinado indicador do que se houvesse uma baixa fidedignidade. Para Hayes (HAYES, 1992), a fidedignidade alta “torna mais provável a descoberta de relacionamentos entre indicadores realmente relacionadas”.

de 0,60 é considerado aceitável desde que os resultados obtidos com esse instrumento sejam interpretados com precaução (DEVELLIS, 1991).

Os níveis de fidedignidade recomendados por esses autores devem servir apenas como uma base de partida e, não, como critério definitivo de classificação da fidedignidade do modelo. Torna-se claro que, quanto mais elevadas forem as covariâncias (ou correlações entre os itens), maior é a homogeneidade dos itens e maior é a consistência com que medem o mesmo construto ou termo teórico.

O Alfa de Cronbach é considerado por alguns autores, como uma opção conservadora e que fornece uma subestimativa da verdadeira fidedignidade do modelo empírico, especialmente quando é empregado em construtos de indicadores heterogêneos ou quando definem modelos multifatoriais. Thompson (1994) e Vacha-Haase (1998) referem que nenhum modelo ou construto é fidedigno, apenas os “dados são fidedignos ou não-fidedignos”. Segundo estes autores, apenas se pode falar de “*score reliability*” e é errado inferir, a partir de uma estimativa de Alfa de Cronbach, que este é ou não fidedigno. Um Alfa de Cronbach é apenas uma estimativa da fidedignidade dos dados obtidos com um determinado modelo empírico.

Segundo DeVellis (1991), a maioria dos pesquisadores tende, não apenas a considerar o Alfa de Cronbach uma medida indicada para o estudo de instrumentos de mensuração, como tendem a percebê-lo como fornecendo “estimativas fiáveis” da fidedignidade do modelo empírico. Para o autor, um modelo empírico com validade reduzida é um instrumento que mede algo que não se pretendia medir.

A Tabela 5 apresenta o resultado da análise de cada construto do Desenvolvimento Social, individualmente, os resultados dos testes *Kaiser-Meyer-Olkin* (KMO) e de esfericidade de Barlett, o Alfa de Cronbach e a variância explicada, assim como, os indicadores pertencentes aos construtos, com respectivos carregamentos. A análise da tabela mostra que, entre os cinco construtos de Desenvolvimento Social do modelo empírico proposto no Capítulo 6, apenas dois construtos apresentam coeficientes de Cronbach inferiores aos 50% recomendado na literatura.

Tabela 5 – Análise multivariável dos cinco construtos de Desenvolvimento Social Brasil, 2000

Construto	Indicador	Carregamento	KMO	Esfericidade de Barlett	Alfa de Cronbach	Variância
I. Dinâmica Populacional	1 - Taxa de crescimento da população	0,560	0,500	79,436 gl 1	0,208	55,998
	2 - Taxa de fecundidade	0,560				
II. Sistema de Produção (Trabalho e Rendimento)	3 - Índice de Gini da distribuição do rendimento	0,200	0,608	14.175,347 gl 6	0,766	59,970
	4 - Taxa de desocupação	0,501				
	5 - Rendimento familiar per capita	0,858				
	6 - Rendimento médio mensal	0,840				
III. Longevidade (Saúde)	7 - Esperança de vida ao nascer	0,953	0,509	11.127,168 gl 6	0,648	49,063
	8 - Taxa de mortalidade infantil	0,953				
	9 - Prevalência de desnutrição total	0,424				
	10 - Doenças relacionadas ao saneamento ambiental inadequado	0,666				
IV. Socialização (Educação)	11 - Taxa de escolarização	0,739	0,637	13.243,470 gl 3	0,895	83,888
	12 - Taxa de alfabetização	0,845				
	13 - Escolaridade	0,933				
V. Segurança	14 - Coeficiente de mortalidade por homicídios	0,507	0,500	0,994 gl 1	-0,010	50,673
	15 - Coeficiente de mortalidade por acidentes de transporte	0,507				

Fonte: Elaboração própria

Os construtos Dinâmica Populacional (I) e Segurança (V) apresentam, respectivamente, coeficientes de Cronbach de 0,208 e -0,010. Esses resultados inexpressivos revelam a baixa consistência interna dos construtos I e V, e a baixa fidedignidade dos dados obtidos com o modelo empírico comum aos Índices e Sistemas de Indicadores de Desenvolvimento Social; não sendo considerados, portanto, consistentes por este método de validação.

Os demais construtos II. Sistema de Produção (Trabalho e Rendimento); III. Longevidade (Saúde) e IV. Socialização (Educação) apresentam valores do Alfa de Cronbach dentro dos parâmetros considerados aceitáveis (HAIR, 1998), o que demonstra uma boa consistência interna dos indicadores que constituem os respectivos construtos.

Da mesma maneira, os valores do KMO para os construtos II, III e IV mostram um índice aceitável para a análise fatorial. O mesmo ocorre com o teste de esfericidade de Barlett, que indica a adequação da análise fatorial à estrutura de dados. A variância explicada dos três construtos pelos itens é significativa para os referidos construtos.

O resultado da função Associação do KDD, proposta e aplicada nesta seção, oferece indícios, por meio das medidas Alfa de Cronbach, KMO e esfericidade de Barlett, para se questionar a existência de hierarquias de níveis mutuamente independentes e lineares nas relações entre os indicadores empregados nos respectivos Índices e Sistemas de Indicadores de Desenvolvimento Social estudados. Faz-se necessário, portanto, o emprego da Análise Fatorial exploratória para determinar o número mínimo de construtos mutuamente independentes e lineares nas relações entre os 15 indicadores necessários para representar o fenômeno Desenvolvimento Social.

7.3.2 Análise Fatorial dos Indicadores associados o termo teórico Desenvolvimento Social

Diante dos resultados encontrados na análise individual dos construtos realizada na sub-seção anterior, optou-se por empregar a função associação para identificar o número de fatores (construtos) que permita descrever o fenômeno Desenvolvimento Social de forma mais simplificada, foram empregadas as técnicas baseadas em Análise de Componentes Principais (ACP) e Análise Fatorial (AF). Estas técnicas são amplamente utilizadas para testar modelos de múltiplos indicadores (HAIR et al. , 1992).

A Análise de Componentes Principais, aplicada aos indicadores selecionados, permite identificar, entre as 15 variáveis originais (indicadores), o subconjunto de indicadores correlacionados com os cinco construtos associados ao Desenvolvimento Social, contribuindo, desta forma, para a verificação dos construtos selecionados – Dinâmica Populacional, Sistema de Produção, Longevidade, Socialização e Segurança –, conforme o modelo conceitual inicialmente proposto no Capítulo 4.

No contexto da etapa de Associação, como é comum nos processos de *KDD*, foi necessário reduzir o número de variáveis (indicadores) através do método de redução de dimensionalidades Análise Fatorial (AF) via ACP. A AF foi empregada com o objetivo de identificar o menor número de fatores comuns (construtos) através da análise das inter-

relações entre os 15 indicadores sociais, pertencentes ao modelo hipotético elaborado para representar o termo teórico Desenvolvimento Social encontrado nos Índices e Sistemas de Indicadores brasileiros.

Com o objetivo de testar a conveniência da técnica fatorial para identificar o menor número de construtos explicativos, conforme discutido no Capítulo 6, inicialmente os dados foram submetidos aos testes de esfericidade de *Bartlett* e, em seguida, foi calculada a medida de *Kaiser-Meyer-Olkin* para todos os 5.483 municípios brasileiros. Os resultados encontrados demonstraram que a Análise Fatorial pode ser utilizada na mineração dos dados. O resultado da medida *KMO* foi 0,882, o que pode ser considerado como muito bom, já o teste de esfericidade de *Bartlett* foi aceito com significância zero para um qui-quadrado 63.027,92 com 105 graus de liberdade. Os resultados encontrados demonstram que a Análise Fatorial é apropriada.

Basicamente, esta etapa de Análise Fatorial dos 15 indicadores associados o termo teórico Desenvolvimento Social, contribuiu para a seleção dos indicadores mais relevantes para o estudo e posterior identificação do modelo empírico comum às ferramentas brasileiras de mensuração do desenvolvimento social. Resumidamente, esta etapa consistiu de três passos: 1) Seleção dos indicadores significantes; 2) Determinação do número de dimensões de análise selecionadas 3) Rotação dos fatores para contribuir com melhor interpretação do modelo empírico identificado.

A partir dos 15 indicadores originais selecionados para análise, procedeu-se à verificação inicial para exclusão daqueles indicadores em que a comunalidade⁸⁹ explicada fosse menor que 0,50 (HAIR et al., 1992), procedimento que permitiu manter, no modelo empírico, apenas os indicadores muito significantes para explicar o fenômeno Desenvolvimento Social.

Como pode ser observado na Tabela 6, seis indicadores entre os 15 indicadores previamente selecionados possuem comunalidade inferior a 0,500. Este primeiro critério oferece a possibilidade de identificar os indicadores que são excluídos do modelo empírico. De acordo com o resultado apresentado, entre os indicadores de Desenvolvimento Social, selecionados e justificados, discutidos anteriormente no Capítulo 6 desta tese, foram excluídos do modelo empírico seis indicadores: Índice de Gini da distribuição do rendimento (3); Taxa de desocupação (4); Prevalência de desnutrição total (9); Doenças relacionadas ao saneamento

⁸⁹ Comunalidade pode ser explicada com sendo a proporção da variância total de cada variável que é explicada pelo conjunto de fatores comuns.

ambiental inadequado (10); Coeficiente de mortalidade por homicídios (14) e Coeficiente de mortalidade por acidentes de transporte (15).

Tabela 6 – Correlação dos 15 indicadores de desenvolvimento social com as dimensões originais – Brasil, 2000

INDICADOR	COMUNALIDADE	Dimensões Originais		
		1	2	3
1 - Taxa de crescimento da população	0,678	0,125	-0,605	0,545
2 - Taxa de fecundidade	0,626	-0,723	-0,235	0,218
3 - Índice de Gini da distribuição do rendimento	0,483	0,381	0,571	0,106
4 - Taxa de desocupação	0,390	0,522	0,342	0,031
5 - Rendimento familiar per capita	0,907	0,942	-0,142	-0,013
6 - Rendimento médio mensal	0,871	0,913	-0,194	0,014
7 - Esperança de vida ao nascer	0,777	0,866	0,157	0,054
8 - Taxa de mortalidade infantil	0,804	0,866	0,130	0,052
9 - Prevalência de desnutrição total	0,283	0,101	0,345	0,391
10 - Doenças relacionadas ao saneamento ambiental inadequado	0,565	-0,121	0,377	0,639
11 - Taxa de escolarização	0,653	0,776	-0,139	-0,178
12 - Taxa de alfabetização	0,846	0,919	-0,008	0,034
13 - Escolaridade	0,867	0,915	-0,167	-0,035
14 - Coeficiente de mortalidade por homicídios	0,047	-0,032	-0,181	0,115
15 - Coeficiente de mortalidade por acidentes de transporte	0,395	-0,456	0,344	-0,263

Fonte: Elaboração própria

Como não foi encontrada correspondência satisfatória entre as relações postuladas no modelo empírico inicial e as correlações observadas na base empírica, o modelo empírico proposto (após a exclusão dos seis indicadores), foi modificado e novamente submetido a testes de ajuste. Tal processo prosseguiu até que se chegasse à configuração que, por se adaptar bem aos dados empíricos, permite representar o modelo empírico comum às ferramentas de mensuração.

O resultado obtido nesta primeira aproximação do modelo empírico corrobora a seleção dos construtos, realizada durante a análise das metodologias empregadas na elaboração de 46 Sistemas ou Índices nacionais que contemplam a dimensão Social do Desenvolvimento, já apresentada no Capítulo 4 desta tese.

Como pode ser observado na Tabela 7, nenhum dos nove indicadores selecionados nesta última etapa, possui comunalidade inferior a 0,5. De acordo com o

resultado encontrado, todos os indicadores selecionados e justificados através do referencial sobre Desenvolvimento Social, discutidos anteriormente nos Capítulos 4 e 6, serão mantidos no modelo. Este resultado corrobora a seleção realizada durante a revisão da literatura desta tese.

Tabela 7 – Correlação dos nove indicadores de desenvolvimento social com as dimensões originais – Brasil, 2000

INDICADOR	COMUNALIDADE	Dimensões Originais	
		1	2
1 - Taxa de crescimento da população	0,934	0,131	0,958
2 - Taxa de fecundidade	0,646	-0,722	0,353
5 - Rendimento familiar per capita	0,905	0,947	0,088
6 - Rendimento médio mensal	0,865	0,918	0,146
7 - Esperança de vida ao nascer	0,761	0,866	-0,105
8 - Taxa de mortalidade infantil	0,796	0,866	-0,102
11 - Taxa de escolarização	0,623	0,789	-0,033
12 - Taxa de alfabetização	0,851	0,922	0,032
13 - Escolaridade	0,875	0,931	0,096

Fonte: Elaboração própria

Diante dos nove indicadores validados por meio da seleção dos mais significantes, investiu-se na verificação do número de construtos (fatores), através da análise dos autovalores⁹⁰ onde o valor fosse superior a 1,0 (HAIR et al., 1992), mantendo-se apenas os fatores que explicaram mais do que um indicador poderia explicar isoladamente.

A Tabela 8 traz os resultados dos autovalores para os indicadores e todos os possíveis construtos do Desenvolvimento Social. Após a extração pelo método da Análise das Componentes Principais, todos os indicadores apresentaram comunalidade maior que 50%, ratificando a presença de nove indicadores no modelo empírico do Desenvolvimento Social. Ao analisar a Tabela 8, percebe-se que existem dois construtos em que o autovalor é maior que 1, o que significa que dois construtos ou fatores são suficientes para reunir os nove indicadores selecionados e, a partir da terceira dimensão em diante, o autovalor passa a ser menor que 1. Como pode ser observado na Tabela 8, o terceiro construto possui autovalor 0,604 e o quarto construto o autovalor é 0,444 e assim distribuídos em ordem decrescente.

⁹⁰ Valores próprios da matriz de correlação (em inglês , *eigenvalue*).

Tabela 8 – Autovalores e percentagens de variância explicada pelos construtos, com base em 9 indicadores

Construtos / Fatores	Autovalores	Percentagem de Variância	Percentagem de Variância Acumulada
1	6,152	68,36	68,36
2	1,104	12,27	80,63
3	0,604	6,72	87,34
4	0,444	4,94	92,28
5	0,300	3,33	95,61
6	0,199	2,21	97,82
7	0,086	0,95	98,77
8	0,062	0,68	99,45
9	0,049	0,55	100,00

Fonte: Elaboração própria

Utilizando-se o critério adotado por vários autores, entre eles, Johnson e Wichern (1998) e Hair et al. (1992), deve-se aceitar apenas dois construtos para compor o modelo empírico comum do Desenvolvimento Social, visto que o primeiro e o segundo construto possuem autovalores, 6,152 e 1,104, respectivamente, ambos superiores à unidade preconizada.

Diante do resultado encontrado, o termo teórico Desenvolvimento Social pode ser representado por apenas dois construtos (construto A e construto B) e não em cinco construtos (I. Dinâmica Populacional; II. Sistema de Produção; III. Longevidade; IV. Socialização e V. Segurança), como anteriormente proposto.

Determinados os dois construtos (construto A e construto B) significativos para a análise pretendida e diante do grupo de nove indicadores mais significantes para compor o modelo empírico do Desenvolvimento Social, foram calculados novamente os autovalores, a inércia total acumulada, a matriz com as cargas fatoriais e as comunalidades para cada indicador e a matriz rotacionada (Tabela 7) pelo método Varimax⁹¹ (HAIR et al., 1992). Este procedimento auxiliou na identificação e associação entre os indicadores com cargas fatoriais altas e seus respectivos construtos (componentes de cada fator).

⁹¹ A rotação ortogonal permite que os coeficientes de correlação entre os indicadores e as dimensões fiquem o mais próximo de zero, 1 ou -1, contribuindo para interpretação do modelo.

A interpretação dos construtos foi feita a partir da análise dos indicadores com maior carga⁹² no fator. Este procedimento foi utilizado porque, quando a análise fatorial faz parte de uma matriz de correlação, as cargas fatoriais são os próprios coeficientes de correlação entre os indicadores e seus construtos. Estes coeficientes representam as próprias cargas fatoriais que, por sua vez, revelam quanto um indicador está relacionado a um construto.

A matriz de correlações dos indicadores, com os construtos de cada um dos nove indicadores selecionados e suas respectivas comunalidades, pode ser observada na Tabela 5. Já a Tabela 9 mostra os resultados para os nove indicadores, após a rotação pelo método Varimax, e as correlações entre cada indicador e seus respectivos construtos Rotacionados.

Tabela 9 – Correlação dos indicadores de desenvolvimento com os fatores rotacionados

INDICADOR	COMUNALIDADE	Fatores Rotacionados	
		1	2
1 - Taxa de crescimento da população	0,934	0,068	0,964
2 - Taxa de fecundidade	0,646	-0,744	0,305
5 - Rendimento familiar per capita	0,905	0,939	0,151
6 - Rendimento médio mensal	0,865	0,907	0,207
7 - Esperança de vida ao nascer	0,761	0,871	-0,048
8 - Taxa de mortalidade infantil	0,796	0,891	-0,043
11 - Taxa de escolarização	0,623	0,789	0,019
12 - Taxa de alfabetização	0,851	0,918	0,093
13 - Escolaridade	0,875	0,922	0,157

Fonte: Elaboração própria

Nota: Extração pelo método da Análise das Componentes Principais. Rotação ortogonal pelo método Varimax com normalização Kaiser.

A análise da Tabela 6, por outro lado, também permite observar que os dois construtos selecionados para o modelo empírico podem explicar 80,63% da variância total dos nove indicadores retidos. Estas afirmações são verificadas a partir dos resultados dos autovalores e da percentagem da inércia total explicada pelos fatores.

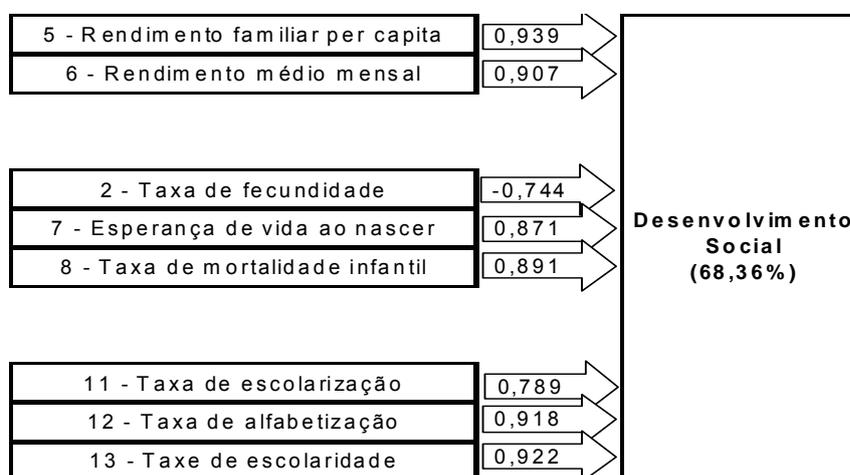
Ao examinar, conjuntamente, as Tabelas 6 e 7, é possível constatar que o construto “A” explica 68,36% da inércia total do conjunto original, o que significa que este é o construto mais explicativo, formado por oito indicadores: 2 - Taxa de fecundidade; 5 - Rendimento familiar per capita; 6 - Rendimento médio mensal; 7 - Esperança de vida ao

⁹² As cargas devem ser analisadas em valores absolutos.

nascer; 8 - Taxa de mortalidade infantil; 11 - Taxa de escolarização; 12 - Taxa de alfabetização; 13 – Escolaridade. O único indicador restante 1 - Taxa de crescimento da população forma o segundo construto “B”, que por sua vez, explica 12,72% da inércia total do conjunto original.

O modelo conceitual do termo teórico Desenvolvimento Social, elaborado a partir da síntese das relações e conceitos identificados na análise de 46 Sistemas de Indicadores ou Índices nacionais, que contemplam a dimensão social do desenvolvimento, apresentado no Capítulo 4, passou por processos de crítica e construção na etapa de Prospecção do KDD, apresentada no Capítulo 6, resultando no modelo hipotético que, por fim, foi testado neste capítulo por meio da mineração dos dados empíricos referentes aos 5.483 municípios brasileiros. Após a aplicação das funções da mineração de dados relativas à estimação, à avaliação de ajuste e às reespecificações do modelo empírico, chegou-se ao resultado apresentado na Figura 16, que pode ser considerado como uma representação aceitável das relações existentes entre as variáveis que compõem o conjunto de indicadores de Desenvolvimento Social estudado.

Figura 16 – Modelo empírico comum do Desenvolvimento Social – Brasil, 2000



Fonte: Elaboração própria

O modelo empírico da Figura 16 apresenta os resultados das análises exploratórias para todos os quinze indicadores selecionados e testados para cada município brasileiro. Indica, portanto, os atributos identificados nos índices e sistemas de indicadores estudados, que mais contribuem com processo de mudança social acelerada em busca do bem-estar

humano. O modelo empírico apresentado permite a avaliação da articulação entre os fatores associados, o que propicia uma melhor compreensão do caráter relacional e complexo, composto por diversos fatores do Desenvolvimento Social que se inter-relacionam e se reforçam mutuamente.

Aplicação do *Knowledge Discovey in Databases* (KDD) nas bases de dados públicas trabalhadas, bem como as respectivas verificações de adequação para seu uso, permitiu que se estabelecesse um modelo empírico validado a partir da análise do *scree plot*, e dos critérios de *Kaiser* (que sugere abandonar todos os fatores cujo autovalor seja inferior a 1, ou seja, autovalores maior que 1) e de Harman (que estabelece que cada fator tenha no mínimo 3% de variância explicada). Além destes critérios, utilizou-se o critério $\lambda_1/\lambda_2 > 5$, para verificação da unidimensionalidade dos fatores com base no *eigenvalue* (Tabela 7.6). De acordo com este critério, quando a razão entre o 1º e o 2º *eigenvalue* é maior ou próxima de cinco, significa que há unidimensionalidade no fator. Caso contrário, um segundo fator deve ser considerado na análise como uma 2ª dimensão, ou seja, não há unidimensionalidade no fator.

A unidimensionalidade encontrada indica que apenas um construto pode ser medido pelo conjunto de indicadores que compõem o Índice ou Sistema de Indicador. Entretanto, não se deve confundir consistência interna com dimensionalidade, porque apesar de unidimensionalidade implicar em alta consistência interna, alta consistência interna não implica em unidimensionalidade. Não é apenas possível, mas também fácil obter alta consistência interna em testes que são multidimensionais.

De certa forma, a unidimensionalidade se enquadra nos chamados estudos de estrutura interna (CRONBACH & MEEHL, 1955), nos quais os resultados dos indicadores são relacionados uns com os outros, pois considera as relações entre os indicadores dentro de um Índice ou Sistema de Indicador. Nesse caso específico, o resultado encontrado indica que todos os indicadores medem o mesmo termo teórico. Nesse contexto, a presença da unidimensionalidade significa que os resultados para os indicadores de desenvolvimento, selecionados das 46 formas de mensuração do desenvolvimento social brasileiro, dependeriam do mesmo termo teórico subjacente, podendo denominá-lo de “unifatorial”.

A Figura 15 fornece uma representação pictórica simplificada das relações resultantes das interligações entre os indicadores selecionados para este estudo, obtida por meio do emprego função Associação do KDD. O emprego do KDD pode permitir uma

reflexão sobre as possibilidades de utilização desse poderoso método em estudos futuros que possam verificar a plausibilidade de tais conjecturas.

O modelo empírico validado apresenta elevada consistência interna (alfa de Cronbach de 0,879), é de fácil aplicação e pode ser utilizado principalmente por governos e institutos oficiais de pesquisa para balizar as atuais formas de mensuração existentes. A unidimensionalidade detectada permite a construção de índices e sistemas de indicadores mais simples e mais efetivos para mensuração do desenvolvimento social. Com base nos oito indicadores sociais validados, conclui-se que esta é uma alternativa para investigações empíricas e trabalhos aplicados em municípios, podendo orientar medidas que visem à promoção do bem-estar da sociedade.

A construção abstrata representada na Figura 7.1, elaborada a partir das definições e convenções terminológicas sobre o conceito de Desenvolvimento Social ao longo desta tese, pode representar a realidade de uma maneira muito aproximada através do modelo empírico do Desenvolvimento Social. Conjugada ao referencial teórico sobre os fatores associados ao desenvolvimento social, a abordagem do contexto brasileiro pode ajudar a fundamentar a interpretação substantiva dos achados do modelo empírico.

Através da análise da estrutura interna do corpo de relações que constituem o fenômeno Desenvolvimento Social, este Capítulo oferece importantes possibilidades para as teorias da Sociologia e de outras áreas do conhecimento.

É inquestionável a relevância dessa forma de representação para os estudos sobre a mensuração do Desenvolvimento Social na atualidade. A necessidade de que a realidade social seja apreendida e representada de maneira apropriada coloca-se imperativa para que se possam desenvolver ações efetivas voltadas à promoção do bem-estar humano e ambiental com potencial de sucesso.

Tendo em vista a evolução recente das tecnologias digitais e dos métodos quantitativos de pesquisa social, deve-se pensar na elaboração de modelos empíricos que permitam a realização de uma análise dinâmica desse fenômeno social, feita a partir da identificação e da contemplação dos indicadores de Desenvolvimento Social, da compreensão de como eles se articulam e se inter-relacionam e quais os seus impactos, isolados ou em conjunto, nas condições de bem-estar das pessoas. Por todos os motivos aqui discutidos, avanços nessa área do conhecimento podem se mostrar promissores para subsidiar o monitoramento dos processos de transformações de natureza econômica, política, social e

ambiental, entre outras, pois podem abrir oportunidades de que sejam enxergados aspectos, até então, pouco explorados dessa questão. A seguir, são apresentadas as considerações finais desta tese.

8 CONSIDERAÇÕES FINAIS

O desejo mais profundo da mente, até em suas operações mais complexas, é o paralelo ao sentimento inconsciente do homem diante de seu universo: é uma insistência na familiaridade, um apetite de clareza. Para o homem, compreender o mundo é reduzi-lo ao humano, imprimindo-lhe sua marca. (ALBERT CAMUS, 1940, p.17)

O estudo realizado nesta tese teve como principal objetivo contribuir com o estudo do Desenvolvimento por meio da aplicação do método *Knowledge Discovery in Databases* (KDD) em bases de dados públicas, para identificar o modelo empírico comum aos instrumentos brasileiros de mensuração do desenvolvimento social. Neste contexto, o desenvolvimento social do Brasil foi mensurado na maior escala possível, segundo os dados disponíveis em bases de dados públicas. Na “jazida de dados”, cada município correspondeu à unidade de observação, e as informações nela contidas corresponderam ao conjunto de indicadores elaborados com base no referencial teórico levantado sobre o desenvolvimento social.

Há dois fatores principais que diferenciam esta tese dos estudos de mensuração do desenvolvimento social encontrados na literatura: o primeiro é a forma como o desenvolvimento social é modelado. Na modelagem proposta por este trabalho não é utilizado um indicador simples, a mensuração é feita de maneira multidimensional através dos indicadores sociais de resultado que representam as características dos cidadãos relacionadas aos municípios onde residem. Esta nova abordagem permitiu contribuir com a modelagem empírica do fenômeno desenvolvimento social, a partir dos indicadores selecionados na literatura e dos dados públicos disponíveis para o Brasil para o ano de 2000⁹³.

O segundo fator diferenciador foi a modelagem e o detalhamento das fases típicas de um processo de *Knowledge Discovery in Databases* destinado a elaboração e validação de instrumento de mensuração de fenômenos sociais a partir de bases de dados públicas. Diversos trabalhos discutem o conceito e mesmo o resultado da aplicação do KDD. Raros falam, mais detalhadamente, sobre as etapas e passos do processo e nenhum dos trabalhos revistos descreve ou analisa a totalidade do processo no nível de detalhe aqui apresentado. Esse detalhamento permitiu também identificar limites e possibilidades do emprego do KDD

⁹³ Ano do último censo demográfico realizado pelo IBGE.

na modelagem do desenvolvimento social, além de oferecer a possibilidade de novos estudos a partir da aferição dos instrumentos de mensuração, aqui apresentados.

Também é incipiente na literatura a avaliação das propriedades, individuais e conjuntas, dos indicadores e dos construtos do Desenvolvimento Social por meio de análise conjugada de considerações substantivas e de resultados empíricos em bases de dados públicas no Brasil, mas certamente, em um curto espaço de tempo, esta prática estará difundida. Diante da escassez de referências na literatura, espera-se que esta tese possa auxiliar aos novos prospectores e mineradores de dados a descobrirem conhecimento relevante em bases públicas e, forneça maior possibilidade de explicação e contribua para enriquecer as formas de racionalização do Desenvolvimento Social.

Neste capítulo, são apresentadas as considerações finais desta pesquisa. Inicialmente, são discutidas algumas limitações e apresentadas as principais contribuições desta tese e, por fim, são tecidas algumas sugestões para futuros estudos do tema proposto.

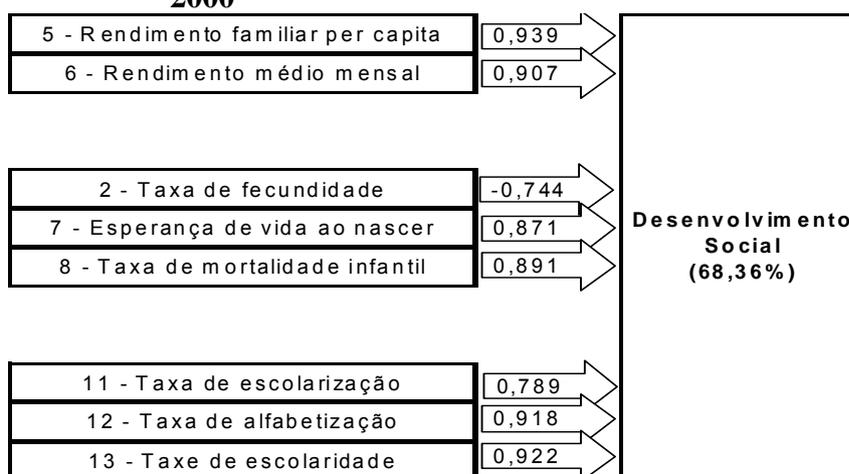
Foi evidenciado que um bom instrumento de mensuração do desenvolvimento raramente é fácil de ser entendido em sua construção. É necessário que os cientistas das áreas social, econômica e ambiental, apoiados na infra-estrutura tecnológica hoje amplamente disponível nas suas respectivas áreas de conhecimento, assumam a responsabilidade da elaboração de construtos explícitos simultaneamente mais ecléticos e rigorosos em termos de sua base metodológica, assim como de um maior cuidado com o uso e a comunicação dos resultados obtidos pelas métricas existentes.

Com o emprego do protocolo de métodos *Knowledge Discovery in Databases* é possível determinar modelos alternativos sobre as relações de influência para o fenômeno Desenvolvimento Social. Diante da variedade de relações possíveis entre os indicadores selecionados pode-se admitir que as relações entre estes indicadores possam ser teoricamente diversas, o que conseqüentemente proporciona a formulação de modelos alternativos que possibilitam a proposição de explicações conjecturais sobre as divergências encontradas.

Espera-se que quanto maior for o nível especulativo das teorias, maiores serão as oportunidades de realizar importantes avanços na formulação de novas teorias do desenvolvimento social. O intuito não é o de promover uma proliferação das teorias especulativas, na medida em que aquelas que representam descrições inadequadas do mundo podem ser refutadas em função do resultado da observação ou de outras provas.

Antes do início deste estudo, foi definido o objetivo geral: contribuir com o estudo do Desenvolvimento por meio da aplicação do método *Knowledge Discovery in Database* (KDD) em bases de dados públicas, para identificar o modelo empírico comum aos instrumentos brasileiros de mensuração do desenvolvimento social (Figura 17). Ao analisar os resultados atingidos, constata-se que objetivo geral proposto foi plenamente satisfeito.

Figura 17 – Modelo empírico comum do Desenvolvimento Social – Brasil, 2000



Fonte: Elaboração própria

Para atender a esse objetivo geral, foram identificados quatro objetivos específicos. O primeiro deles procurou analisar a trajetória do conceito de desenvolvimento, buscando compreender a sua utilização nas principais formas de mensuração encontradas nos índices e sistemas de indicadores de desenvolvimento utilizados no Brasil e no mundo, para isso foram empreendidos estudos de aspectos teóricos apresentados no Capítulo 2 - Desenvolvimento: as trajetórias de um conceito e aspectos metodológicos no Capítulo 3 – Aspectos teórico-metodológicos da mensuração do desenvolvimento social.

O estudo realizado no Capítulo 3 também contribuiu para atender ao segundo objetivo específico – Estruturar os conceitos e metodologias para construção de índices e sistemas de indicadores sociais que possam mensurar conceitos relevantes de um sistema social. No Capítulo 3, foram estudados os aspectos teórico-metodológicos da mensuração de fenômenos sociais e sua conexão com as principais propostas de formulação de Índices Sociais disponíveis na literatura, a fim de subsidiar a construção de categorias analíticas para o estudo dos instrumentos de medição do desenvolvimento por meio de um levantamento bibliográfico e documental dos mais conhecidos Índices e Sistemas de Indicadores de

Desenvolvimento disponíveis na literatura nacional e internacional, realizado no Capítulo 4, para subsidiar a identificação e representação gráfica do termo teórico, dos respectivos construtos e dos principais conceitos empregados nos instrumentos analisados para elaboração do modelo conceitual do desenvolvimento social no Brasil.

Por fim, o objetivo específico quatro – Aplicar o KDD para confrontar o modelo conceitual do desenvolvimento social sobre os dados disponibilizados pelo IBGE e Ministério da Saúde para 5.483 municípios brasileiros, a fim de detectar e estabelecer o modelo empírico comum aos índices e sistemas de indicadores de desenvolvimento social brasileiros – foi atingido com a aplicação do KDD e suas respectivas funções Associação e Sumarização, ambas apresentadas no Capítulo 6 – o protocolo *Knowledge Discovery in Database* (KDD) e os modelos empíricos e no Capítulo 7 – Modelos empíricos do desenvolvimento social. O quadro 8.1 apresenta o resumo da abordagem dos capítulos 2; 4; 6 e 7 da tese.

Este estudo também trouxe outras contribuições relevantes ao conhecimento teórico e empírico sobre os fatores que influenciam no Desenvolvimento Social, além de contribuições metodológicas relacionadas com a modelização e validação de instrumentos de mensuração de fenômenos sociais.

A seguir são apresentadas os limites e as possibilidades identificadas na aplicação do *Knowledge Discovery in Database* (KDD).

Quadro 4 – Quadro resumo da abordagem dos capítulos 2; 4; 6 e 7 da tese

Capítulo	Objetivo	Dimensão	Modelo	Unidade observacional	Método	Fonte	Principal Contribuição
2 DESENVOLVIMENTO: AS TRAJETÓRIAS DE UM CONCEITO	apresentar aspectos relativos às interpretações dominantes do conceito Desenvolvimento procurando agrupá-los segundo a similitude de suas abordagens, visando subsidiar o levantamento de questões que envolvem a instrumentalização desses conceitos e suas formas de mensuração.	Termo Teórico	Teórico	Teorias do Desenvolvimento	Revisão bibliográfica	Bibliografia sobre Desenvolvimento	identificar as principais diferenças e similitudes nas abordagens sobre desenvolvimento e reuni-las por correntes ou escolas de pensamento. analisar o termo teórico desenvolvimento e seus desdobramentos.
4 ANÁLISE DOS ÍNDICES E SISTEMAS DE INDICADORES DE DESENVOLVIMENTO	apresentar e analisar as principais formas de mensuração encontradas nos Índices e Sistemas de Indicadores de Desenvolvimento utilizados no Brasil e no mundo nos últimos 45 anos. Para consecução do objetivo proposto para este capítulo, foram definidos os seguintes objetivos específicos: 1. Levantar, por meio de pesquisa bibliográfica e documental, os mais conhecidos Índices e	Conceitos	Conceitual	Instrumentos de mensuração	Infometria	Documentos dos Índices e Sistemas de Indicadores	identificar e representar graficamente o termo teórico, os respectivos construtos e conceitos empregados nos instrumentos analisados. Elaborar o modelo conceitual.
6 O PROTOCOLO KNOWLEDGE DATA DISCOVERY E OS MODELOS HIPOTÉTICOS	descrever os passos do KDD empregados na elaboração do modelo hipotético do Desenvolvimento Social que é testado no Capítulo 7.	Indicadores	Hipotético	Bases de Dados	KDD - Prospecção	IBGE; MS; MJ	foi elaborado o modelo gráfico hipotético das relações entre os indicadores para esta pesquisa, a partir dos conceitos levantados com o estudo das principais propostas de Índices e Sistemas de Indicadores de desenvolvimento Social no Brasil, disponíveis na literatura.
7 MODELOS EMPÍRICOS DO DESENVOLVIMENTO SOCIAL	estabelecer um modelo empírico, com base nos dados disponibilizados pelo IBGE e Ministério da Saúde para 5.483 municípios brasileiros, a fim de detectar e estabelecer conjecturas explicativas sobre este fenômeno no Brasil.	Variáveis	Empírico	Jazida de Dados	KDD - Mineração	Elaboração própria	modelo empírico apresentado permite a avaliação da articulação entre os fatores associados, o que propicia uma melhor compreensão do caráter relacional e complexo, composto por diversos fatores do Desenvolvimento Social que se inter-relacionam e se reforçam mutuamente.

Fonte: Elaboração própria

8.1 LIMITES E POSSIBILIDADES DO PROTOCOLO DE MÉTODOS KDD

Embora as possibilidades sejam muitas, podem-se encontrar aplicações do KDD nas áreas de finanças e detecção de fraudes; saúde; marketing e vendas; jogos e esportes; recursos humanos; serviços de infra-estrutura: energia, água, transporte, telecomunicação, Internet; ciências; governo e muitas outras.

No âmbito da construção da mensuração do Desenvolvimento Social, a literatura traz poucas referências ao emprego do KDD. O processo basicamente tem sido empregado no setor governamental de arrecadação de impostos, visando aumentar a receita e detectar fraudes do tipo transferências de fundos para “lavagem” de dinheiro e sonegação de impostos, como no caso clássico da Angoss Software Corporation em que o governo dos Estados Unidos utilizou a ferramenta *Knowledge Seeker* em um sistema detector de fraudes fiscais.

Nesta pesquisa, diversas possibilidades do KDD foram identificadas:

- i) O KDD mostra que é possível realizar estudos com múltiplas fontes de informação. Diante da complexidade do fenômeno desenvolvimento social e da situação atual das estatísticas oficiais e registros administrativos disponíveis, nenhuma fonte isolada seria suficiente;
- ii) Diferentemente dos métodos empregados nos principais Sistemas e Índices de Desenvolvimento Social estudados nesta tese, o KDD permite identificar com clareza as fontes de informação que estão sendo usadas e quais seus pontos fortes e fracos. É importante destacar que omissões como essas, observadas em alguns índices estudados nesta pesquisa, só contribuem para minar a credibilidade e a possibilidade de reprodução do trabalho de mensuração;
- iii) Como foi visto, grande parte dos Índices de Desenvolvimento Social estudados recorre a artifícios matemáticos para apresentar uma exatidão da medida maior do que realmente as escalas utilizadas para coletar os dados. Dentre os Índices e Sistemas de Indicadores analisados, em muitos casos, viu-se uma escala ordinal (menor exatidão) transformada equivocadamente numa escala de nível intervalar (maior exatidão). O KDD permite que sejam observadas as medidas apropriadas para cada tipo de escala de medida empregada nas variáveis disponíveis, permitindo atender aos pressupostos dos

algoritmos empregados e, conseqüentemente, uma maior aproximação do fenômeno;

- iv) O aumento da clareza dos procedimentos proporcionados pelo KDD permite melhorar a mensuração em todos os níveis de qualidade da informação. A abordagem multidimensional do KDD permite captar a complexidade dos construtos e dos conceitos de uma maneira mais adequada que a das medidas existentes;
- v) Finalmente, a mensuração deve ser completada pela etapa de disseminação dos resultados para análise ou avaliação política, seja pelo governo, sociedade civil ou entidades do terceiro setor.

Diante da ubiqüidade do KDD, é mais difícil identificar setores que não permitem sua aplicação do que aqueles que dela podem beneficiar-se, entretanto existem algumas barreiras que podem inviabilizar seu uso. Entre as principais, podem ser destacadas: necessidade de elevado conhecimento sobre o domínio a ser estudado; necessidade de grandes volumes de dados armazenados em poderosos servidores; e a complexidade das ferramentas de *Data Mining (DM)*, além do desafio de preparar os dados para mineração. Dificilmente será possível encontrar uma base de dados já filtrada, limpa, integrada com dicionário de dados e elaborada com tamanho rigor técnico-científico necessários ao processo KDD; por fim e não menos importante, a dificuldade de se obter uma análise custo/benefício bem fundamentada do projeto de implantação do KDD, assim como do DM.

Também foram identificadas algumas limites do KDD:

Embora o estudo desenvolvido nesta tese tenha focado em todos os municípios de 168.370.893 habitantes, distribuídos nos 44.795.101 domicílios particulares permanentes existentes nos 5.561 municípios brasileiros, a jazida de dados “minerada” foi relativamente pequena e desatualizada (embora a mais recente disponível), contemplando apenas 15 indicadores de desenvolvimento social de 5.483 municípios. Apesar de ter sido feita a opção por “minerar” os dados do IBGE e do Ministério da Saúde e não por “minerar” um *Data Warehouse* tradicional, foi possível delimitar uma base de dados pública (jazida de dados) de porte e relevância e cumprir as principais etapas do KDD propostas pelos principais autores da área.

As tarefas de modelização e de elaboração de instrumentos de mensuração envolvem conhecimento para reconhecer e identificar a informação, assim sendo, a participação de equipes multidisciplinares é fundamental para o sucesso do processo. Além disso, o apoio computacional foi imprescindível devido à complexidade dos algoritmos e do volume de dados trabalhado. O uso de pacotes computacionais adequados a cada necessidade torna-se uma questão crucial.

Ficou evidenciado, durante a pesquisa, que as principais dificuldades na elaboração de instrumentos de mensuração de fenômenos sociais estão atreladas à definição de um modelo hipotético adequado e à seleção dos respectivos indicadores disponíveis em bases de dados públicas para caracterizar o fenômeno a ser estudado. Tanto a etapa de Prospecção, como a etapa de Mineração de Dados (mais complexa), demonstraram necessitar de conhecimento profundo da teoria e dos conceitos disponíveis como da técnica escolhida para mineração dos dados. Nesta tese, caracterizou-se a etapa de Prospecção como sendo de responsabilidade de um chamado “especialista de domínio” e a etapa de Mineração ficou sob o encargo do aqui chamado “especialista em dados”.

A “jazida de dados” construída para esta pesquisa contemplou apenas os dados do Instituto Brasileiro de Geografia e Estatística e Ministério da Saúde de todos os 5.561 municípios brasileiros. Entretanto, esta “jazida”, poderia ter sido enriquecida com outras bases de dados públicas caso a escala de estudo fosse menor, estadual ou nacional, por exemplo. Este procedimento favoreceria a análise de outras dimensões do desenvolvimento social estudado.

8.2 PRINCIPAIS CONTRIBUIÇÕES

Entre as principais contribuições apresentadas por esta tese, destacam-se as seguintes:

- a) O estudo demonstrou a possibilidade de utilização das informações de Bases de Dados Públicas, especialmente do IBGE, do Ministério da Justiça e do Ministério da Saúde, como ferramenta importante na identificação dos modelos hipotético e empírico do desenvolvimento social para os municípios brasileiros. A utilização do KDD e a facilidade e rapidez do acesso a dados

secundários potencializam o seu uso enquanto instrumento de produção de conhecimento científico.

- b) A modelagem do fenômeno Desenvolvimento Social, objeto desta tese, mostrou-se adequada à proposta do trabalho, permitindo uma razoável aproximação, respeitando a escala escolhida. Além disso, a confrontação do modelo hipotético com os dados disponíveis na jazida de dados, resultou na identificação do modelo empírico validado comum às ferramentas brasileiras de mensuração do desenvolvimento social, por meio da aplicação do método KDD em bases de dados públicas. Espera-se que outros estudos semelhantes e mais sofisticados venham a ser desenvolvidos em futuro próximo.
- c) O uso do diagrama de caminhos para representar o modelo empírico oferece uma análise ampla e sistêmica do construto desenvolvimento social existente nos índices estudados e poderá ser útil no desenvolvimento de novos índices na esfera das investigações da realidade social em afinidade com as formulações teóricas mais recentes.
- d) as discussões sobre a natureza conceitual do construto Desenvolvimento Social, bem com o conjunto de testes utilizados nesta pesquisa para testar empiricamente sua operacionalização, foram consideravelmente mais extensos e abrangentes do que já publicado na literatura de Administração Pública.
- e) No Capítulo 5, foi feito um breve histórico dos principais acontecimentos ocorridos antes do surgimento do KDD até a sua utilização no contexto público, histórico este não encontrado em qualquer literatura isolada.
- f) Outra importante contribuição está relacionada à proposta de caracterização e sistematização de etapas para o processo KDD em bases de dados públicas, resultado do processo dialógico de análise dos trabalhos correlatos nas diversas áreas estudadas no processo de elaboração da tese.
- g) Esta proposta traz mais uma contribuição relacionada à demonstração da importância de validar o algoritmo selecionado para processar os dados. Como é sabido, todos os algoritmos possuem limitações e foi fundamental conhecê-

las. Como foi visto no Capítulo 6, o processo de escolha pode ser visto mais como uma “arte” do que como uma ciência.

- h) O uso do KDD abre perspectivas interessantes na produção de conhecimento científico a partir na análise de bases de dados públicos, particularmente no sentido de gerar referências que possibilitem o estudo de índices e sistemas de indicadores de desenvolvimento social. O emprego de métodos de *Data Mining* que possibilitem uma análise simultânea de construtos que caracterizam o objeto de estudo, oferece ao pesquisador uma visão bastante ampla do fenômeno estudado.
- i) Esta pesquisa trouxe uma terceira dimensão ao KDD, que a maioria dos autores aqui citados não apresenta. Com o avanço dos *softwares*, foi possível potencializar o uso das técnicas de mineração de dados, facilitando sobremaneira a análise dos resultados encontrados desde as etapas preliminares à construção do diagrama final para representar os modelos do desenvolvimento social medidos pelos índices e sistemas de indicadores estudados.

A aplicação de KDD em bases de dados públicas é recente, e os limites levantados nesta tese podem ser generalizados em futuros estudos de mesma natureza, aos quais podem ser agregadas novas dimensões do fenômeno desenvolvimento além das aqui propostas.

O sucesso dos resultados deveu-se, fundamentalmente, à compreensão do domínio estudado através do aprofundamento no assunto e da busca por especialistas para melhor compreender o Desenvolvimento Social, da seleção correta das fontes ao treinamento realizado para manejar adequadamente as bases de dados e, principalmente, pelo entendimento dos métodos disponíveis, o que possibilitou a escolha adequada das técnicas empregadas na mensuração de fenômenos complexos. Este último fator também possibilitou reduzir o número de iterações na busca por métodos mais apropriados, entretanto a proposta aqui apresentada pode ser replicada com o acréscimo de novas informações e a utilização de outros modelos.

- j) Por fim, foi demonstrado nesta pesquisa que o KDD, bem empregado em bases de dados públicas, não só pode contribuir para modelização de fenômenos sociais e com o aprimoramento dos métodos empregados na mensuração do desenvolvimento social como também pode ser empregado para outras

dimensões relevantes do desenvolvimento como ambiental, política, econômica etc.

8.3 SUGESTÕES PARA PESQUISAS FUTURAS

Certamente o conhecimento adquirido e acumulado nesta tese permitirá avançar na direção de aprimorar o estudo dos instrumentos de mensuração de fenômenos sociais e efetivamente poderá auxiliar os responsáveis na produção de índices e indicadores de Desenvolvimento Social e principalmente a sociedade civil organizada a acompanhar e fiscalizar as ações governamentais através das informações disponibilizadas em bases de dados públicas.

O principal requisito para viabilizar a aplicação ampla da modelização de fenômenos sociais por meio da aplicação do protocolo KDD está relacionado a três fatores principais: o primeiro reza sobre a necessidade de um conhecimento aprofundado de cada uma das fases do processo de modelização e de extração de conhecimento apresentados nesta tese, o segundo trata da sintaxe (termo teórico, construtos, conceitos e indicadores) e da semântica (relações) do fenômeno estudado e sua representação por meio das informações existentes nas bases de dados e o terceiro e último está relacionado ao conhecimento do domínio a ser estudado.

Como foi visto anteriormente, de maneira geral, a modelização de fenômenos sociais pode ser feita com o apoio do KDD e das técnicas de DM que, podem ser aplicados a qualquer conjunto de dados estruturados, independentemente de estarem armazenados em um Sistema Gerenciador de Banco de Dados. Entretanto, deve-se ressaltar que os *Data Warehouses* solucionam algumas questões inerentes à fase de Prospecção (modelo hipotético) e tornam a fase de Mineração (modelo empírico) mais eficiente.

Com a possibilidade de enriquecer a “jazida de dados” com outras informações, certamente será possível, por exemplo, utilizando a função Regressão do DM, identificar por meio do modelo empírico, qual a contribuição que cada indicador e sua respectivo construto trazem para a dinâmica do Desenvolvimento Social. Assim, as mudanças sintáticas e semânticas promovidas nas componentes do modelo empírico podem explicar as alterações no processo de mudança social em busca de um de bem-estar humano e ambiental. Em outras palavras, são claras as linhas para futuras pesquisas na área.

Entre as várias alternativas existentes, futuros trabalhos de pesquisa poderiam utilizar outras técnicas⁹⁴ (Redes Neurais Artificiais – RNA, Algoritmos Genéticos – AG, Lógica Fuzzy etc.), ao invés da tradicional Análise Estatística utilizada neste trabalho. Deve ser observada a adequabilidade das técnicas a determinados contextos, pois não se pode garantir a priori a técnica que daria o melhor resultado. O importante é escolher a forma de análise que apresente os melhores resultados possíveis, em cada caso, de tal forma que as percepções obtidas sejam as mais consistentes.

Estas questões não foram abordadas neste estudo. E, pela vastidão dos temas envolvidos, justifica-se a iniciativa de realização de pesquisas específicas para sua elucidação. As respostas a serem obtidas aumentarão a capacidade de implementação do protocolo de métodos KDD proposto nesta tese.

Sob o aspecto dos instrumentos de mensuração do desenvolvimento, foi possível constatar que adoção indistinta de índices ou sistemas de indicadores pode ter, não apenas uma função interpretativa, mas também forte influência formadora, podendo contribuir com a eliminação sumária de características endógenas de povos ou regiões e atingir diretamente a diversidade cultural. É claro que antes de se adotar qualquer indicador de desenvolvimento, é fundamental conhecer previamente o conceito e as metodologias adotados em sua construção, no entanto, grande parte dos índices ou sistemas de indicadores (e quase a totalidade daqueles aqui apresentados) é resultado de proxies de renda, indicadores socioeconômicos e ambientais. Ou seja, as métricas mais usuais são freqüentemente uma reunião de elementos de correntes de pensamento distintas e, portanto, parecem ser eventualmente incompatíveis. Esta hipótese poderá ser investigada em detalhe na continuação desta pesquisa.

Observou-se ainda que a visão do desenvolvimento de uma sociedade não pode ficar restrita a de um conceito operatório, puramente descritivo, relacionado às hipóteses induzidas, que culminam, em última análise, em um modelo de padrões culturais eventualmente estranhos às sociedades em que são aplicados.

É importante salientar que, ainda assim, os indicadores de desenvolvimento de mais largo escopo teórico devem ser vistos, não apenas como instrumentos de uso eventual ou figurativo, mas sim como elementos importantes para fundamentar a tomada de decisão na esfera pública. A escolha de um indicador de desenvolvimento para subsidiar decisões governamentais, em última análise, portanto, significa a adoção de um modelo de

⁹⁴ Os apêndices X e V apresentam alguns exemplos de modelagem utilizando outras técnicas de modelagem para o Desenvolvimento Social brasileiro.

desenvolvimento mimético e algumas vezes coercitivo, podendo potencialmente favorecer a consolidação de dependências que por vezes são a própria antítese do desenvolvimento e que podem, desta forma, inibir as transformações estruturais de muitas economias e suas sociedades.

Os índices e sistemas de indicadores, desta forma, prescrevem normas e procedimentos para o que é considerado um percurso normal de desenvolvimento. Conclui-se então que para se estudar e instrumentalizar o desenvolvimento por meio de métricas sintéticas é necessário um olhar crítico, inquisitivo e transdisciplinar. Esse paradigma exige, por sua vez, a construção de categorias analíticas que integrem aportes de diferentes campos de conhecimento.

8.4.1 Perspectivas

No início desta tese, ao se apresentar a proposta de trabalho e delimitar o estudo, foi feita a opção de não aprofundar a discussão de alguns temas que, sem dúvida, pela sua riqueza, dificultariam a convergência em torno do foco central. Entre os principais, destacam-se: Métodos e Técnicas de Investigação Social; Teoria do Conhecimento; Sistemas de Apoio à Decisão; *Data Warehouse*; outras técnicas de *Data Mining* (Redes Neurais Artificiais, Algoritmos Genéticos etc.); *Business Intelligence* e outros.

No que concerne à capacidade de ultrapassar os limites da simples quantificação dos aspectos relativos ao desenvolvimento social, pode-se dizer que os índices e os sistemas de indicadores sociais estudados não satisfazem a esse atributo, uma vez que estão limitados a simples agregação de dados. Os resultados oferecidos por tais métodos são números ou coeficientes derivados de processos de contagem ou de algum outro tipo de quantificação. Essas quantificações são vistas como insuficientes para o estudo do desenvolvimento, devendo ser apenas um estágio inicial das pesquisas. A idéia é que as investigações prossigam rumo a um entendimento mais profundo do fenômeno estudado, fornecendo conhecimentos sobre a identificação e intensidade da influência dos fatores que afetam o bem-estar tanto do homem como do ambiente.

Em função da escassez de recursos públicos, os governos procuram formas de gestão mais eficientes e conduzidas a partir de critérios claros e objetivos, visando uma real melhoria das condições de vida da população. Com o uso do KDD na elaboração e validação

de índices e sistemas de indicadores de desenvolvimento, podem ser desenvolvidos mecanismos de monitoramento das ações governamentais, principalmente no tocante à avaliação de resultados para a sociedade e à eficiência das ações. Outra perspectiva para estudos futuros estaria relacionada à avaliação de projetos e programas e à análise de políticas públicas.

O método KDD surgiu em 1996, tendo, portanto, menos de 15 anos de existência. Apesar de ser uma metodologia relativamente recente, recorre a conceitos e técnicas que remontam ao início do século passado. Entre os principais desafios ao KDD/DM, estão questões como: qualidade dos dados originais, integração de sistemas, condições de pré-processamento, complexo relacionamento entre as variáveis estudadas, dinâmica nos dados e no conhecimento, além da interação homem-máquina e questões éticas na obtenção/utilização do conhecimento.

Além de dados convencionais, as Bases de Dados possibilitam o armazenamento de informações de diversas origens, inclusive aquelas inerentes aos aspectos fisiológicos básicos relativos ao processo de percepção dos seres humanos quanto às mudanças que ocorrem em seu ambiente. O avanço tecnológico possibilitará a fertilização do terreno onde o KDD irá prosperar. O aumento da sua utilização pelas mais diferentes áreas resultará no aperfeiçoamento dos conceitos utilizados. Trata-se de um fenômeno crescente que não será detido e, como seu conceito não é estático, certamente amadurecerá e mudará continuamente.

A crescente disponibilização de informações públicas, índices e sistemas de indicadores, principalmente disponibilizados pelos governos através da Internet, irá propiciar aplicações cada vez mais freqüentes do KDD. O processo KDD, conforme descrito neste trabalho, pode ser considerado um ponto de partida para estudos sobre o desenvolvimento social e diversas outras temáticas que podem ser exploradas em bases de dados públicas.

Os avanços tecnológicos vistos sob essa ótica apresentam contribuições muito positivas. Entretanto deve-se observar a necessidade de preservar a autonomia do pesquisador em detrimento da alternativa tecnológica oferecida pelos Sistemas de Apoio à Decisão e pelo KDD.

Concluindo, a mensuração de fenômenos sociais com o apoio do KDD é um desafio que envolve a participação de diversas áreas do conhecimento. Uma nova abordagem para o processo KDD, servindo-se das contribuições das ciências sociais, certamente irá auxiliar os pesquisadores da área a enfrentar este desafio. Esta tese levantou e estruturou

questões acerca da extração de conhecimento, destacando a importância da participação de equipes multidisciplinares, com objetivos definidos e, principalmente, usuárias responsáveis das técnicas disponíveis de análise dados. Desta forma, os resultados deste estudo demonstram, para os analistas de dados e especialistas do domínio, que a utilização dos instrumentos de mensuração de fenômenos sociais exige uma postura responsável e conhecimento aprofundado de cada etapa do processo de modelagem e validação, bem como dos conceitos e da teoria que estão sendo estudados. Espera-se que este trabalho contribua para uma reflexão acerca da forma com que esses instrumentos de mensuração do desenvolvimento social e as bases de dados públicas vêm sendo utilizadas.

REFERÊNCIAS

- ADRIAANS, P.; ZANTIGE, D. **Data Mining**. Harlow: Addison-Wesley, 1996.
- AIDAR, Tirza. **As causas externas e o perfil de mortalidade da população residente no Município de Campinas entre 1980 e 2000**. Disponível em: <http://www.abep.nepo.unicamp.br/docs/anais/pdf/2002/GT_SAU_ST11_Aidar_texto.pdf> Acesso em: 23 de jul. 2009.
- AGRAWAL, R.; MANNILA, H.; SRIKANT, R.; TOIVONEN, H.; VERKAMO, A. I. Fast discovery of association rules. In: _____. **Advances in knowledge discovery and data mining**. Menlo Park, CA: AAAI/MIT Press, 1996. Chapter 12. p. 307-328.
- ALBUQUERQUE, Roberto Cavalcanti de. **A questão social no Brasil**. Um balanço do século XX. Rio de Janeiro: IPEA, 2004.
- ALMEIDA, J. A. A problemática do desenvolvimento sustentável. In: BECKER, Dinizar F. (Org.). **Desenvolvimento sustentável: necessidade ou possibilidade**. Santa Cruz do Sul: Edunisc, 1999.
- ALMEIDA, Paulo Roberto. **A cláusula social no comércio internacional**. RBCE Comércio Exterior. Rio de Janeiro: Funcex, 1994, nº 40, jul-ago-set, p. 52-60.
- ALTER, Steven. **Information systems: a management perspective**. 2 nd ed. Menlo Park, CA: Benjamin & Cummings, 1996.
- AMARAL, Fernanda Cristina. **Data Mining: técnicas e aplicações para o marketing direto**. São Paulo: Berkeley Brasil, 2001.
- ANAND, Sudhir. **Inequality and poverty in Malasya: measurement and decomposition**. Oxford, UK: Oxford University Press, 1977.
- ÁRIAS, A. R. **Estimativas de indigência e pobreza no Brasil no período 1990-1996: resumo metodológico e resultados**. Brasília: IPEA, 1999.
- BARATA, Rita Barradas; RIBEIRO, Manoel Carlos Sampaio de Almeida. **Relação entre homicídios e indicadores econômicos em São Paulo, Brasil, 1996**. *Rev Panam Salud Publica* [online]. 2000, vol.7, n.2, pp. 118-124.
- BECK, Ulrich. **Die Eröffnung des Welthorizontes: Zur Soziologie der Globalisierung**. Globalisierung. Baden Baden: Nomos, 1997.
- BECKER, B.K. 1993. **A Amazônia pós ECO-92: por um desenvolvimento regional responsável**. In: Bursztyn, Marcel (org.), 1993, p. 129-143.
- BECKERMAN, W., & Bacon, R. **International comparisons of income levels: a suggested new measure**. *Economic Journal*, 76(303), 519-536, 1996.

BENNETT, M. K. (1951). **International disparities in consumption levels**. American Economic Review, 41(4), 632–649.

BOOTH, Andrew D. A. A “law” of occurrences for words of low frequency. **Information and Control**, v. 10, n.4, p.386-393, April 1967.

BOSSSEL, Hartmut. **Indicators for sustainable development**: theory, method, applications. Winnipeg : International Institute for Sustainable Development, 1999.

BRAGA, Gilda Maria. A representação da informação na reconstrução do texto. In **COLÓQUIO: ORGANIZAÇÃO DO CONHECIMENTO E REPRESENTAÇÃO DA INFORMAÇÃO: ENFOQUES E PERSPECTIVAS**. Resumos... Rio de Janeiro: IBICTCNPq, 1996.

BRAGA, T. M.; FREITAS, A. G.; DUARTE, G. S. **Índice de sustentabilidade urbana**. In: 1º Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Ambiente e Sociedade (AMPAS), São Paulo, 2002, Paper. São Paulo: AMPAS (2002).

BRASIL. **Brasil 2002 a sustentabilidade que queremos**. Brasília: Gráfica Positiva, 2002.

BRUYNE, Paul de ; HERMAN, Jaques ; SCHAUTHEETE, Marc de. **Dinâmica da pesquisa em ciências sociais**: os polos da prática metodológica Rio de Janeiro: FRANCISCO ALVES, 1982. 251 p.

BRUYNE, Paul de ; HERMAN, Jaques ; SCHOUTHEETE, Marc de. **Dynamique de la recherche en sciences sociales**: les pôles de la pratique méthodologique. Paris: Presses Universitaires de France, 1974.

BRUYNE, Paul de, et al. **Dinâmica da pesquisa em ciências sociais**: os pólos da prática metodológica. 5. ed. Rio de Janeiro: Francisco Alves, 1991.

BUARQUE, Cristovam. **A desordem do progresso**: o fim da era dos economistas e a construção do futuro. Rio de Janeiro: Paz e Terra, 1990.

BUNGE, M. **La investigación científica**. Barcelona: Ariel, 1969.

BUNGE, M. **La ciencia**: su método y su filosofía. Buenos Aires: Siglo Veinte, 1972

BUNGE, M. **Teoria e realidade**. São Paulo: Perspectiva, 1974.

BUNGE, M. **Exploring the world**. Treatise on Basic Philosophy, v. 5. Dordrecht: D. Reidel, 1983.

BUNGE, M. **Seudociência e ideología**. Madrid: Alianza, 1985.

BUNGE, M. **Sociología de la ciencia**. Buenos Aires: Siglo Veinte, 1993.

BURSZTYN, Marcel (Org.) **Para pensar o desenvolvimento sustentável**. Brasília: IBAMA/ENAP. Brasiliense, 1993

BUSTELO, P. **Teorías contemporâneas del desarrollo económico**. Madrid: Editorial Síntesis. Development: an Analysis of Concepts, Measurement and Indicators BAR, Curitiba, v. 5, n. 2, art. 2, p. 104-124, Apr./June 2008. Disponível em: <www.anpad.org.br/bar> Acesso em: 25 de jun. 2009.

CAMINO, R. de; MÜLLER, S. **Sostenibilidad de la agricultura y los recursos naturales: bases para establecer indicadores**. San José, Costa Rica: IICA, 1993. 134p.

CAMUS, A. **The myth of sisyphus**. Trad. Justin O'Brien. New York: Vintage Books, 1991.

CASTELLS, M. **A era da informação: economia, sociedade e cultura**. São Paulo: Paz e Terra, 2000.

CAVALCANTI, Clóvis. **Meio ambiente, desenvolvimento sustentável e políticas públicas**. São Paulo: Cortez e Fundação Joaquim Nabuco, 1997.

CHÂTELET, François. **Uma história da razão: entrevistas com Émile Noel**. Rio de Janeiro: Jorge Zahar Editor, 1997.

COBB, C. et al. **The genuine progress indicator: summary of data and methodology**. San Francisco: Redefining Progress, 1995.

CRESWELL, John W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. trad. Luciana de Oliveira da Rocha. 2. ed. Porto Alegre: Artmed, 2007.

CRONBACH, Lee J. **Coefficient Alpha and the Internal Structure of tests**. University of Illinois, *Psychometria-Vol.*, 16, N.3 September, 1951. *Essentials of Psychological Testing*. 4th ed. New York: Harper & Row, 1984

CUSTANCE J., HILLER H. **Statistical issues in developing indicators of sustainable development**. In *Journal of the Royal Statistical Society*, 1998, 161/3, pp.281-302.

DALY, H.; GAYO, D. **Significado, conceptualización y procedimientos operativos del desarrollo sostenible: posibilidades de aplicación a la agricultura**. In: MARÍN, A. C. (Org.). *Agricultura y Desarrollo Sostenible*. Madrid: Ministerio da Agricultura e Pescas, 1995, p.19-38.

DAROLT, M.R. **As dimensões da sustentabilidade: um estudo da agricultura orgânica na região metropolitana de Curitiba-PR**. Curitiba, 2000. Tese (Doutorado em Meio Ambiente e Desenvolvimento) – Universidade Federal do Paraná, Paraná.

DIENER, E. **A value based index for measuring national quality of life**. *Social Indicators Research*, 36(2), 107–127, 1995.

DUCHET, Michèle. **Le partage des savoirs: discours historique, discours ethnologique**. Paris: Editions La Découverte, 1984.

EPSTEIN, Isaac. **Revoluções Científicas**. São Paulo: Ática, 1988. p. 40.

ESPÍRITO-SANTO, P.S. M. F. OLIVEIRA, P. T. RIBEIRO, D. F.. **O conceito de desenvolvimento social sob a ótica do pensamento complexo**. Anais do 4º Congresso Brasileiro de Sistemas – Centro Universitário de Franca Uni-FACEF – 29 e 30 de outubro de 2008.

ESTES, R. J. **The social progress of nations**. New York: Praeger, 1984.

FAORO, Raymundo. **A questão nacional: a modernização**. In: Estudos Avançados, Jan./Apr. 1992, vol.6, no.14, p.7-22.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **Advances in knowledge discovery and data mining**. Menlo Park, CA: AAAI/MIT Press, 1996.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**. New York, ACM Press. v. 39, n. 11, p. 27-34, Nov. 1996.

FERREIRA, F. H. G., LANJOUW, P., & Neri, M. C. (2003). **A robust poverty profile for Brazil using multiple data sources**. Revista Brasileira de Economia, 57(1), 59–92.

FERREIRA, Francisco H. G. **Os determinantes da igualdade de renda no Brasil: lutas de classes ou heterogeneidade educacional?** Disponível em: <<http://www.econ.puc-rio.br/pdf/td415.pdf>> Acesso em 22 de jul. 2009.

FOSTER, J., GREER, J; THORBECKE, E. **A class of decomposable poverty measures**. Econometrica, Ann Arbor, MI, v. 52, n. 3, p. 761-766, May 1984.

FURTADO, C. **O mito do desenvolvimento econômico**. São Paulo: Paz e Terra, 1974.

FRANCA, Elisabeth et al. **Associação entre fatores sócio-econômicos e mortalidade infantil por diarreia, pneumonia e desnutrição em região metropolitana do Sudeste do Brasil: um estudo caso-controle**. *Cad. Saúde Pública* [online]. 2001, vol.17, n.6, pp. 1437-1447.

FUNDAÇÃO DE ECONOMIA E ESTATÍSTICA. **Índice de desenvolvimento socioeconômico (IDESE)**. Porto Alegre: FEE, 2003.

FUNDAÇÃO GETÚLIO VARGAS. **Mapa do fim da fome**. Rio de Janeiro: Author, 2001.

FURTADO, Celso. **Introdução ao desenvolvimento**. Enfoque Histórico-Estrutural. Rio de Janeiro: Paz e Terra, 2000.

GAWRYSZEWSKI, Vilma Pinheiro; COSTA, Luciana Scarlazzari. **Homicídios e desigualdades sociais no município de São Paulo**. *Rev. Saúde Pública* [online]. 2005, vol.39, n.2, pp. 191-197.

GINSBERG, N., OSBORN, J.; BLANK, G. **Geographic perspectives on the wealth of nations**. Chicago: University of Chicago, 1986.

GOMES, Jaime de O.; SANTO, Augusto H. **Mortalidade infantil em município da região Centro-Oeste Paulista, Brasil, 1990 a 1992.** *Rev. Saúde Pública* [online]. 1997, vol.31, n.4, pp. 330-341.

GWARTNEY, J., LAWSON, R.; BLOCK, W. **Economic freedom of the world (1975–1995).** Johannesburg: Free Market Foundation, 1996.

HAYES, Bob E. **Measuring Customer Satisfaction: Development and Use of Questionnaires.** Milwaukee, Wisconsin: ASQC Quality Press, 1992.

HARBISON, F.; MYERS, C. A. (1964). **Education, manpower, and economic growth.** New York: McGraw-Hill, 1964.

HART, R. **Summary, conclusions and lesson learned from the SANRM/ INFORUM Eletronic Conference on Indicators of Sustainability.** Retrieved May 14, 2006. Disponível em: <<http://tdg.uoguelph.ca/www/FSR/collection/indicator/summary02.txt>> Acesso em: 23 de ago. 2008.

HARTZ, Zulmira Maria de Araújo; CHAMPAGNE, François; LEAL, Maria do Carmo and CONTANDRIOPOULOS, André-Pierre. **Mortalidade infantil "evitável" em duas cidades do Nordeste do Brasil:** indicador de qualidade do sistema local de saúde. *Rev. Saúde Pública* [online]. 1996, vol.30, n.4, pp. 310-318.

HESS, P. **The military burden, economic growth, and the human suffering index:** evidence from the LDCs. *Cambridge Journal of Economics*, 13(4), 497–515, 1989.

HODGE, T.; HOLTZ, S.; SMITH, C.; BAXTER, K. H. **Pathways to sustainability:** assessing our progress. Ottawa: NRTEE, 1995.

HOFFMANN, Rodolfo. **Distribuição de renda e crescimento econômico.** *Estud. av.*, São Paulo, v. 15, n. 41, Apr. 2001. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142001000100007&lng=en&nrm=iso>. access on 28 July 2009. doi: 10.1590/S0103-40142001000100007.

HOLTZ, Suzan. **Approaches to reporting on human well-being.** In: HODGE, R.A. et al. *Pathways to sustainability: assessing our progress.* Canada: NRTEE, 1995, p.104-125.

HORN, Robert V. **Statistical indicators for the economic and social sciences.** Cambridge: University Press, 1993.

IGNÁCIO, S. A. **Tipologia dos municípios paranaenses, segundo indicadores socioeconômicos e sociodemográficos:** uma análise estatística. 89f. Tese (Professor titular) – Pontifícia Universidade Católica do Paraná, Curitiba, 2002.

IISD. **The Dashboard of Sustainability.** International Institute for Sustainable Development, 2003.

IMBACH, A.; DUDLEY, E.; ORTIZ, N.; SÁNCHEZ, H. **Mapeo analítico, reflexivo y participativo de la sostenibilidad – MARPS.** Cambridge: UICN, 1997.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Indicadores de desenvolvimento sustentável**. Rio de Janeiro: Author, 2004

INTERNATIONAL INSTITUTE FOR SUSTAINABLE DEVELOPMENT. **The dashboard of sustainability**. Manitoba, Canada: Author, 2003.

IRSCHMANN, A. **A Estratégia do desenvolvimento**. São Paulo: Fundo de Cultura. 1961.

JANNUZZI, P. M. **Repensando a prática de uso de indicadores sociais na formulação e avaliação de políticas públicas municipais**. Anais do Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Administração, Campinas, SP, Brasil, 25, 2001.

JOHNSTON, B. T.; SHEEHY, T. P. **The index of economic freedom**. Washington: Heritage Foundation, 1995.

KLIKSBERG, Bernardo. **Falácias e mitos do desenvolvimento social**. Brasília: UNESCO, São Paulo: Cortez Editora, 2003.

LEMOS, M. B.; ESTEVES, O. A.; SIMÕES, R. F. **Uma metodologia para construção de um índice de qualidade de vida urbana**. Nova Economia, 5(2), 1995. p.157–176.

LIMA, Maria Luiza C de et al. **Evolução de homicídios por área geográfica em Pernambuco entre 1980 e 1998**. *Rev. Saúde Pública* [online]. 2002, vol.36, n.4, pp. 462-469.

LOPES, S. B. **Arranjos institucionais e a sustentabilidade de sistemas agroflorestais: uma proposição metodológica**. Porto Alegre: PGDR-UFRGS, 2001. (Dissertação de Mestrado).

MACHADO FILHO, A.; SOARES JÚNIOR, J. S.; ZEBENDE, G. F. **DFA and DCCA applied in a time series of thefts and robbery of vehicles (Salvador-BA)** In: XXXII Encontro Nacional de Física da Matéria Condensada, Águas de Lindóia, 2009.

MACHADO FILHO, A.; ZEBENDE, G.F.; SOARES JÚNIOR, J. S. **Estatística policial de Salvador-BA: uma proposta de modelagem utilizando DFA**. In: XL Simpósio Brasileiro de Pesquisa Operacional, João Pessoa, 2008.

MACHADO FILHO, A.; ZEBENDE, G.F.; SOARES JÚNIOR, J. S. **Detrended Fluctuation Analysis e Detrended Cross-Correlation Analysis: uma aplicação em indicadores sociais de criminalidade**. In: XLI Simpósio Brasileiro de Pesquisa Operacional, Porto Seguro, 2009.

MAIA, Paulo Borlina. **O impacto das mortes violentas na esperança de vida do município de São Paulo e de quatro zonas sócio-econômicas**. Disponível em: < http://www.abep.nepo.unicamp.br/docs/anais/pdf/2000/Todos/saut18_2.pdf> Acesso em 15 de jun. 2009.

MACEDO, Adriana Carneiro. PAIM, Jairnilson Silva. SILVA, Lúcia Maria Vieira da. COSTA, Maria da Conceição Nascimento. **Distribuição espacial da mortalidade por homicídios na cidade de Salvador, segundo condições de vida 1991-1994**. Disponível em: < [http://www.observatorioseguranca.org/pdf/01%20\(15\).pdf](http://www.observatorioseguranca.org/pdf/01%20(15).pdf)> Acesso em 22 de jul. 2009.

MCGRANAHAN, D. V.; RICHARD-PROUST, C.; SOVANI, N. V.; SUBRAMANIAN, M. **Contents and measurement of socioeconomic development**. New York: Praeger, 1972.

MACIAS-CHAPULA, Cesar A. **O papel da informetria e da cienciometria e sua perspectiva nacional e internacional**. Ci. Inf., Brasília, v. 27, n. 2, 1998 . Available. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200005&lng=en&nrm=iso>. Acesso em 19 de Jul. 2009.

MARZALL, K.; ALMEIDA, J. **Parâmetros e indicadores de sustentabilidade na agricultura**: limites, potencialidades e significado no contexto do desenvolvimento rural. Extensão Rural, DEAER/CPGER-CCR-UFSM, n.5, 1998.

MATTOS, Carlos. **Desenvolvimento sustentável nos territórios da globalização**: alternativa de sobrevivência ou nova utopia? In: BECKER, Bertha K., MIRANDA, Mariana (orgs). A geografia política do desenvolvimento sustentável. Rio de Janeiro: editora UFRJ, 1997, p. 103– 126.

MAXWELL, S. **The meaning and measurement of poverty**. Poverty Briefin. Retrieved January 10, 2004. Disponível em: <<http://www.odi.org.uk/publications/briefing/pov3.html>> Acesso em: 23 de jun. 2008.

MORRIS, M. D. **Measuring the condition of the world's poor: the physical quality of life index**. Pergamon Policy Studies nº 42, pp. 20–56. New York: Pergamon Press, 1979.

NAREDO, José Manuel. **Qué pueden hacer los economistas para ocuparse de los recursos naturales?** Desde el Sistema Económico hacia la Economía de los Sistemas. Pensamiento Iberoamericano, n. 12, jul.-dez. 1987, pp. 61-74.

NEWMAN, P. W. G. **Sustainability and cities**: extending the metabolism model. Australia, 1999.

NISBET, Robert A. **Social change and history**. aspects of the western theory of development. New York: Oxford University Press, 1970.

NOBRE, M. **Desenvolvimento sustentável**: origens e significado atual. In: NOBRE, M. Nobre e M. C. Amazonas (orgs.), Desenvolvimento sustentável: a institucionalização de um conceito. Brasília: Ibama, 2002.

NOGUEIRA, Roberto Passos. **Mortalidade por três grandes grupos de causa no Brasil**. Disponível em: <http://www.ipea.gov.br/sites/000/2/publicacoes/bpsociais/bps_09/ENSAIO1_Roberto.pdf> Acesso em: 23 de jul. 2009.

NOLL, Heinz H. ZAPF, Wolfgan. **Social indicators research**: societal monitoring and social reporting. In: Ingwer Borg e Peter Ph. Mohler (orgs.), Trends and perspectives in empirical social research. Berlin: Gruyter, 1994, p.1-17.

PNUD, Programa das Nações Unidas para o Desenvolvimento. Disponível em: <<http://www.pnud.org.br/>>. Acesso em 10 de set. de 2005.

POCHMANN, M.; AMORIM, R. G. **Atlas da exclusão social no Brasil**. São Paulo: Cortez, 2003.

POLANYI, M. **Personal Knowledge: Towards a post-critical philosophy**. London: Routledge & Kegan Paul, 1958.

POPPER, K. **Conocimiento objetivo**. Madrid, Editorial Tecnos, 2. ed., 1982.

PRESCOTT-ALLEN, R. **Barometer of sustainability: measuring and communicating wellbeing and sustainable development**. Cambridge: IUCN, 1997.

PRESCOTT-ALLEN, R. **Assessing progress toward sustainability: The System Assessment Method illustrated by the Wellbeing of Nations**. Cambridge: IUCN, 1999.

PRESCOTT-ALLEN, R. **The wellbeing of nations: a country-by-country index of quality of life and the environment**. Washington: Island Press, 2001.

PRESTON, Peter W. **Development theory: an introduction**. Oxford: Blackwell Publishers, 1996.

PROGRAMA DAS NAÇÕES UNIDAS PARA O MEIO AMBIENTE. Projeto GEO Cidades (Relatório Ambiental Urbano Integrado). Rio de Janeiro, RJ: Consórcio Parceria 21, 2002.

QUINTELLA, Rogério Hermida ; SOARES JUNIOR, J. S. . **Sistemas de apoio à decisão e descoberta de conhecimento em bases de dados: uma aplicação potencial em políticas públicas**. O&S. Organizações & Sociedade, Salvador - BA, v. 10, n. 28, p. 83-98, 2003.

RAM, R. **Composite indices of physical quality of life, basic needs fulfilment, and income: a principal component representation**. Journal of Development Economics, 11(2), 1982, p. 227–247.

RAYNAUT, C. **Les Sabels: diversité et dynamiques des relations sociétés-natures**. Paris: Kartjal, 1997.

RAYNAUT, C.; LANA, P. ; ZANONI, M. **Pesquisa e formação na área de meio ambiente e desenvolvimento: novos quadros de pensamento, novas formas de avaliação**. Desenvolvimento e meio ambiente, v. 1, jul-dez. Curitiba: Editora da UFPR, 2000.

RIBEIRO, Adagenor Lobato. **Modelo de indicadores para mensuração do desenvolvimento sustentável na Amazônia**. Tese (Doutorado Ciências: Desenvolvimento Sócio-Ambiental). Belém: Núcleo de Altos Estudos Amazônicos, NAEA, Universidade Federal do Pará, UFPA, 2002.

RIBEIRO, Maria Teresa Franco. **Notas de aula do curso de doutorado em administração**. Salvador: EADM – UFBA, 2005. Notas de aula.

RIST, Gilbert. **Le Développement: une histoire de croyance occidentale**. Paris: Presses de Science Po. 2001.

RIVERO, Oswaldo de. **O mito do desenvolvimento**. Os países inviáveis no século XXI. Petrópolis: Editora Vozes, 2002.

ROCHA, Sônia. Estimacão de linhas de indigência e de pobreza: opções metodológicas no Brasil. In R. O. Henriques (Ed.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000, p. 109-127.

ROCHA, Sônia. **Pobreza no Brasil**: afinal, de que se trata? Rio de Janeiro: FGV, 2003.

ROSTOW, W. W. **Etapas do desenvolvimento econômico**. Rio de Janeiro: Zahar Editores, 1971.

SACHS, Ignacy. **Desenvolvimento**: incluyente, sustentável, sustentado. Rio de Janeiro: Garamond, 2004.

SANTAGADA, Salvatore. **Indicadores Sociais**: contexto social e breve histórico. Indicadores Econômicos. n.20, v.4. p245-255, 1993.

SANTOS, Boaventura de Sousa. **O Fórum Social Mundial**. São Paulo: Cortez, 2005.

SCHUMPETER, J. **Capitalismo, socialismo e democracia**. Rio de Janeiro: Zahar Editores, 1984.

SCHUMPETER, J. **A teoria do desenvolvimento econômico**. São Paulo: Nova Cultural, 1985.

SCHUMPETER, J. **Business cycles**: a theoretical, historical and statistical analysis of the capitalist process. Philadelphia: Porcupine, 1989.

SEGNINI, Liliana Rolfsen Petrilli. **Educação e trabalho**: uma relação tão necessária quanto insuficiente. Disponível em: <<http://www.scielo.br/pdf/spp/v14n2/9791.pdf>> Acesso em 22 de jul. 2009.

SEN, Amartya K. **Desenvolvimento como liberdade**. São Paulo: Companhia das Letras, 2000.

SEN, Amartya K. **Desigualdade reexaminada**. São Paulo: Editora Record, 2001.

SEN, Amartya K. **Poverty**: an ordinal approach to measurement. *Econometrica*, Ann Arbor, MI, v. 44, n. 2, p. 219-231, Mar. 1976.

SEVERINO, Antonio J. **Educação trabalho e cidadania**: a educação brasileira e o desafio da formação humana no atual cenário histórico. Disponível em: <http://biblioteca.universia.net/html_bura/ficha/params/id/567193.html> Acesso em: 12 de ago. 2008.

SLOTTJE, D. J. **Measuring the quality of life across countries**. *Review of Economics and Statistics*, 73(4), 1991, p. 684–693.

SMITH, Adam; RICARDO, David. **Riqueza das nações**. Princípios de Economia Política. Salvador: Abril Cultural, 1978.

SOARES JUNIOR, J. S.; COSTA, E. C. C.; NAVARRO, L. M. . **Produto municipal do Estado da Bahia - 2000**. Conjuntura & Planejamento, Salvador, BA, v. 101, p. 23-28, 2002.

SOARES JUNIOR, J. S.; HORIGOSHI, Mitiko . **Padrões de desenvolvimento socioeconômico dos municípios baianos**. Conjuntura & Planejamento, v. 158, p. 22-30, 2008.

SOARES JUNIOR, J. S.; MACHADO FILHO, A. **Séries temporais de indicadores sociais: uma abordagem detrended fluctuation analysis**. In: XI Escola de Modelos de Regressão, 2009, Recife. Anais da XI Escola de Modelos de Regressão, 2009.

SOARES JUNIOR, J. S.; MACHADO FILHO, A.; QUINTELLA, Rogério Hermida; ZEBENDE, G. F. **Knowledge discovery in databases: uma análise estatística dos indicadores de criminalidade de Salvador**. Conjuntura & Planejamento, v. 161, p. 12, 2008.

SOARES JUNIOR, J. S.; MACHADO FILHO, A. **Indicadores de criminalidade e políticas públicas: knowledge discovery in databases como instrumento de apoio à decisão na segurança pública**. In: XI Encontro de Modelagem Computacional, 2008, Volta Redonda. XI Encontro de Modelagem Computacional, 2008.

SOARES JUNIOR, J. S.; QUINTELLA, R. H. **Descoberta de conhecimento em bases de dados públicas: uma proposta de estruturação metodológica**. Revista de Administração Pública, 39(5), 2005, p.1077–1107.

SOARES JUNIOR, J. S.; QUINTELLA, R. H. **Indicadores sociais de baixo custo e sua utilidade na gestão da interface entre os governos estadual e municipal**. Organizações e Sociedade, 9(25), 2002, p. 45-60.

SOARES JUNIOR, J. S.; QUINTELLA, Rogério Hermida . **Development: an Analysis of Concepts, Measurement and Indicators**. BAR. Brazilian Administration Review, v. 5, p. 104-124, 2008.

SOARES JUNIOR, J. S.; QUINTELLA, Rogério Hermida . **Instrumentalização do desenvolvimento: teorias, conceitos e indicadores**. O&S. Organizações & Sociedade, v. 45, p. 61-78, 2008.

SOARES JUNIOR, J. S.; QUINTELLA, Rogério Hermida . **Descoberta de conhecimento em bases de dados públicas: uma proposta de estruturação metodológica**. RAP. Revista Brasileira de Administração Pública, v. 39, p. 1077, 2005.

SOARES JUNIOR, J. S.; QUINTELLA, Rogério Hermida . **Descoberta de conhecimento em bases de dados públicas: o caso de uma iniciativa governamental na Bahia**. Bahia Análise & Dados, v. 14, p. 273-286, 2004.

SOARES JUNIOR, J. S. ; QUINTELLA, Rogério Hermida. **Descoberta de Conhecimento em bases de dados públicas: uma proposta de modelagem estatística para o**

subdesenvolvimento. In: X Escola de Modelos de Regressão, 2007, Salvador. X Escola de Modelos de Regressão, 2007.

SOARES JUNIOR, J. S. ; SANTOS, Ernani Marques . **Governança eletrônica**: uma perspectiva sociotécnica das organizações públicas a partir da padronização e interoperabilidade. In: XXXI Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Administração, 2007. XXXI EnANPAD, 2007.

SOARES JÚNIOR, J. S.; MACHADO FILHO, A.; QUINTELLA, Rogério Hermida; ZEBENDE, G. F. **Uma análise estatística dos indicadores de criminalidade de Salvador**. Conjuntura & Planejamento, v. 161, p. 40-49, 2008.

SOUTH COMMISSION. **Rapport de la commission, défis au sud**. Paris: Econômica, 1990.

SUNKEL, Osvaldo. “A sustentabilidade do desenvolvimento atual”, in: ARBIX, Glauco, Mauro Zilbovicius & Ricardo Abramovay (orgs.). **Razões e ficções do desenvolvimento**. São Paulo: Editora UNESP & EDUSP, 2001, pp. 267-310.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. Índices de desenvolvimento econômico e social da Bahia: 1998. Salvador: SEI, 2000.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. Índices de desenvolvimento econômico e social da Bahia: 2000. Salvador: SEI, 2002.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. Índices de desenvolvimento econômico e social da Bahia: 2002. Salvador: SEI, 2004.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. Índices de desenvolvimento econômico e social da Bahia: 2004. Salvador: SEI, 2006.

SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. Índices de desenvolvimento econômico e social da Bahia: 2006. Salvador: SEI, 2008.

SZWARCWALD, Célia Landmann; CASTILHO, Euclides Ayres de. **Estimativas da mortalidade infantil no Brasil, década de oitenta**: proposta de procedimento metodológico. *Rev. Saúde Pública* [online]. 1995, vol.29, n.6, pp. 451-462.

SZWARCWALD CL, BASTOS FI, ESTEVES MAP, ANDRADE CLT, PAEZ MS, MEDICI EV, DERRICO M. **Desigualdade de renda e situação de saúde**: o caso do Rio de Janeiro. *Cad Saúde Pública* 1999;15(1):15-28

TAYLOR, C. L.; JODICE, D. A. **World handbook of political and social indicators**. New Haven: Yale University Press, 1983.

TEIXEIRA, Luciana da Silva. **Determinantes da violência no Brasil**. Disponível em: <http://apache.camara.gov.br/portal/arquivos/Camara/internet/publicacoes/estnottec/pdf/2004_11402.pdf> Acesso em: 22 de jun. 2009.

TIRONI, L. F., SILVA, L. C. E. Silva; VIANNA, S. M.; MÉDICI, A. C. **Critérios para geração de indicadores de qualidade e produtividade no serviço público**. Brasília: IPEA/MEFP, 1991.

TRUMAN, Harry S. **Inaugural address** – January 20, 1949. Disponível em: <<http://www.harrystruman.us/TrumanInauguralSpeech.html>>. Acesso em 19 de fev. de 2006.

TRZESNIAK, P. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. **Revista de Ciência da Informação**, Brasília, v. 27, n. 2, p. 159-164, maio/ago. 1998.

TURBAN, E.; RAINER, R. K.; POTTER, R. E. **Administração de tecnologia de informação**. Rio de Janeiro: Campus, 2003.

UNICAMP. Índice DNA-Brasil. Disponível em: <<http://www.dnabrasil.org.br>>. Acesso em 25 de jan. de 2006.

UNITED NATIONS DEVELOPMENT PROGRAMME. **Human development report**. New York: Oxford University Press, 1991.

UNIVERSIDADE ESTADUAL DE CAMPINAS. **Índice DNA-Brasil**. Retrieved January 25, 2006. Disponível em <<http://www.dnabrasil.org.br>> Acesso em: 23 de jun. 2008.

VEIGA, José Eli da. **Desenvolvimento sustentável: o desafio do século XXI**. Rio de Janeiro: Garamond, 2005.

WACKERNAGEL, M.; REES, W. **Our ecological footprint**. Gabriola Island, BC and Stony Creek, CT: New Society Publishers, 1996.

WORLD BANK. **The World Bank**. Disponível em: <<http://web.worldbank.org>>. Acesso em 12 de set. de 2005.

World Economic Forum. **Global competitiveness report**. Geneva: Author, 1996.

YALE CENTER FOR ENVIRONMENTAL LAW AND POLICY & CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK. **Environmental Sustainability Index: Benchmarking National Environmental Stewardship**. New Haven, U.S.: Author, Yale University and New York, U.S.: Author, Columbia University, 2005.

ZEBENDE, G. F.; MACHADO FILHO, A. **Detrended Cross-Correlation Analysis: measuring cross-correlation between mathematical constants**. In: XXXII Encontro Nacional de Física da Matéria Condensada, Águas de Lindóia, 2009.

ZEBENDE, G. F.; MACHADO FILHO, A. Cross-correlation between time series of vehicles and passengers. *Physica A*, n.3, v.88, p. 4863-4866, 2009.

APENDICE A – Sistemas e Índices Nacionais de Desenvolvimento, 1991 – 2005

ANO	AUTOR	TÍTULO
1991	CEPAL	Linha da indigência da Cepal
1991	Árias	Linha da Indigência do IBGE, IPEA, Cepal
1995	Prescott-Allen	Barômetro de Sustentabilidade
1995	Cobb et al	Indicador de Progresso Genuíno (GPI)
1995	PMBH e PUC/MG	Índice de Qualidade de Vida Urbana (IQVU)
1995	SOUTO ET AL	Índice Municipal (IM)
1995	SOUTO ET AL	Índice de Sustentabilidade Municipal (ISM)
1995	SOUTO ET AL	Índice Social Municipal (ISM)
1995	SOUTO ET AL	Índice de Gestão Municipal (IGM)
1997	PNUD	Glossário da Pobreza e Desenvolvimento Humano
1998	FJP, IPEA, PNUD	Índice de Desenvolvimento Humano Municipal (IDH-M)
1998	FJP, IPEA, PNUD	Índice de Condições de Vida (ICV)
1998	Seattle	Indicadores de comunidade sustentável
1998	ARANHA	Indicador Ponderado de Carência (IPC)
1998	ARANHA	Índice de Potencial de Consumo (IPC)
1999	RONDONIA	IR do ICMS Índice de Repartição do ICMS
1999	KAGEYAMA; LEONE	Tipologia de Economias Regionais (TER)
2000	Rocha	Linha da Indigência do IPEA
2000	FEE	Índice de Desenvolvimento Social (IDS)
2001	Ferreira et al	A Robust Poverty Profile for Brazil Using Multiple Data Sources
2001	FGV	Mapa do fim da fome
2001	PARÁ	Índice de Potencialidade Socioeconômica dos Municípios (IPSM)
2001	PARÁ	Índice Geral de Qualidade de Vida Urbana (IGQVU)
2001	IPES	Índice de Desenvolvimento dos Municípios (IDM)
2001	FEE	Índice Social Municipal Ampliado (ISMA)
2002	Ribeiro	Indicadores do estado do Pará
2002	SEI	Índice de Desenvolvimento Econômico (IDE)
2002	SEI	Índice de Desenvolvimento Social (IDS)
2002	Braga et al	Índice de Sustentabilidade Urbana (ISU)
2002	CAMPUS ET AL	Índice de Exclusão Social (IES)
2002	RIBEIRO	Índice Agregado de Sustentabilidade da Amazônia (IASAM)
2002	SEI	Índice de Desenvolvimento Econômico (IDE)
2002	SEI	Índice de Desenvolvimento Social (IDS)
2002	CIDE	Índice de Qualidade dos Municípios (IQM)
2002	CIDE	Índice de Qualidade dos Municípios (IQM-C)
2002	CIDE	Índice de Qualidade dos Municípios (IQM-V)
2002	CIDE	Índice de Qualidade dos Municípios (IQM-SF)
2002	SANTA CATARINA	Índices de Desenvolvimento Social e Humano (IDS)
2002	IGNACIO	Tipologia dos Municípios Paranaenses (TMP)
2003	TOCANTINS	Índice do ICMS Ecológico (ICMS Eco)
2003	UFPI	Índice de Bem-Estar Social para os Municípios (IBRG)
2003	IEME	Índice Geral de Competitividade (IGC)
2004	Pochmann et al	Atlas da Exclusão Social
2004	SMA/SP/MG	Geo Cidades
2004	FEE	Índice de Desenvolvimento Socioeconômico (IDESE)
2004	IBGE	Índice de Desenvolvimento Sustentável (IDS)
2004	HADDAD	Índice de Qualidade Institucional dos Municípios (IQIM)
2004	HADDAD	Índice do Potencial de Desenvolvimento do Município (IPDM)
2004	SILVA E RIBEIRO	Índice de Degradação (ID)
2004	IPECE	Índice de Desenvolvimento Social de Resultados (IDS-R)
2004	IPECE	Índice de Desenvolvimento Social de Oferta (IDS-O)

ANO	AUTOR	TÍTULO
2004	EMBRAPA	Índice de Sustentabilidade Ambiental do Uso da Água (ISA-Água)
2004	GOIAS	Índice de Desenvolvimento Econômico (IDE)
2004	GOIAS	Índice de Desenvolvimento Social (IDS)
2004	IPES	Índice de Carência em Saneamento Básico (ICSB)
2004	LEMOS ET AL	Índice de Inadequabilidade Habitacional (IIH)
2004	PMPA	Índice de Vulnerabilidade Social Infanto-Juvenil (IVS-IJ)
2004	SANTA CATARINA	Indicadores Educacionais dos Municípios de Santa Catarina (IDE)
2005	Soares Jr; Quintella	Mapeamento da Pobreza
2005	UNICAMP	Índice DNA-Brasil
2005	UFRJ	Índice de Carência Habitacional (ICH)
2005	CEDEPLAR	Índice para Mapeamento da Qualidade de Vida (IQV)
2005	SEADE	Índice Paulista de Responsabilidade Social (IPRS)
2005	SEADE	Índice Paulista de Vulnerabilidade Social (IPVS)

APENDICE B – Sistemas e Índices Internacionais de Desenvolvimento por autor e título, 1964 – 1999

Ano	Autor	Título
1976	Abt, Clark C. (ed).	The Evaluation of Social Programs. Beverly Hills: Sage Publications.
		Boundary Crossers: Case Studies of How Ten of Americas Metropolitan Regions Work. College Park, MD: The James MacGregor Burns Academy of Leadership, University of Maryland, College Park.
1997	Adams, Bruce and John Parr.	Alternative Economic Indicators. New York: Routledge.
1991	Anderson, Victor.	"Long-term Social Development of Hungary, Measured by Social Indicators." Social Indicators Research
1980	Andorka, Rudolf.	Social Indicators of Well-being: Americans' Perceptions of Life Quality. New York: Plenum Press.
1976	Andrews, F. W. and S. B. Withey.	Social Indicators of Well-being. New York: Plenum Press.
1976	Andrews, Frank A. and Stephen B. Withey.	Research on the Quality of Life. Ann Arbor: University of Michigan Press.
1986	Andrews, Frank A., ed.	"Social Indicators and Health for All." Social Science and Medicine
1981	Andrews, Frank M.	"Population Issues and Social Indicators of Well-being." Population and Environment
1983	Andrews, Frank.	"Going by the Numbers: Using Indicators to Know Where You've Been—and Where You're Going." Planning 62: 14-8.
1996	Andrews, James H.	"Social Indicators." in Implementing Change in Communities: A Collaborative Process. Edited by Sarah Ellen Archer. St. Louis, KS: The C.V. Mosby Company.
1984	Archer, Sarah Ellen, Carole D. Kelly and Sally Ann Bisch.	"Developing Indicators of Sustainable Community: Lessons from Sustainable Seattle." Environmental Impact Assessment Review 16: 337- 350.
1996	Atkisson, Alan.	Florida State University.
1998	Audirac, Ivonne.	Take Charge: Economic Development for Small Communities. Ames, IA: North Central Regional Center for Rural Development.
	Ayres, Janet, Robert Cole, Clair Hein, Stuart Huntington, Wayne Kobberdahl, Wanda Leonard and Dale Zetocha.	
1990	Leonard and Dale Zetocha.	

- 1996 Azar, Christian, John Holmberg and Kristian Lindgren. "Methodological and Ideological Options Socio-ecological Indicators for Sustainability." Ecological Economics
Sustainable Cities. Research Triangle Park, NC: Research Triangle Institute.
- 1999 Barnett, Camille Cates and Francine Luloffs. The Social Indicators Movement: Problems and Contradictions. Sotsiologicheskie Issledovaniya 7
- 1980 Batygin, G. S. and Shchelkin, A. G. Social Indicators. Cambridge, MA: M.I.T. Press.
- 1966 Bauer, Raymond. Social Indicators: Bibliography I. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology, Sociology Report No. 92.
- 1971 Beal, G. M., R. M. Brooks, L. D. Wilcox and G. E. Klonglan. "Measuring Quality Of Life." Social Indicators Research 42: 21-39.
- 1997 Beckie, T.M. and L.A. Hayduk.

Ano	Autor	Título
1997	Bernard, James R.	Frameworks for Categorizing and Organizing Environmental Indicators. Montpelier, VT: Green Mountain Institute for Environmental Democracy.
1981	Bertrand, William, Nancy Beth Mock and Robert R. Franklin.	“Social Indicators: Their Use in Evaluation Research.” Pp. 77-90 in Methodological Advances in Evaluation Research. Edited by Ross F. Conner. Beverly Hills, CA: Sage Publications.
1994	Bertuglia, C. S., G. P. Clarke and A. G. Wilson.	Modeling the City: Performance, Policy, and Planning. London and New York: Routledge.
1997	Besleme, Kate and Megan Mullin.	“Community Indicators and Healthy Communities.” National Civic Review 86 (1): 43-52.
1999	Besleme, Kate, Elisa Maser and Judith Silverstein.	A Community Indicators Case Study: Addressing the Quality of Life in Two Communities. San Francisco: Redefining Progress.
1982	Bestuzhey, L. and Igor Vasilevich.	Social Indicators of the Way of Life of Society: Comparative Study. International Sociological Association.
1979	Bestuzhey, L. and Igor Vasilevich.	Systems of Social Indicators: Theory and Practice. Sotsiologicheskies Issledovaniya
1983	Bird, Carolyn.	The Good Years: Your Life in the Twenty-First Century. New York: E. P. Dutton, Inc.
1983	Bloom Bernard L.	“The Use of Social Indicators in the Estimation of Health Needs.” Pp. 147-162 in Assessing Health and Human Service Needs: Concepts, Methods and Applications. Edited by Roger Bell, Martin Sundel, Joseph Aponke, Stanley A. Murrell and Elizabeth Lin. New York: Human Sciences Press Inc.
1998	Boonchit, W. and S. Natenuj.	The Eighth National Economic and Social Development Plan and Current Economic Adjustment and Indicators for Monitoring and Evaluation of the Eighth Plan. Working Paper No. 64, Tokyo, Japan: Department of Research Cooperation, Economic Research Institute, Economic Planning Agency.
1989	Boyer, Rick and David Savageau.	Places Rated Almanac: Your Guide to Finding the Best Places to Live in America. New York: Prentice Hall.

1985	Boyer, Rick and David Savageau.	Places Rated Almanac: Your Guide to Finding the Best Places to Live in America. New York: Prentice Hall.
1981	Boyer, Rick and David Savageau.	Places Rated Almanac: Your Guide to Finding the Best Places to Live in America. New York: Prentice Hall.
1971	Brenner, Harvey M.	Time Series Analysis of Relationships between Selected Economic and Social Indicators. Washington, D.C.: U.S. Department of Labor Manpower Administration.
1998	Brown, Lester R. and Ed Ayres (eds).	Worldwatch Reader on Global Environmental Issues. New York: W. W. Norton and Co.
1998	Brown, Lester R., Christopher Flavin, Hilary F. French, Janet Abramovitz, Chris Bright, Seth Dunn, Gary Gardner, Anne McGinn, Jennifer Mitchell, Michael Renner, David Roodman, John Tuxill and Linda Starke. 1998. State of the World	A Worldwatch Institute Report on Progress Toward a Sustainable Society. New York: W.W. Norton and Co.

Ano	Autor	Título
1998	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1997	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1996	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1995	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1994	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1993	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1992	Brown, Lester, Christopher Flavin and Hal Kane.	Vital Signs: The Trends that are Shaping our Future. New York: W.W. Norton and Co.
1993	Brown, Valerie A.	“The uses of social and environmental health indicators in monitoring the effects of climate change.” Climatic Change 25 (3-4): 389-404.
1996	Brutschy, Susan, Therese Doherty, Yolanda Goda, Kelly Pleskunas and Dan Mason.	Community Assessment Project—Santa Cruz County, California. La Selva Beach, CA: Applied Survey Research and Santa Cruz County United Way.
1976	Byrnes, K.J. and J. Shadi-Talab.	A Methodology for Indicators for Social Development: Supplemental Report 3: The Small Farmer Agricultural Sector. Sociology Report 127-3. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1984	Cagle, Laurence T.	“Using Social Indicators to Assess Mental Health Needs: Lessons from a Statewide Study.” Evaluation Review 8
1974	Callaghan, J., L. D. Wilcox, K. Byrnes, S. M. Hunter and S. M. Kim.	A Methodology for Indicators of Social Development: Report 4: Health Sector Information System. Sociology Report 121. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1974	Callaghan, John Oliver.	Indicators of Social Development: Conceptual and Methodological Considerations in Monitoring the

		Health Sector. Unpublished Doctoral Dissertation. Iowa State University.
1976	Campbell, Angus, Philip Converse and Willard L. Rodgers.	The Quality of American Life: Perceptions, Evaluations and Satisfaction. New York: Russell Sage Foundation.
1981	Campbell, Angus.	The Sense of Well-being in America. New York: McGraw-Hill Book Company.
1991	Caplow, T., H. M. Bahr, J. Modell, B. A. Chadwick.	Recent Social Trends in the United States 1960-1990. Montreal and Kingston, ON, Canada: McGill-Queen's University Press.
1981	Carley, M.	Social Measurement and Social Indicators: Issues of Policy and Theory. London: George Allen and Unwin.
1984	Carley, Michael, J. and Eduardo S. Bustelo.	Social Impact Assessment and Monitoring: A Guide to the Literature. Boulder and London: Westview Press.
1996	Carlson, Ulrika, John Homberg and Goran Berndes.	"Socio-ecological Indicators for Sustainability for Gotland, Sweden." Chalmers University of Technology, Goteborg University, Institute of Physical Resource Theory. Institute Report 1996: 07.
1984	Carr-Hill, Roy A.	"The Political Choice of Social Indicators." Quality and Quantity 18

Ano	Autor	Título
1977	Carter, Keith, Rodney Ganey, Frank Fear and Chris Marshall.	A Social Report for Greene County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1977	Carter, Keith, Rodney Ganey, Frank Fear and Chris Marshall.	A Social Report for Webster County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1981	Chandraskar, Krishnamurti.	“Productivity and Social Indicators.” Annals of the American Academy of Political and Social Science 453 (Jan):
1992	Clavijo, Sergio.	“Variations on the basic needs yardstick: an application to Colombia.” World Development, 20: Lessons Learned from the History of Social Indicators. San Francisco: Redefining Progress.
1998	Cobb, Clifford W. and Craig Rixford.	“If the Economy is Up, Why is America Down?” Atlantic Monthly, see
1995	Cobb, Clifford, Ted Halstead and Jonathan Rowe.	The Genuine Progress Indicator: Summary of Data and Methodology. San Francisco: Redefining Progress.
1995	Cobb, Clifford, Ted Halstead and Jonathon Rowe.	Urban Life in Texas: A Statistical Profile and Assessment of the Largest Cities. Austin, TX: University of Texas Press.
1986	Cole, Richard L., Ann Crowley Smith and Delbert A. Taebel.	Subjective Social Indicators: A Theoretical and Conceptual Exploration into Human Well-being. Unpublished Masters Thesis. Iowa State University
1976	DeMarzo, Robert Salvatore.	Social Impact Analysis and Development Planning in the Third World. Boulder: Westview Press.
1985	Derman, William and Scott Whiteford.	“A Broader Perspective for Social Indicators.” Australian Journal of Social Issues 16
1981	Devin, Marianne.	“Assessing Subjective Well-being: Progress and Opportunities.” Social Indicators Research 31
1994	Diener, Ed.	Rural Society in the U.S.: Issues for the 1980s. Boulder, CO: Westview Press.
1982	Dillman, D. A. and D. J. Hobbs, eds.	“Suggestions for a Model to Measure Quality of Life through Social Indicators (in Italian).” Studi-di-
1983	Di-Orio, Ferdinando.	

		Sociologia 21: 29-42.
1997	Doak, Sam and Jonathon Kusel.	Well-being of Communities in the Klamath Region. United States Forest Service, Klamath National Forest.
1984	Dobrianov, Velichko.	Social Indicators and Social Management. Social Indicators Research 14
1995	Dorgan, Charity, ed. 1995.	Gale Country and World Rankings Reporter. New York: Gale Research Inc.
1986	Drewnowski, Jan.	“The Level of Civilization: A New Field for the Application of Social Indicators.” Social Indicators Research 18
1970	Drewnowski, Jan.	Studies in the Measurement of Levels of Living and Welfare. Report No. 70.3, Geneva: United Nations Research Institute for Social Development.
1998	Eckhardt, C.	The Human Factor in Ecological Research: An Annotated Bibliography. Portland, OR: Pacific Northwest Research Station, U.S. Department of Agriculture, Forest Service.
1974	Edwards, Bernard.	Sources of Social Statistics. London: Heinemann.
1996	Edwards, Kevan.	Resources for Social Researchers. Minneapolis, MN: University of Minnesota.

Ano	Autor	Título
1984	Estes, Richard.	The Social Progress of Nations. New York: Praeger.
1994	Evans, David R.	“Enhancing Quality of Life in the Population at Large.” Social Indicators Research 33: 47-88.
1985	Evans, David R., Joan E. Burns, Wendy E. Robinson and Owen J. Garrett.	“The Quality of Life Questionnaire: A Multidimensional Measure.” American Journal of Community Psychology 13
1990	Eyles, John.	“Objectifying the Subjective: The Measurement of Environmental Quality.” Social Indicators Research, 22
1982	Fabio, Taiti.	From Quantity to Quality: New Social Indicators for an Advanced Country: Italy toward the Year 2000. International Sociological Association.
1977	Fear, Frank Anthony.	Exploratory Social Indicator Model Building and Testing using Multiple Indicators with a Block Recursive System. Unpublished Doctoral Dissertation. Iowa State University.
1977	Fear, Frank, Chris Marshall, Keith Carter and Rodney Ganey.	A Social Report for Boone County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1977	Fear, Frank, Chris Marshall, Keith Carter and Rodney Ganey.	A Social Report for Hamilton County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1980	Felson, Marcus and Kenneth C Land.	Social Indicators Time Series Archive for the United States, 1946-1980. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
1984	Felson, Marcus, and Michael Gottfredson.	“Social Indicators of Adolescent Activities Near Peers and Parents.” Journal of Marriage and the Family 46
1981	Fischer, Dietrich.	Major Global Trends and Causal Interactions Among Them. Tokyo, Japan: United Nations University.
1972	Flax, M. J.	A Study in Comparative Urban Indicators: Conditions in 18 Large Metropolitan Areas. Washington, D.C.: The Urban Institute.
1972	Flax, Michael J.	A Comparative Study of Urban Indicators: Conditions

		in 18 Metropolitan Areas. Springfield, VA: National Technical Information Service. U.S. Department of Commerce.
1999	Flora, Cornelia B., Kelly Goddard, Michael Kinsley, Vicki Luther, Milan Wall, Susan Odell, Shanna Ratner, Janet Topolsky.	Measuring Community Success and Sustainability: An Interactive Workbook. Ames, IA: North Central Regional Center for Rural Development, Resource Conservation and Development Program and the Social Sciences Institute, USDA-NRCS, and Rural Community Assistance of the USDA Forest Service.
1995	Flora, Cornelia Butler, Margaret Kroma and Alison Meares.	Indicators of Sustainability: Community and Gender. Proceedings of the Indicators of Sustainability Conference and Workshop. SANREM Research Report No. 1-95. Cullowhee, NC: Center for PVO University Collaboration.
1997	Flora, Cornelia Butler.	“Quality of Life.” Encyclopedia of Rural America: Volume 2, M-Z. Edited by Gary A. Goreham. Santa Barbara, CA: ABC-CLIO.

Ano	Autor	Título
1997	Flower, Joe.	“Beyond Economics: Healthy Communities and Healthy Economies.” National Civic Review 86
1997	Force, Jo Ellen and Gary Machlis.	“The Human Ecosystem Part II: Social Indicators in Ecosystem Management.” Society and Natural Resources 10
1985	Fox, Karl A.	Social Systems Accounts: Linking Social and Economic Indicators Through Tangible Behavior Settings. Dordrecht, The Netherlands: Reidel Publishing Co.
1974	Fox, Karl.	Social Indicators and Social Theory: Elements of an Operational System. New York: John Wiley and Sons.
1989	Fox, Karl.	“Environmental Quality in a New System of Social Accounts.” Pp. 189- 201. In Economy and Ecology: Towards Sustainable Development, R. Archibugi and P. Nijkamp (eds). Pp. 189-201. Boston: Kluwer Academic Publications.
1998	Freshley, Craig.	Sustain Western Main indicators.
1997	Frey, R. S. and F. Song.	“Human Well-Being in Chinese Cities.” Social Indicators Research: an International and Interdisciplinary Journal for Quality Measurement 42 (1): 41-76.
1997	Friedman, M. I.	Improving the Quality of Life: A Holistic Scientific Strategy. Westport, CT: Praeger.
1996	Frizzel, A. and J.H Pammett, eds. 1996.	Social Inequality in Canada. Ottawa: Carleton University Press.
1997	Gabriel, Roy, M.	“Community Indicators of Substance Abuse: Empowering Coalition Planning and Evaluation.” Evaluation and Program Planning 20 (3): 335-343.
1977	Ganey, Rodney, Frank Fear, Chris Marshall and Keith Carter.	A Social Report for Humboldt County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1994	Garoogian, Rhoda, Andrew Garoogian and Joy Fromm, eds.	America's Toprated Smaller Cities: A Statistical Handbook. Boca Raton, FL: Universal Reference Publications.

1998	Garrison, Betsy M.E.	“Determinants of the Quality of Life of Rural Families.” <i>The Journal of Rural Health</i> 14
1981	Gilbert, Gwendolyn C.	“The Political Use of Social Indicators Research to Combat Racism in the 1980’s.” <i>Social Development Issues</i> 5
1990	Gilford, Dorothy M. and Ellen Tenenbaum (eds). 1990.	<i>Precollege Science and Mathematics Teachers: Monitoring Supply, Demand, and Quality.</i> Washington, D.C.: National Academy Press.
1981	Gilford, Dorothy M., Glenn L. Nelson and Linda Ingram (eds).	<i>Rural America in Passage: Statistics for Policy.</i> Washington, D. C.: National Academy Press.
1979	Gilmartin, K. J., R. J. Rossi, L. S. Lutomski and D. F. B. Reed.	<i>Social Indicators: An Annotated Bibliography of Current Literature.</i> New York: Garland Publishing Company.
1982	Goldstein, Jeanette, Alex Zautra and Darlene Goodhard.	“A Test of the Utility of Social Indicators for Behavioral Health Service Planning.” <i>Social Indicators Research</i> 10
1997	Gruenwald, Paul, Andrew J. Treno, Gail Taff and Michael Klitzner.	<i>Measuring Community Indicators: A Systems Approach to Drug and Alcohol Problems.</i> Three Oaks, CA: Sage Publications.

Ano	Autor	Título
1983	Gumprecht, Nancy.	“Block Grants and Funding Decisions: The Case for Social Indicators.” <i>Social Service Review</i> 57
1986	Hafstrom, Jeanne.	search. Urbana: IL. Illinois Agricultural Experiment Station.
1978	Hall, Orman and David Royce.	“Mental Health Assessment with Social Indicators: An Empirical Case Study.” <i>Administration in Mental Health</i> 15: 36-46.
1997	Hart, Maureen.	“Guide to Sustainable Community Indicators. North Andover, MA: Hart Environmental Data Research.
1980	Harvey, Andrew S., E. Grant MacDonald, W. Stephen Macdonald, David H. Elliot, Susan M. Clark and Scott Wood.	A Blueprint for Core Social Indicators: Meeting Social Data Needs for the 1980's. Halifax, N.S.: Institute of Public Affairs, Dalhousie University.
1997	Hayes, Michael V. and Sharon Manson Willams.	“Healthy Community Indicators: The Perils of the Search and the Paucity of the Find.” <i>Health Promotion International</i> 5: 161-166.
1996	Henderson, Hazel	Building a Win-Win World: Life Beyond Global Economic Warfare. San Francisco: Berrett-Koehler Publishers.
1985	Hilhorst, J. G. M. and M. Klatter.	Social Development in the Third World: Level of Living indicators and Social Planning. London and Dover, NH: Croom Helm.
1994	Hoff, Marie D. and John G. McNutt,	The Global Environmental Crisis: Implications for Social Welfare and Social Work. Brookfield, VT: Avebury/Ashgate Publishing Limited.
1998	Hoff, Marie, ed.	Sustainable Community Development: Studies in Economic, Environmental, and Cultural Revitalization. Boca Raton: CRC Press.
1993	Horn, Robert Victor.	Statistical Indicators for the Economic and Social Sciences. Cambridge and New York: Cambridge University Press.
1980	Hsiao,-Hsin-huang-Michael.	“Social Indicators and Quality of Life: A Critical Review.” <i>National Taiwan University Journal of Sociology</i> 14 (Nov)
1975	Hunter, MacDonald Sandra.	Theoretical Considerations for the Development of Social Indicators: The Educational Sector.

		Unpublished Doctoral Dissertation. Iowa State University.
1998	Illner, Michal.	“The Changing Quality of Life in a Post-Communist Country: The Case of Czech Republic.” Social Indicators Research 43
1975	Innes de Neufville, J. E.	Social Indicators and Public Policy: Interactive Processes of Design and Application. New York: Elsevier Scientific Publishing Company.
1982	Innes de Neufville, Judith.	“Social Indicators of Basic Needs: Quantitative Data for Human Rights Policy.” Social Indicators Research 11: 383-403.
1990	Innes, J. E.	Knowledge and Public Policy: The Search for Meaningful Indicators. New Brunswick: Transaction Publishers.
1975	Jenkins, Thomas and Robert Seufert	Theory, Research, Policy and Action: A Bibliography of Planning and Social Analysis. Monticello, IL: Council of Planning Libraries.

Ano	Autor	Título
1977	Johnston, D.	Basic Disaggregations of Main Social Indicators. Paris: Organization for Economic Co-operation and Development.
1981	Johnston, Denis F. and Michael J. Carley.	"Social Measurement and Social Indicators." Annals of the American Academy of Political and Social Science 453
1981	Johnston, Denis F., ed.	Measurement of Subjective Phenomena: Special Demographic Analysis. Washington, D.C.: Center for Demographic Studies. Bureau of Census.
1981	Jordakel, J.	"Analysis of Development by Means of Social Indicators." Socioloski Pregled 15
1985	Jorgensen, Joseph G., Richard McCleary and Steven McNabb.	"Social Indicators in Native Village Alaska." Human Organization 44: 2-17.
1981	Juster Thomas F. and Kenneth C. Land, eds.	Social Accounting Systems: Essays on the State of the Art. New York: Academic Press.
1996	Kibel, Barry, M.	An Introduction to Results Mapping: Working for Maximum Results. December 1996, Bethesda, MD: Pacific Institute for Research and Evaluation (PIRE).
1992	Kim, Kwangkee and Philip M. Moody.	"More Resources Better Health: A Cross-national Perspective." Social Science and Medicine 34 (8):
1982	Kiuranov, Chaydar.	Social Indicators as an Instrument of Social Change. International Sociological Association.
1993	Kline, Elizabeth and Nancy Goodman.	Defining a Sustainable Community. Medford, MA: Center for Environmental Management, Tufts University.
1995	Kline, Elizabeth.	Sustainable Community Indicators. Medford, MA: Consortium for Regional Sustainability, Global Development and Environment Institute, Tufts University.
1995	Kline, Elizabeth. February	"Sustainable Community Indicators: Examples from Cambridge, MA." Medford, MA: Consortium for Regional Sustainability, Global Development and Environment Institute, Tufts University.
1976	Klonglan, Gerald E., George M. Beal, Leslie D. Wilcox, Frank A. Fear and K. William Wasson.	Toward A Methodology for Social Indicators in Rural Development: Summary Report. Sociology Report No.

		124 (January 1976). Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1976	Klonglan, Gerald E., Richard D. Warren, Frank A. Fear, Rodney F. Ganey, Christopher E. Marshall and Keith A. Carter.	Social Indicators for Rural Development: Strategies and Approaches: A Progress Report. Sociology Report 132 (November 1976). Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1975	Knox, Paul.	Social Well-being: A Spatial Perspective. London: Oxford University Press.
1974	Kogan, Leonard S. and Shirley Jenkins.	Indicators of Child Health and Welfare: Development of the DIPOV Index. New York: Columbia University Press.
1985	Koran, Lorrin M. and Kenneth Meinhardt.	“Social Indicators in Statewide Mental Health Planning: Lessons from California.” Social Indicators Research 15: 131-144.
1979	Kurian, George Thomas.	The Book of World Rankings. New York: Facts on File, Inc.

Ano	Autor	Título
1996	Kusel, Jonathon.	“Well-Being in Forest Dependent Communities, Part I: A New Approach.” Pp. 361-374 in Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Assessments and scientific basis for management options. Davis: University of California, Centers for Water and Wildland Resources.
1983	Kutta, Fratisek, Michal Illner.	System of Social Indicators and Elaboration of the Information Basis for Social Planning and Forecasting.” Sociologicky-Casopis 19
1980	Land, K. C. and McMillen, M. M.	“Demographic Data and Social Indicators.”Sociology and Social Research 64
1975	Land, Kenneth and Seymore Spilerman.	Social Indicator Models. New York: Russell Sage Foundation.
1983	Land, Kenneth C.	Social Indicators.” Annual Review of Sociology 9:
1990	Landecker, Werner S., Angus Campbell and Robert W. Marans.	Detroit Area Study, 1980: The Sociology of Knowledge and the Quality of Life in Detroit. ICPSR 9302. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
1975	Liu, Ben-Chieh.	Quality of Life Indicators in the U.S. Metropolitan Areas, 1970: A Comprehensive Assessment. Washington, D.C.: U.S. Environmental Protection Agency, Washington Environmental Research Center.
1990	Loftin, Colin.	Detroit Area Study, 1979: A Study of Metropolitan Issues. ICPSR 9301. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
1992	MacBeath, John.	“Learning to Use the Indicators.” Times Educational Supplement 3945: 10.
1995	Maclaren, Virginia W.	Developing Indicators of Urban Sustainability: The Canadian Experience. Toronto: State of the Environment Directorate, Environment Canada; Canada Mortgage and Housing Corporation; Intergovernmental Committee on Urban and Regional Research.
1985	MacRae, Duncan.	Policy Indicators: Links Between Social Science and Public Debate. Chapel Hill: University of North

Carolina Press.		
1982	Majumdar, Atreyi.	Development and Welfare Indicators: A Critical Appraisal. New Delhi: Allied.
1979	Mancini, J. A.	“Social Indicators of Family Life Satisfaction: A Comparison of Husbands and Wives.” International Journal of Sociology of the Family 9: 221- 231.
1983	Mangahas, Mahar.	“Measurement of Poverty and Equity: Some ASEAN Social Indicators Experience.” Social Indicators Research 13
1978	Marshall, Chris, Keith A. Carter, Frank A. Fear, Rodney F. Garney and Erik R. Anderson.	Needs Assessment for Social Planning: Ideas and Approaches. Monticello,
1977	Marshall, Chris, Keith Carter, Rodney Ganey and Frank Fear.	A Social Report for Carroll County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1977	Marshall, Chris, Keith Carter, Rodney Ganey and Frank Fear.	A Social Report for Pocahontas County: Social Indicators for Rural Development. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.

Ano	Autor	Título
1983	Matejko, Alexander J.	1983. "Note on Social Indicators." <i>Revue Internationale de Sociologie</i> 19: 201-209.
1998	Matutinovic, Igor.	"Quality of Life in Transition Countries: Central East Europe with Special Reference to Croatia." <i>Social Indicators Research</i> 43: 97-119.
1985	McGranahan, Donald, Eduardo Pizarro and Claude Richard.	Measurement and Analysis of Socioeconomic Development: An Enquiry into International Indicators or Development and Quantitative Interrelations of Social and Economic Components of Development. Geneva: United Nations Research Institute for Social Development.
1975	McIntosh, W. A., L. Wilcox, G. M. Beal and J. Callaghan.	A Methodology for Indicators of Social Development: Report 5: Social Information and Development Planning in Thailand. <i>Sociology Report</i> 126. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1976	McIntosh, William Alex, Susan Evers, John Callaghan and Leslie D. Wilcox.	A Methodology for Indicators of Social Development: Health and Nutrition; Supplementary Report 1: <i>Sociology Report</i> No. 127 (January 1976). Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1975	McIntosh, William Alexander.	Social Indicators for the Monitoring of the Nutritional Aspects of Societal Well-being. Unpublished Doctoral Dissertation. Iowa State University.
1997	McMurtry, R. Y. and A. D. Brown.	"The Bank of Canada as a Determinant of Health." <i>Social Indicators Research</i> 40: 179-187.
1979	Merryman, John Henry, David S. Clark and Lawrence M. Friedman.	Law and Social Change in Mediterranean Europe and Latin America. Stanford, CA: Stanford Law School.
1976	Merwin, Donna J. 1976. The Quality of Life: A Bibliography of Objective and Perceptual Social Indicators. Monticello, IL:	Council of Planning Librarians.
1998	Metzenbaum, Shelly.	Making Measurement Matter: The Challenge and Promise of Building a Performance-Focused Environmental Protection System. Washington, D.C.: Center for Public Management, The Brookings Institution.

1980	Michalos, A. C.	North American Social Report: A Comparative Study of the Quality of Life in Canada and the U.S.A. from 1964-1974 (volumes I-V). Boston: D. Reidel Publishing Company.
1980	Mills, Cadman Atta.	On Social Indicators and Development. HSDRGPID-33/UNUP-147. Tokyo: United Nations University.
1978	Moland, John Jr.	"Social Indicators with Emphasis on the Quality of Life among the Black Aged in Rural Louisiana." Black Sociologist 8:
1998	Moller, Valerie.	"Quality of Life in South Africa: Post-Apartheid Trends." Social Indicators Research 43:
1997	Moore, James.	"Raising the Bar: Bringing Accountability to Health Care." National Civic Review 86: 23-32.
1982	Morgan, David.	Handbook of State Policy Indicators. Norman, Oklahoma: Bureau of Research, University of Oklahoma.

Ano	Autor	Título
1992	Morreim, E. Haavi.	“The Impossibility and the Necessity of Quality of Life Research.” <i>Bioethics</i> , 6
1988	Moum, Torbjorn.	“Yea-saying and Mood-of-the-day Effects in Self-reported Quality of Life.” <i>Social Indicators Research</i> 20
	Muoghalu, Leonard N.	“Subjective Indices of Housing Satisfaction as Social Indicators for Planning Public Housing in Nigeria.” <i>Social Indicators Research</i> 15: 145-164.
1973	Mushkin, Selma J.	National Assessment and Social Indicators, January 1973. Washington, D.C.: National Center for Educational Statistics, Superintendent of Documents, Printing Office.
1995	Oropesa, R.S.	“Consumer Possessions, Consumer Passions, and Subjective Wellbeing.” <i>Sociological Forum</i> 10
1985	Osman, Tony.	Facts of Everyday Life. Boston: Faber and Faber.
1984	Perry, Charles S.	“Economic Activity and Social Indicators: A Rural-Urban Discontinuum?” <i>American Journal of Economics and Sociology</i> 43
1998	Raphael, Dennis, Brenda Steinmetz and Rebecca Renwick.	How to Carry Out a Community Quality of Life Project: A Manual—A Health Promotion Approach to Understanding Communities. To receive copies of the report contact Dennis Raphael, University of Toronto, McMurrich Building, Toronto, Ontario, M5S 1A8. (416) 978-7567; fax (416) 978-2087.
1992	Rastogi, P.N.	Policy Analysis and Problem-Solving for Social Systems: Toward Understanding, Monitoring, and Managing Complex Real World Problems. Newbury Park, CA: Sage Publications Inc.
1996	Reynolds, Farley.	The New American Reality: Who We Are, How We Got Here, Where We Are Going. New York: Russell Sage Foundation.
1977	Rist, Ray C. and Ronald J. Anson.	Education, Social Change and the Judicial Process. New York: Teachers College Press.
1997	Roberts, Eric.	“Neighborhood Social Environments and the Sistribution of Low Birthweight in Chicago.” <i>American-Journal-of-Public-Health</i> 87: 597-603.

1985	Robertson, A. and A. Osborn.	Planning To Care. Brookfield, VT: Gower Publishing Company.
1998	Roseland, Mark.	Toward Sustainable Communities. New Haven, CT: New Society Publishers.
1979	Ross, P. J., H. Bluestone and F. K. Hines. 1979. Indicators of Social Well-Being for U.S.	Counties. Research Development Research Report No. 10, Economics, Statistics, and Cooperative Service, United States Department of Agriculture. Washington, D.C.: United States Department of Agriculture.
1993	Ryan, Michael.	Social Trends in Contemporary Russia: A Statistical Source-Book. New York: St. Martin's Press.
1982	Safir, Nadji.	Social Indicators of Development by the Year 2000: The Algerian Case. International Sociological Association.
1985	Schuessler, K.F. and G.A. Fisher.	"Quality of Life Research and Sociology." Annual Review of Sociology 11: 129-149.
1989	Schumacher, R., G. Sevens, T. O'Donnell, L. Torrence and K. Carney.	World Quality of Life Indicators. Santa Barbara, CA: ABC-CLIO.

Ano	Autor	Título
1998	Schyns, Peggy.	“Crossnational Difference in Happiness: Economic and Cultural Factors Explored.” Social Indicators Research 43.
1983	Scott, Wolf and N. T. Mathew.	A Development Monitoring Service at the Local Level, Volume 2: Levels of Living and Poverty in Kerala. Geneva: United Nations Research Institute for Social Development.
1973	Scott, Wolf, Helen Argalias and D.V. McGranahan.	The Measurement of Real Progress at the Local Level: Examples From the Literature and a Pilot Study. Geneva: United Nations Research Institute for Social Development.
1998	Segedy, James A. and Bradley E. Johnson.	The Neighborhood Charrette Handbook: Visioning & Visualizing Your Neighborhood's Future.
1985	Selen, Jan.	“Multidimensional Description of Social Indicators.” Social Indicators Research 17 (4):
1986	Shanks, J. Merrill.	Social Indicators: Bay Area Survey II. Ann Arbor, MI: Inter- University Consortium for Political and Social Research.
1968	Sheldon, E. B. and W. E. Moore (eds).	Indicators of Social Change: Concepts and Measurements. New York: Russell Sage Foundation.
1998	Sirgy, Joseph M.	Quality-of-Life Studies and Social Indicators Research: An Annotated Bibliography of Selected Works (1984-98). Blacksburg, VA: Department of Marketing, Pamplin College of Business, Virginia Tech University, The international Society for Quality-of-Life Studies.
1985	Smith, A. Wade.	“Social Class and Racial Cleavages on Major Social Indicators.” Research in Race and Ethnic Relations 4:
1973	Smith, David M.	The Geography of Social Well-being in the United States: An Introduction of Territorial Social Indicators. New York: McGraw-Hill Book Company.
1998	Smith, George, R.	“Are We Leaving the Community Out of Rural Community Sustainability? An examination of approaches to development and implementation of indicators of rural community sustainability and related public participation.” International Journal of

		Sustainable Development and World Ecology 5:
		Measuring Outcome in the Public Sector. London: Taylor and Francis.
1996	Smith, Peter, ed.	
1981	Smith, Tom W.	“Social Indicators: A Review Essay.” Journal of Social History 14
1980	Snider, E. L.	“Some Social Indicators for Multiple Family Housing.” Social Indicators Research 8
1982	Snider, Earle L.	“Social Indicators, Health Policy, and the Elderly.” Social Indicators Research 11
1981	Soderstrom, Edward Jonathan.	Social Impact Assessment: Experimental Methods and Approaches. New York: Praeger Press.
1986	Straus, Murray A.	“Cross-National Research on Gender Equality Using Sub- National Social Indicators.” International Sociological Association (Association Paper).

Ano	Autor	Título
1974	Strumpel, Burkhard, ed.	Subjective Elements of Well-being. Paris: Organization for Economic Cooperation and Development.
1976	Strumpel, Burkhard, ed.	Economic Means for Human Needs: Social Indicators of Well-Being and Discontent. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.
1976	Strumpel, Burkhard, Gerald Gurin and Richard T. Curtin.	Economic Incentives, Values and Subjective Well-being. Ann Arbor, MI: ISR Social Science Archive.
1991	Sullivan, Michael J. III.	Measuring Global Values: The Ranking of 162 Countries. New York: Greenwood Press.
1978	Taeuber, Conrad	“America in the Seventies: Some Social Indicators.” The Annals of the American Academy of Political and Social Sciences 435: 1-355.
1982	Taeuber, Richard C. and Richard C. Rockwell.	National Social Data Series: A Compendium of Brief Descriptions. Washington, D.C.: Social Science Research Council, Center for Coordination of Research.
1998	Tang, Kwong-Leung.	“East Asian Newly Industrializing Countries: Economic Growth and Quality of Life.” Social Indicators Research 43: 69-96.
1987	Tauber, Conrad	America Enters the Eighties: Some Social Indicators. In The Annals of the American Academy of Political and Social Science 453 (January 1981): 1-257.
1982	Taylor, C. L. and D. A. Jodice.	World Handbook of Political and Social Indicators: Volume I—Cross-National Attributes and Rates of Change. New Haven, CT.: Yale University Press.
1968	Taylor, Charles Lewis, ed.	Aggregate Data Analysis: Political and Social Indicators in Cross-National Research. Paris: Mouton and Co.
1980	Taylor, Charles Lewis, ed.	Indicator Systems for Political, Economic, and Social Analysis. Cambridge, MA: Oelgeschlager, Gunn and Hain.
1994	Tepperman, L. and S.J. Wilson.	Choices and Changes Sociology for Everyday Life. Boulder, CO: Harcourt Brace and Company Canada, Ltd.

1975	Terleckyj, Nestor.	
1998	Toupal, Rebecca S. and Michael D. Johnson.	Conservation Partnerships: Indicators of Success. Technical Report, Social Science Institute, Madison, WI: USDA-NRCS, Social Sciences Institute.
1985	Ujimoto, K. Victor.	“The Allocation of Time to Social and Leisure Activities as Social Indicators for the Integration of Aged Ethnic Minorities.” Social Indicators Research
1997	Valerio, C.	Elderly Americans: Where They Choose to Live. New York: Garland Publishing.
1995	Van Dieran, Wouter	Taking Nature into Account, Toward a Sustainable National Income: A Report of the Club of Rome. New York: Copernicus.
1997	Van Dijk, Frans.	Social Ties and Economic Performance. Dordrecht: The Netherlands.
1978	Van Vliet, Willem.	Theories, Methods and Applications of Activities Analysis. Monticello, IL:Vance Bibliographies.

Ano	Autor	Título
1983	Vance, Mary.	Social Indicators: A Revision of P-1259. Monticello, IL: Vance Bibliographies. 1983 1979
1983	Vance, Mary.	Social Indicators: Monographs. Public Administration Series. Bibliography #1259. Monticello, IL: Vance Bibliographies
1984	Verwayen, Henri.	“Social Indicators: Actual and Potential Uses.” Social Indicators Research 14
1983	Visbalingam, M.	“Area-based Social Indicators: Signed Chi-Squared as an Alternative to Ratios.” Social Indicators Research 13
1995	Waddell, Steve.	“Lessons from the Healthy Cities Movement for Social Indicator Development.” Social Indicators Research 34
1977	Warner, Sam Bass Jr. and Sylvia Fleisch.	Measurements for Social History. Beverly Hills: Sage Publications.
1993	Westendorff, David G. and Dharam Ghai, eds	Monitoring Social Progress in the 1990s: Data Constraints, Concerns and Priorities. Brookfield, VT: Avebury.
1978	Wilcox, L.	Methodology for Indicators of Social Development: Education. Supplemental Report 4: Education. Sociology Report 127-4. Ames, IA: Department of Sociology and Anthropology, Iowa State University.
1977	Wilcox, L.	A Methodology for Indicators for Social Development: Final Report. Social Report No. 127. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1972	Wilcox, L. D., R. M. Brooks, G. M. Beal and G. E. Klonglan.	Social Indicators and Societal Monitoring: An Annotated Bibliography. New York: Elsevier Scientific Publishing Company.
1974	Wilcox, L., W. A. McIntosh and J. Callaghan	A Methodology for Indicators of Social Development: Report 3: Toward an Integrated Social Information System. Sociology Report No. 117. Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1973	Wilcox, L., W. A. McIntosh, K. J. Byrnes, J. Callaghan, S. M. Hunter, S-M. Kim, J. Rowena.	A Methodology for Indicators of Social Development: Report 2: An Analysis of Selected Operational Indicators and Concepts. Sociology Report No. 116.

		Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1976	Wilcox, Leslie D., K. William Wassan, Frank A. Fear, Gerald E. Klonglan, George M. Beal.	Toward a Methodology for Social Indicators in Rural Development: Base Report. Sociology Report 125 (January 1975). Ames, IA: Department of Sociology and Anthropology, Iowa State University of Science and Technology.
1998	Williams, W.	Honest Numbers and Democracy: Social Policy Analysis in the White House, Congress, and the Federal Agencies. Washington, D.C.: Georgetown University Press.
1988	Wilson, Stephan M. and Gary W. Peterson.	"Life Satisfaction Among Young Adults from Rural Families." Family Relations 37: 84-91.

Ano	Autor	Título
1982	Wimberley, Ronald C. and James A. Christenson.	“Civil Religion, Social Indicators, and Public Policy.” Social Indicators Research 10
1983	Wolf, Herbert F.	“Subjective Social Indicators and the Socialist Way of Life.” Wissenschaftliche Zeitschrift der Karl Marx Universität Leipzig Gesellschafts Sprachwissenschaftliche Reihe 32
1995	Yoakam, Cy, ed.	Urban Quality Indicators. Ann Arbor, MI: Urban Quality Publications.
1997	Zamfir, E.	“Social Services for Children at Risk: the Impact on the Quality of Life.” Social Indicators Research: an International and Interdisciplinary Journal for Quality Measurement 42: 41-76.
1977	Zenhner, Robert.	Indicators of the Quality of Life in New Communities. Cambridge, MA: Ballinger Publishing
1998	Zvidrins, Peteris.	“Changes in Living Standards and Depopulation in Latvia in the 1990s.” Social Indicators Research 43: 121-140.
1999	State of Utah, Governor’s Office of Planning and Budget.	Economic Report to the Governor Social Indicators. Salt Lake City: Office of Demographic and Economic Analysis, Governor’s Office of Planning and Budget, State of Utah.
1998	Alliance for Community Education.	This Place Called Home; Tools for Sustainable Communities. New Haven, CT: New Society Publishers.
1997	America’s Children: Key National Indicators of Well-being.	Washington, D.C.: Federal Interagency Forum on Child and Family Statistics.
1999	Anchorage Department of Health and Human Services Community Health Promotion Program.	Healthy Anchorage Indicators: Helping Anchorage to Measure its Health and Quality of Life.
1998	Bahá’í International Community.	Initial Considerations Regarding the Creation of Spiritually Based Indicators for Development. London: “World Faiths and Development Dialogue.”
1998	Balance America 2100.	Sustainable Indicators Project. Rockville, MD: Balance America 2100.
1995	Brookings Institution.	Coping with Austerity: Poverty and Inequality in Latin

		America. Washington, D.C.: Brookings Institution.
1973	Bureau of Census.	Social and Health Indicator System. Washington, D.C.: Bureau of Economic Opportunity.
1973	Bureau of Census. United States Government.	Census Use Study: Social and Health Indicators System. Atlanta: Part II. Washington, D.C.: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Census.
1973	Bureau of Census. United States Government.	Census Use Study: Social and Health Indicators System. Los Angeles: Part II. Washington, D.C.: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Census.
1973	Bureau of Census. United States Government.	Census Use Study: Social and Health Indicators System. Part II: Rural—Mound Bayou, Mississippi. Washington, D.C.: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Census.
1973	Bureau of Census. United States Government.	Census Use Study: Social and Health Indicators System. Phoenix: Part I. Washington, D.C.: U.S. Department of Commerce, Social and Economic Statistics Administration, Bureau of Census.

Ano	Autor	Título
1998	Center of Excellence for Sustainable Development.	Boulder, CO: Department of Energy.
1999	City of Albuquerque.	Five-Year Goals Progress Indicators. Albuquerque, NM: City of Albuquerque.
1999	City of Austin, Texas and University of Texas, Department of Planning.	Sustainability Indicators Project (SIP). Austin, TX: City of Austin.
1996	Commonwealth of Independent States.	The Statistical Handbook of Social and Economic Indicators for the Former Soviet Union: 1980, 1990-1992. New York: Norman Ross Publication, Inc.
1996	Commonwealth of Virginia.	Performance Measures. Richmond, VA: Office of Governor George Allen.
1996	Community Planning Council.	Community Indicators: A Report Card on Greenville County. Greenville, SC: United Way of Greenville County.
1995	Concern, Inc.	Sustainability in Action: Profiles of Community Initiatives Across the United States. Washington, D.C.: Urban and Economic Development Policy Division of the U.S. Environmental Protection Agency.
1993	Corporation for Enterprise Development.	The Regional Performance Benchmarks System: Policymaker's Guide and User's Manual. Washington, D.C.: Corporation for Enterprise Development.
1986	D'Sa, E. R.	"Social Indicators and Spatial Disparities in Papua New Guinea." Social Indicators Research 18: 285-319.
1999	Data and Program Library Service.	UW-Madison DPLS Catalog of Holdings. Data and Program Library Service, University of Wisconsin-Madison.
1978	Department of Economic and Social Affairs, United Nations.	Systematic Monitoring and Evaluation of Integrated Development Programmes: A Source-Book. ST/ESA.78. New York: United Nations.
1992	Department of International Economic and Social Affairs, United Nations.	World Population Monitoring 1991: With Special Emphasis on Age Structure. Population Studies No. 126. ST/ESA/SER.A/126. New York: United Nations.
1978	Department of International Economic and Social Affairs.	Social Indicators: Preliminary Guidelines and Illustrative Series. Statistical Papers, Series M, No. 63. New York: United Nations.

1997	Developing Ideas Digest.	Issue # 9: Indicators of Sustainable Development. Approaches and Methods used in Longterm Social Planning and Policy-Making: Conclusions of the Seminar at Arc-et- Senans, France and Collection of Working Papers Submitted to the Seminar. New York: United Nations.
1973	Economic Commission for Europe.	United Nations.
1971	Economic Development Division, Economic Research Service, U. S. Department of Agriculture.	The Economic and Social Condition of Rural America in the 1970s. Washington, D.C.: Superintendent of Documents, U. S. Government Printing Office.
1975	Educational Testing Service Invitational Conference.	Educational Indicators: Monitoring the State of Education. Princeton, NJ: Educational Testing Service.
1998	Environment Canada and Canada Mortgage and Housing Corporation.	Sustainable Community Indicators Software.
1998	Environmental Protection Agency.	Index of Watershed Indicators.
1984	Family Services America.	The State of Families: 1984-1985. New York: Family Services America

Ano	Autor	Título
1998	Florida Keys National Marine Sanctuary.	Linking the Economy and Environment of Florida Keys/Florida Bay.
1999	Florida Sustainable Communities Center.	Focusing Sustainability Indicators: Workshop Helps City Measure Progress.
1997	Georgia Department of Community Affairs.	Community Indicators.
1997	Green Mountain Institute.	The Resource Guide to Indicators, First Edition. Montpelier, VT: Green Mountain Institute for Environmental Democracy.
1998	Green Mountain Institute.	GMI: The Resource Guide to Indicators: Second Edition. Montpelier, VT: The Green Mountain Institute,
1997	Hart Environmental Data.	Indicators of Sustainability.
1998	Indicators for a Sustainable San Mateo County: A Report Card of our County's Quality of Life.	San Mateo, CA: Sustainable San Mateo County.
1982	Indicators of Urban Condition.	Washington, D.C.: Government Finance Research Center, Municipal Finance Officers Association.
1997	International Institute for Sustainable Development.	Compendium of Sustainable Development Indicator Initiatives and Publications.
1995	International Union for the Conservation of Nature (IUCN).	A Sustainable World: Defining and Measuring Sustainable Development. Sacramento: Published for IUCN—the World Conservation Union by the International Center for the Environment and Public Policy, California Institute of Public Affairs. (
	Interrain Pacific and Ecotrust Columbia Pacific.	Columbia and Pacific Indicators: Assessing Community Sustainability for the Region. Interrain Pacific.
1996	Interrain.	Columbia-Pacific Indicators: Assessing Community Sustainability for the Region.
1997	Jacksonville Community Consortium, Inc.	Jacksonville Quality of Life Index. Jacksonville, FL: Jacksonville Community Consortium Inc.
1998	Kentucky Institute for the Environment and Sustainable Development.	SUSTAIN: A Journal Of Environmental And Sustainability Issues.
1998	Livable Tucson.	Livable Tucson Indicators Report.
1999	Louisiana State University.	Louisiana Population Data Center.

1998	Metro Washington Community Indicators Project.	State of the Community Report. Washington, D.C.: 1999 Integrated Strategies Forum,
1998	Minnesota Planning.	Minnesota Milestones 1998: Measures that Matter.
1997	National Key Centre for the Social Applications of Geographical Information Systems.	National Urban Indicators Project.
1998	New Zealand Ministry for the Environment.	Environmental Performance Indicators.
1999	North Carolina Department of Commerce.	North Carolina Consolidated Plan.
1999	North Dakota State University, Department of Agricultural Economics.	The North Dakota State Data Center. Fargo: North Dakota State University.
1973	Office of Management and Budget. Statistical Policy Division.	United States Government. 1973. Social Indicators, 1973: Selected Statistics on Social Conditions and Trends in the United States. Washington, D. C.: Superintendent of Documents, U.S. Government Printing Office.
	Office of Social and Economic Analysis, University of Missouri System.	Missouri State Census Data Center. Columbia, MO: University of Missouri.

Ano	Autor	Título
1982	Organization for Economic Cooperation and Development.	OECD Social Indicator Development Program Special Studies: McWhinnie, John R. Measuring Disability. Patrik Tornudd. Measuring Victimization. Dagfinn As. Measuring the Use of Time. And Catherine Pelli. Measuring Housing Conditions. Paris: OECD.
1977	Organization for Economic Cooperation and Development.	Basic Disaggregations of Main Social Indicators. OECD Special Studies No. 4. Paris: OECD.
1976	Organization for Economic Cooperation and Development.	Data Sources for Social Indicators on Victimization Suffered by Individuals. The OECD Social Indicators Program. Special Studies No. 3. Paris: OECD.
1973	Organization for Economic Cooperation and Development.	The OECD Social Indicator Program: List of Social Concerns Common to Most OECD Countries. Paris: OECD.
1997	Organization for Economic Cooperation and Development.	Better Understanding of our Cities: The Role of Urban Indicators. Paris: Publications Service, OECD.
1986	Organization for Economic Cooperation and Development.	Living Conditions in OECD Countries: A Compendium of Social Indicators. Social Policy Studies No. 3. Paris: OECD.
1982	Organization for Economic Cooperation and Development.	The OECD List of Social Indicators. Paris: OECD.
1976	Organization for Economic Cooperation and Development.	Measuring Social Well-being: A progress Report on the Development of Social Indicators. Paris: OECD.
1973	Organization for Economic Cooperation and Development.	The OECD Social Indicator Program: List of Social Concerns Common to Most OECD Countries. Paris:OECD.
1996	Passamaquoddy Tribe at Pleasant Point.	Passamaquoddy Community Vision, 1996. Perry, ME: White Owl Press.
1998	Pierce County Department of Community Services.	The Pierce County Quality of Life Benchmark Project. Tacoma, WA: Pierce County.
1980	President's Commission for a National Agenda for the	Report of the Panel on the Quality of American Life. 1980. The Quality of American Life in the Eighties. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office.
1960	President's Commission on National Goals.	Goals for Americans; programs for action in the sixties, comprising the report of the President's

		Commission on National Goals and chapters submitted for the consideration of the Commission. Englewood Cliffs, N.J.: Prentice-Hall.
1997	President's Council on Sustainable Development.	Sustainable Communities Task Force Report. Washington, D.C.: President's Council on Sustainable Development.
1997	Redefining Progress, Tyler Norris Associates, Sustainable Seattle.	The Community Indicators Handbook: Measuring Progress Toward Healthy and Sustainable Communities. San Francisco: Redefining Progress.
1998	Redefining Progress.	Community Indicators Project Directory.
1998	Redefining Progress.	The RP-CINet Listserve Archives.
1985	Research and Development Group, Iowa Development Commission.	Iowa's Quality of Life As Compared to the Nation. Des Moines: Iowa Development Commission.
1974	Research Committee, The Council of National Living.	Social Indicators in Japan. Tokyo, Japan: Printing Bureau, Ministry of Finance.
1996	Rural Economic Policy Program.	Measuring Community Capacity Building: A Workbook-in-Progress for Rural Communities. Queenstown, MD: The Aspen Institute.

Ano	Autor	Título
1997	Rural Partners.	State of Illinois: Strategic Plan for Rural Development. Springfield, IL: Rural Partners,
1964	Russett, B. M., H. R. Alker Jr., K. W. Deutsch and H. D. Lasswell.	World Handbook of Political and Social Indicators. New Haven, CT: Yale University Press.
1999	San Diego Natural History Museum— Environmental Science Education Center.	Regional Indicators Coordination Project.
1999	Santa Barbara South Coast Community Indicators Project.	Santa Barbara South Coast Community Indicators Report. 930 Miramont Dr., Santa Barbara, CA: Community Environmental Council.
1998	SINET.	Social Indicators Network News. Duke University.
1985	Social Monitoring Group.	From Birth to Death. Wellington, New Zealand: New Zealand Planning Council.
1980	Statistical Office of the European Communities.	Social Indicators for the European Community, 1960-1978 = Indicateurs sociaux pour la Communauté européenne. Washington, D.C.: European Community Information Service.
1983	Statistics Division, Economic and Social Development Department.	Community- Level Statistics. FAO Economic and Social Development Paper 33. Rome: Food and Agriculture Organization of the United Nations.
1998	Sustainable Communities Network.	Sustainability In Action: Profiles Of Community Initiatives Across The United States.
1995	Sustainable Seattle.	Indicators of Sustainable Community 1995: A Status Report on Long-term Cultural, Economic, and Environmental Health (Sustainable Seattle Indicators 1995). For ordering information contact Sustainable Seattle, metrocenter YMCA, 909 Fourth Ave., Seattle, WA 98104; (206) 382-5013
1998	The Canadian Council on Social Development.	Redefining Wealth and Progress: New Ways to Measure Economic, Social and Environmental Change.
1989	The Caracas Report on Alternative Development.	New York: The Bootstrap Press.
1998	The Ford Foundation.	“Exploring Our Future: Conservation-Based Development.”
1999	The Maine Economic Growth Council.	Measures of Growth
1999	The New York State Council on Children and	Kids Well-being Indicator Clearinghouse. Albany, NY:

- Families.
- 1999 The Oregon Progress Board and The Governor's Oregon Shines Task Force.
- 1998 The Puget Sound—Sustainable Community Roundtable.
- 1998 The Puget Sound—Sustainable Community Roundtable.
- 1998 The Sonoran Institute.
- 1996 The Thomas Jefferson Sustainability Council.
- 1995 The University of Wisconsin Extension.
- 1998 The Yale Social Science Statistical Laboratory.
- Center for Technology in Government.
- Oregon Progress Board Home Page. Salem, OR: Government of Oregon.
- State of the Community.
- State of the Community.
- The Community Stewardship Exchange.
- Indicators of Sustainability: Interim Report. Charlottesville, VA: Thomas Jefferson Planning District Commission.
- <http://monticello.avenue.gen.va.us/Gov/TJPDC>.
- “Promoting Agricultural and Business Competitiveness and a Sustainable Environment.”
- New Haven On Line Community

Ano	Autor	Título
1978	U.S. Commission on Civil Rights.	Social Indicators of Equality for Minorities and Women: A Report of the United States Commission on Civil Rights. Washington, D.C.: Commission on Civil Rights.
1978	U.S. Department of Housing and Urban Development.	The 1978 HUD Survey on the Quality of Community Life: Data Book. U.S. Department of Housing and Urban Development. Office of Policy Development and Research.
1991	United Nations	Compendium of Social Statistics and Indicators. New York,: The Statistical Office, Department of International Economic and Social Affairs of the United Nations.
1984	United Nations Economic and Social Council (UNESCO).	Social Science Methods, Decision-making and Development Planning. Paris: UNESCO.
1988	United Nations.	Improving Statistics and Indicators on Women Using Household Surveys. New York: United Nations Department of International Economic and Social Affairs Statistical Office and the International Research and Training Institute for the Advancement of Women. Series F, No. 48.
1984	United Nations.	Improving Concepts and Methods for Statistics and Indicators on the Situation of Women. New York: United Nations Department of International Economic and Social Affairs Statistical Office and the International Research and Training Institute for the Advancement of Women. Series F. No. 33.
1978	United Nations.	Social Indicators: Preliminary Guidelines an Illustrative Series. New York: United Nations. (
1995	United Nations.	The World's Women: Trends and Statistics. New York: United Nations.
1994	United Nations.	World Social Situation in the 1990s. New York: United Nations.
1988	United Nations.	Handbook on Social Indicators. New York: United Nations Department of International Economic and Social Affairs. Series F. No.
1984	United Nations.	Compiling Social Indicators on the Situation of Women. New York: United Nations Department of

- International Economic and Social Affairs Statistical Office and International Research and Training Institute for the Advancement of Women. Series F. No. 32.
- Demographic Indicators of Countries: Estimates and Projections as Assessed in 1980. New York: United Nations.
- 1982 United Nations.
United States Agency for International Development.
- 1998 Development.
United States Agency for International Development.
- 1972 Development.
United States Agency for International Development.
- 1980 United States Department of Commerce.
Bureau of the Census. United States Government. 1980. Social Indicators III: Social Indicators, 1980: Selected statistics on social conditions and trends in the United States. Washington, D. C.: Superintendent of Documents, U.S. Government Printing Office.

Ano	Autor	Título
1976	United States Department of Commerce. Bureau of the Census.	United States Government. 1976. Social Indicators II: Social Indicators, 1976: Selected Statistics on Social Conditions and Trends in the United States. Washington, D. C.: Superintendent of Documents, U.S. Government Printing Office.
1969	United States Department of Health, Education, and Welfare.	Toward a Social Report. Washington, D.C.: Superintendent of Documents, U.S. Government Printing Office.
1971	Universal Engineering Corporation. University of Mississippi.	Social Characteristics of Neighborhoods as Indicators of the Effects of Highway Improvements. Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration. State Data Center of Mississippi. Jackson, MS: University of Mississippi.
1998	University of Washington—Department of Public Health.	The MCBES Project: Methodology of Community-Based Evaluation Studies. Seattle: University of Washington, National Cancer Institute.
1998	University of Wisconsin-Extension Cooperative Extension.	Sustainable Community Development Helping the Future of Wisconsin's Communities. Madison, WI: University of Wisconsin-Extension.
1996	Urban Sustainability Learning Group.	Staying in the Game: Exploring the Options for Urban Sustainability. Chicago, IL: The Tides Center.
1998	Western Rural Development Center.	Coping With Change.
1998	Willapa Indicators for a Sustainable Community.	The 1998 Willapa Indicators. Big Bend, OR: Willapa Alliance/Ecotrust.
1992	Wistat: Women's Indicators and Statistics Database (Version 2).	New York: United Nations.
1997	Wistat: Women's Indicators and Statistics Database (Version 3).	New York: United Nations.
1998	Working Partnerships.	Community Economic Indicators—Growing Together or Drifting Apart? Working Families and Business in the New Economy. Silicon Valley at Work.
1989	World Bank.	World Tables of Economic and Social Indicators, 1950-1987. ICPSR 9300. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research.

- 1996 World Bank. Social Indicators of Development. Baltimore, MD: Johns Hopkins Press.
- 1976 World Bank. World Tables: Volume II- Social Data. Baltimore, MD: Johns Hopkins Press.
- 1998 World Bank. World Development Indicators.
- 1998 World Resources Institute. Developing Ecological Materials Flow Indicators. Yampa Valley Partners-Community Indicators Project: 1997 Report. Craig, CO: Yampa Valley Partners
- 1997 Yampa Valley Partners. Colorado Healthy Communities Initiative, Community Indicators Project.
- 1997 Yampa Valley Partners. Report to the Adirondack Council on Community Capacity Indicators. St. Albans, VT: Yellow Wood Associates.
- 1994 Yellow Wood Associates, Inc.

Apêndice C – Publicações Internacionais Por Assunto

Tabela 1 - Publicações internacionais sobre indicadores sociais por assunto, 1960-1999

Ano	Teóricos	Metodológicos	Políticas Públicas	Índices e Indicadores	Qualidade de Vida	Padrões de Vida	Bibliografia comentada	Internet
1960	1	-	-	-	-	-	-	-
1964	-	-	-	1	-	-	-	-
1966	1	-	-	-	-	-	-	-
1968	-	-	2	-	-	-	-	-
1969	1	-	-	-	-	-	-	-
1970	-	-	1	-	-	-	-	-
1971	-	-	1	-	2	-	1	-
1972	-	-	-	-	2	1	2	-
1973	-	-	2	1	9	1	-	-
1974	3	-	5	1	-	-	-	-
1975	1	-	2	1	1	4	1	-
1976	1	-	6	-	4	4	1	-
1977	-	-	5	1	8	1	-	-
1978	-	-	4	-	5	-	2	-
1979	1	-	-	1	2	1	1	-
1980	3	-	3	2	4	1	-	-
1981	3	-	5	5	2	1	-	-
1982	2	-	6	3	5	1	-	-
1983	3	-	5	1	2	2	2	-
1984	3	-	7	1	2	-	-	-
1985	-	-	7	2	6	2	-	-
1986	-	-	2	-	4	2	-	-
1987	-	-	-	-	1	-	-	-
1988	-	-	2	-	2	-	-	-
1989	1	-	1	-	2	1	-	-
1990	-	-	2	1	3	-	-	-
1991	-	-	-	-	4	-	-	-
1992	-	-	3	-	5	-	-	-
1993	-	-	4	-	3	-	-	-
1994	1	-	1	-	5	2	-	-
1995	4	-	2	-	10	-	1	1
1996	2	-	4	1	13	-	1	1
1997	2	-	3	-	19	6	-	6
1998	1	-	8	1	12	5	2	25
1999	-	-	1	-	10	-	-	6

Fonte: Elaboração do autor

APENDICE D – Sistemas e Índices Nacionais de Desenvolvimento por categorias de análise, 1991 – 2005

Descrição	Número de variáveis (componentes)	Método de Seleção	Escala	Ponderação	Agregação	Cobertura (Países)
1. Combined Consumption Level Index (Bennett, 1951)	19(6)	<i>ad hoc</i>	0-100 medidas de posição	média aritmética	aditivo	31
2. Human Resources Development Index (HRDI) (Harbison and Myers, 1964)	2	<i>ad hoc</i>	nenhuma	<i>ad hoc</i>	aditivo	75
3. Real Index of Consumption (Beckerman and Bacon, 1966)	7	<i>ad hoc</i>	nenhuma	coeficientes de correlação	funcional	80
4. UNRISD General Index of Development (GID) (McGranahan et al., 1972)	18(2)	matriz de correlações (73)	0-100 medidas de posição	coeficientes de correlação	aditivo	58
5. Physical Quality of Life Index (PQLI) (Morris, 1979)	3	<i>ad hoc</i>	0-100 medidas de posição	média aritmética	aditivo	150
6. Composite Basic Needs Indices (Ram, 1982)	5	<i>ad hoc</i>	nenhum e 0-100 medidas de posição	análise de componentes principais	aditivo	82
7. Index of Social Progress (ISP) (Estes, 1984)	44(11)	<i>ad hoc</i>	escore padronizado	média aritmética	aditivo	107

Descrição	Número de variáveis (componentes)	Método de Seleção	Escala	Ponderação	Agregação	Cobertura (Países)
8. World Standard Distance Scales (Ginsberg et al., 1986)	42/22(5)	<i>ad hoc</i>	desvio padrão	média aritmética	aditivo	143
9. Human Suffering Index (HSI) (Camp and Speidel, as related in Hess, 1989)	10	<i>ad hoc</i>	0-10 contínua	média aritmética	aditivo	mais de 60
10. Quality of Life Rankings (Slottje, 1991)	20	<i>ad hoc</i>	nenhuma	misto	funcional	126
11. Combined Quality of Life Indices (CQLI) (Diener, 1995)	14(7)	<i>ad hoc</i>	escore padronizado	média aritmética	aditivo	77
12. Index of Economic Freedom (Johnston and Sheehy, 1995)	10	<i>ad hoc</i>	1-5 contínua	média aritmética	aditivo	mais de 100
13. Economic Freedom Indices (EFIs) (Gwartney et al., 1996)	17(4)	<i>ad hoc</i>	0-10 contínua	pesos iguais e <i>ad hoc</i>	aditivo	103
14. Human Development Index (HDI) (UNDP, 1996)	4(3)	<i>ad hoc</i>	0-100 medidas de posição	média aritmética e <i>ad hoc</i>	aditivo	174
15. Capability Poverty Measure (CPM) (UNDP, 1996)	3	<i>ad hoc</i>	nenhuma	média aritmética	aditivo	101

Descrição	Número de variáveis (componentes)	Método de Seleção	Escala	Ponderação	Agregação	Cobertura (Países)
16. Gender-related Development Index (GDI) (UNDP, 1996)	4(3)	<i>ad hoc</i>	ajustado 0-100 medidas de posição	média aritmética	aditivo	163
17. Gender Empowerment Measure (GEM) (UNDP, 1996)	4(3)	<i>ad hoc</i>	ajustado 0-100 medidas de posição	média aritmética	aditivo	100
18. Global Competitiveness Indices (World Economic Forum, 1996)	155(8)	cobertura de dados (300+)	escore padronizado	<i>ad hoc</i> e coeficientes de correlação	aditivo	49
19. Human Poverty Index (HPI-1) for developing countries (UNDP, 1999)	5(3)	<i>ad hoc</i>	nenhuma	média aritmética	funcional	92
20. Human Poverty Index (HPI-2) for developed countries (UNDP, 1999)	4	<i>ad hoc</i>	nenhuma	média aritmética	funcional	17

Nome dos Municípios	Variações dos nomes dos municípios
Barra do Rocha	Barra do Rocha
Barreiras	Barreiras
Barro Alto	Barro Alto
Barro Preto	Barro Preto
Barro Preto	Barro Preto/Lomanto Junior
Barro Preto	Gov Lomanto Jr
Barro Preto	Gov Lomanto Jr.
Barro Preto	Gov Lomanto Junior
Barro Preto	Gov Lomanto Júnior
Barro Preto	Gov. Lomanto Jr
Barro Preto	Gov. Lomanto Jr.
Barro Preto	Gov. Lomanto Junior
Barro Preto	Gov. Lomanto Júnior
Barro Preto	Governador Lomanto Jr
Barro Preto	Governador Lomanto Jr.
Barro Preto	Governador Lomanto Junior
Barro Preto	Governador Lomanto Júnior
Barro Preto	Lomanto Junior
Barro Preto	Lomanto Júnior
Barrocas	Barroca
Barrocas	Barrocas
Belmonte	Belmonte
Belo Campo	Belo Campo
Biritinga	Biritinga
Boa Nova	Boa Nova
Boa Vista do Tupim	Boa Vista do Tupim
Boa Vista do Tupim	Boa Vista Tupim
Bom Jesus da Lapa	Bom J da Lapa
Bom Jesus da Lapa	Bom J. da Lapa
Bom Jesus da Lapa	Bom Jesus da Lapa
Bom Jesus da Serra	Bom J da Serra

Nome dos Municípios	Variações dos nomes dos municípios
Bom Jesus da Serra	Bom J. da Serra
Bom Jesus da Serra	Bom Jesus da Serra
Boninal	Boninal
Bonito	Bonito
Boquira	Boquira
Botuporã	Botupora
Botuporã	Botuporã
Botuporã	Botuporan
Brejões	Brejoes
Brejões	Brejões
Brejolândia	Brejolândia
Brejolândia	Brejolândia
Brotas de Macaúbas	Brotas de Macaubas
Brotas de Macaúbas	Brotas de Macaúbas
Brumado	Brumado
Buerarema	Buerarema
Buritirama	Buritirama
Caatiba	Caatiba
Caatiba	Catiba
Cabaceiras do Paraguaçu	Cabac. do Paraguaçu
Cabaceiras do Paraguaçu	Cabac. do Paraguaçu
Cabaceiras do Paraguaçu	Cabaceira do Paraguaçu
Cabaceiras do Paraguaçu	Cabaceiras do Paraguaçu
Cabaceiras do Paraguaçu	Cabaceiras do Paraguaçu
Cabaceiras do Paraguaçu	Cabeceira de Paraguaçu
Cabaceiras do Paraguaçu	Cabeceira do Paraguaçu
Cachoeira	Cachoeira
Caculé	Cacule
Caculé	Caculé
Caém	Anselmo da Fonseca
Caém	Caem

Nome dos Municípios	Variações dos nomes dos municípios
Cocos	Cocos
Cocos	Côcos
Conceição da Feira	Conc. de Feira
Conceição da Feira	Conceicao da Feira
Conceição da Feira	Conceição da Feira
Conceição da Feira	Conceicao de Feira
Conceição da Feira	Conceição de Feira
Conceição da Feira	Conceição de Feira
Conceição do Almeida	Conc. do Almeida
Conceição do Almeida	Conceicao Almeida
Conceição do Almeida	Conceição Almeida
Conceição do Almeida	Conceicao do Almeida
Conceição do Almeida	Conceição do Almeida
Conceição do Coité	Conceicao do Coite
Conceição do Coité	Conceição do Coite
Conceição do Coité	Conceição do Coité
Conceição do Coité	Conceição do Coité
Conceição do Jacuípe	Conc. do Jacuípe
Conceição do Jacuípe	Conc. do Jacuípe
Conceição do Jacuípe	Conceicao do Jacuípe
Conceição do Jacuípe	Conceição do Jacuípe
Conceição do Jacuípe	Conceição do Jacuípe
Conceição do Jacuípe	Conceição do Jacuípe
Conde	Conde
Condeúba	Condeuba
Condeúba	Condeúba
Contendas do Sincorá	Contendas do Sincora
Contendas do Sincorá	Contendas do Sincorá
Contendas do Sincorá	Contendas Sincora
Contendas do Sincorá	Contendas Sincorá

Nome dos Municípios	Variações dos nomes dos municípios
Coração de Maria	Coracao de Maria
Coração de Maria	Coração de Maria
Cordeiros	Cordeiros
Coribe	Coribe
Coronel João Sá	Coronel Joao Sa
Coronel João Sá	Coronel Joao Sá
Coronel João Sá	Coronel João Sa
Coronel João Sá	Coronel João Sá
Correntina	Correntina
Cotegipe	Cotegipe
Cotegipe	Cotejipe
Cravolândia	Cravolandia
Cravolândia	Cravolândia
Crisópolis	Crisopolis
Crisópolis	Crisópolis
Crisópolis	Crisópolis
Cristópolis	Cristopolis
Cristópolis	Cristópolis
Cruz das Almas	Crus das Almas
Cruz das Almas	Cruz das Almas
Curaçá	Curaca
Curaçá	Curaçá
Dário Meira	Dario Meira
Dário Meira	Dário Meira
Dias D'Ávila	Dias D Avila
Dias D'Ávila	Dias D Ávila
Dias D'Ávila	Dias D'Avila
Dias D'Ávila	Dias D'Ávila
Dias D'Ávila	Dias D'Ávila
Dias D'Ávila	Dias D'Avila
Dias D'Ávila	Dias D'Ávila

Nome dos Municípios	Variações dos nomes dos municípios
Dias D'Ávila	Dias D'Ávila
Dom Basílio	Dom Basilio
Dom Basílio	Dom Basílio
Dom Macedo Costa	Dom Macedo Costa
Dom Macedo Costa	Dom Macêdo Costa
Elísio Medrado	Elisio Medrado
Elísio Medrado	Elísio Medrado
Encruzilhada	Encruzilhada
Entre Rios	Entre Rios
Érico Cardoso	Água Quente
Érico Cardoso	Água Quente
Érico Cardoso	Erico Cardoso
Érico Cardoso	Érico Cardoso
Esplanada	Esplanada
Euclides da Cunha	Euclides da Cunha
Eunápolis	Eunapolis
Eunápolis	Eunápolis
Eunápolis	Eunapólis
Fátima	Fatima
Fátima	Fátima
Feira da Mata	Feira da Mata
Feira de Santana	Feira de Santana
Filadélfia	Filadelfia
Filadélfia	Filadélfia
Firmino Alves	Firmino Alves
Floresta Azul	Floresta Azul
Formosa do Rio Preto	Formosa do Rio Preto
Formosa do Rio Preto	Formosa Rio Preto
Gandu	Gandu
Gandu	Gandú
Gavião	Gaviao

Nome dos Municípios	Variações dos nomes dos municípios
Gavião	Gavião
Gentio do Ouro	Gentio de Ouro
Gentio do Ouro	Gentio do Ouro
Glória	Gloria
Glória	Glória
Gongogi	Gongogi
Gongogi	Gongoji
Governador Mangabeira	Gov Mangabeira
Governador Mangabeira	Gov. Mangabeira
Governador Mangabeira	Governador Mangabeira
Guajeru	Guajeru
Guajeru	Guajerú
Guanambi	Guanambi
Guaratinga	Guaratinga
Heliópolis	Heliopolis
Heliópolis	Heliópolis
Heliópolis	Heliópolis
laçu	lacu
laçu	lacú
laçu	laçu
laçu	laçu
Ibiassucê	Ibiassuce
Ibiassucê	Ibiassucê
Ibicaraí	Ibicarai
Ibicaraí	Ibicaraí
Ibicoara	Ibicoara
Ibicuí	Ibicui
Ibicuí	Ibicuí
Ibipeba	Ibipeba
Ibipitanga	Ibipitanga
Ibiquera	Ibiquera

Nome dos Municípios	Variações dos nomes dos municípios
Ibirapitanga	Ibirapitanga
Ibirapoã	Ibirapoa
Ibirapoã	Ibirapoã
Ibirapoã	Ibirapoan
Ibirapoã	Ibirapoan
Ibirapoã	Ibirapua
Ibirapoã	Ibirapuã
Ibirapoã	Ibirapuan
Ibirataia	Ibirataia
Ibitiara	Ibitiara
Ibititá	Ibitita
Ibititá	Ibititá
Ibotirama	Ibotirama
Ichu	Ichu
Ichu	Ichú
Igaporã	Igapora
Igaporã	Igaporã
Igaporã	Igaporan
Igaporã	Igapurã
Igaporã	Igapuran
Igrapiúna	Igrapiuna
Igrapiúna	Igrapiúna
Iguaí	Iguai
Iguaí	Iguaí
Ilhéus	Ilheus
Ilhéus	Ilhéus
Inhambupe	Inhambupe
Ipecaetá	Ipecaeta
Ipecaetá	Ipecaetá
Ipiaú	Ipiaú
Ipiaú	Ipiaú

Nome dos Municípios	Variações dos nomes dos municípios
Ipirá	Ipira
Ipirá	Ipirá
Ipupiara	Ipupiara
Irajuba	Irajuba
Iramaia	Iramaia
Iraquara	Iraquara
Irará	Irara
Irará	Irará
Irecê	Irece
Irecê	Irecê
Irecê	Irecê
Itabela	Itabela
Itaberaba	Itaberaba
Itabuna	Itabuna
Itacaré	Itacare
Itacaré	Itacaré
Itaeté	Itaete
Itaeté	Itaeté
Itaeté	Itaetê
Itaeté	Itaite
Itaeté	Itaité
Itaeté	Itaitê
Itagi	Itagi
Itagibá	Itagiba
Itagibá	Itagibá
Itagibá	Itajiba
Itagibá	Itajibá
Itagimirim	Itagimirim
Itaguaçu da Bahia	Itaguacu da Bahia
Itaguaçu da Bahia	Itaguacú da Bahia
Itaguaçu da Bahia	Itaguaçu da Bahia

Nome dos Municípios	Variações dos nomes dos municípios
Itaguaçu da Bahia	Itaguaçu da Bahia
Itaguaçu da Bahia	Itaguacu da Baia
Itaguaçu da Bahia	Itaguacú da Baia
Itaguaçu da Bahia	Itaguaçu da Baia
Itaguaçu da Bahia	Itaguaçu da Baia
Itaju do Colônia	Itaju do Colonia
Itaju do Colônia	Itaju do Colônia
Itaju do Colônia	Itajú do Colônia
Itajuípe	Itajuípe
Itajuípe	Itajuípe
Itamaraju	Itamaraju
Itamaraju	Itamarajú
Itamari	Itamari
Itambé	Itambe
Itambé	Itambé
Itanagra	Itanagra
Itanhém	Itanhem
Itanhém	Itanhém
Itaparica	Itaparica
Itapé	Itape
Itapé	Itapé
Itapebi	Itapebi
Itapetinga	Itapetinga
Itapicuru	Itapicuru
Itapicuru	Itapicurú
Itapitanga	Itapitanga
Itaquara	Itaquara
Itarantim	Itarantim
Itatim	Itatim
Itiruçu	Itirucu
Itiruçu	Itiruçu

Nome dos Municípios	Variações dos nomes dos municípios
Itiúba	Itiuba
Itiúba	Itiúba
Itororó	Itororo
Itororó	Itororó
Ituaçu	Ituacu
Ituaçu	Ituaçu
Ituaçu	Ituaçu
Ituberá	Itubera
Ituberá	Ituberá
Iuiu	Iuiu
Iuiu	Iuiu
Jaborandi	Jaborandi
Jacaraci	Jacaraci
Jacobina	Jacobina
Jaguaquara	Jaguaquara
Jaguarari	Jaguarari
Jaguaripe	Jaguaripe
Jandaíra	Jandaira
Jandaíra	Jandaíra
Jequié	Jequie
Jequié	Jequié
Jeremoabo	Jeremoabo
Jiquiriçá	Jequirica
Jiquiriçá	Jequiriçá
Jiquiriçá	Jiquirica
Jiquiriçá	Jiquiriça
Jiquiriçá	Jiquiriçá
Jitaúna	Jitauna
Jitaúna	Jitaúna
João Dourado	Joao Dourado
João Dourado	João Dourado

Nome dos Municípios	Variações dos nomes dos municípios
Luís Eduardo Magalhães	Deputado Luíz Eduardo Magalhães
Luís Eduardo Magalhães	Luis Eduardo Magalhaes
Luís Eduardo Magalhães	Luis Eduardo Magalhães
Luís Eduardo Magalhães	Luís Eduardo Magalhães
Luís Eduardo Magalhães	Luiz Eduardo Magalhaes
Luís Eduardo Magalhães	Luiz Eduardo Magalhães
Luís Eduardo Magalhães	Luíz Eduardo Magalhães
Macajuba	Macajuba
Macarani	Macarani
Macarani	Macaraní
Macaúbas	Macaubas
Macaúbas	Macaúbas
Macururé	Macurure
Macururé	Macururé
Madre de Deus	Madre de Deus
Maetinga	Maetinga
Maetinga	Maitinga
Maiquinique	Maiquinique
Mairi	Mairi
Mairi	Mairí
Malhada	Malhada
Malhada de Pedras	Malhada de Pedras
Manoel Vitorino	Manoel Vitorino
Mansidão	Mansidao
Mansidão	Mansidão
Maracás	Maracas
Maracás	Maracás
Maragogipe	Maragogipe
Maraú	Marau
Maraú	Maraú
Marcionílio Souza	Marcionilio Sousa

Nome dos Municípios	Variações dos nomes dos municípios
Marcionílio Souza	Marcionílio Sousa
Marcionílio Souza	Marcionilio Souza
Marcionílio Souza	Marcionílio Souza
Mascote	Mascote
Mata de São João	Mata de Sao Joao
Mata de São João	Mata de Sao João
Mata de São João	Mata de São Joao
Mata de São João	Mata de São João
Matina	Matina
Medeiros Neto	Medeiros Neto
Miguel Calmon	Miguel Calmon
Milagres	Milagres
Mirangaba	Mirangaba
Mirante	Mirante
Monte Santo	Monte Santo
Morpará	Morpara
Morpará	Morpará
Morpará	Morpara
Morro do Chapéu	Morro do Chapeu
Morro do Chapéu	Morro do Chapéu
Mortugaba	Mortugaba
Mucugê	Mucuge
Mucugê	Mucugê
Mucuri	Mucuri
Mulungu do Morro	Mulungu do Morro
Mundo Novo	Mundo Novo
Muniz Ferreira	Muniz Ferreira
Muquém do São Francisco	Muquem de Sao Francisco
Muquém do São Francisco	Muquem de São Francisco
Muquém do São Francisco	Muquém de São Francisco
Muquém do São Francisco	Muquem do Sao Francisco

Nome dos Municípios	Variações dos nomes dos municípios
Muquém do São Francisco	Muquem do São Francisco
Muquém do São Francisco	Muquém do São Francisco
Muquém do São Francisco	Muquem S. Francisco
Muquém do São Francisco	Muquém S. Francisco
Muquém do São Francisco	Muquem Sao Francisco
Muquém do São Francisco	Muquem Sao Francisco
Muquém do São Francisco	Muquem São Francisco
Muquém do São Francisco	Muquem São Francisco
Muquém do São Francisco	Muquém São Francisco
Muquém do São Francisco	Muquém São Francisco
Muritiba	Muritiba
Mutuípe	Mutuipe
Mutuípe	Mutuípe
Nazaré	Nazare
Nazaré	Nazaré
Nilo Peçanha	Nilo Pecanha
Nilo Peçanha	Nilo Peçanha
Nordestina	Nordestina
Nova Canaã	Nova Canaa
Nova Canaã	Nova Canaã
Nova Fátima	Nova Fatima
Nova Fátima	Nova Fátima
Nova Ibiá	Nova Ibia
Nova Ibiá	Nova Ibiá
Nova Itarana	Nova Itarana
Nova Redenção	Nova Redencao
Nova Redenção	Nova Redenção
Nova Redenção	Nova Redenção
Nova Soure	Nova Soure
Nova Viçosa	Nova Vicosa
Nova Viçosa	Nova Viçosa

Nome dos Municípios	Variações dos nomes dos municípios
Novo Horizonte	Novo Horizonte
Novo Triunfo	Novo Triunfo
Olindina	Olindina
Oliveira dos Brejinhos	Oliveira Brejinhos
Oliveira dos Brejinhos	Oliveira dos Brejinhos
Ouriçangas	Ouricangas
Ouriçangas	Ouriçangas
Ourolândia	Ourolandia
Ourolândia	Ourolândia
Palmas de Monte Alto	Palmas de Monte Alto
Palmas de Monte Alto	Palmas do Monte Alto
Palmas de Monte Alto	Palmas Monte Alto
Palmeiras	Palmeiras
Paramirim	Paramirim
Paratinga	Paratinga
Paripiranga	Paripiranga
Pau Brasil	Pau Brasil
Paulo Afonso	Paulo Afonso
Pé de Serra	Pe de Serra
Pé de Serra	Pé de Serra
Pedrao	Pedrao
Pedrao	Pedrao
Pedro Alexandre	Pedro Alexandre
Piatã	Piata
Piatã	Piatã
Piatã	Piatan
Pilão Arcado	Pilao Arcado
Pilão Arcado	Pilão Arcado
Pindaí	Ouro Branco
Pindaí	Pindai
Pindaí	Pindaí

Nome dos Municípios	Variações dos nomes dos municípios
Pindobaçu	Pindobacu
Pindobaçu	Pindobaçu
Pindobaçu	Pindobaçú
Pintadas	Pintadas
Piraí do Norte	Pirai do Norte
Piraí do Norte	Piraí do Norte
Piripá	Piripa
Piripá	Piripá
Piritiba	Piritiba
Planaltino	Planaltino
Planalto	Planalto
Poções	Pocoos
Poções	Pocões
Poções	Poções
Pojuca	Pojuca
Ponto Novo	Ponto Novo
Porto Seguro	Porto Seguro
Potiraguá	Potiragua
Potiraguá	Potiraguá
Prado	Prado
Presidente Dutra	Pres Dutra
Presidente Dutra	Pres. Dutra
Presidente Dutra	Presidente Dutra
Presidente Jânio Quadros	Pres Janio Quadros
Presidente Jânio Quadros	Pres Jânio Quadros
Presidente Jânio Quadros	Pres. Janio Quadros
Presidente Jânio Quadros	Pres. Jânio Quadros
Presidente Jânio Quadros	Presidente Janio Quadros
Presidente Jânio Quadros	Presidente Jânio Quadros
Presidente Tancredo Neves	Pres Tancredo Neves
Presidente Tancredo Neves	Pres. Tancredo Neves

Nome dos Municípios	Variações dos nomes dos municípios
Presidente Tancredo Neves	Presidente Tancredo
Presidente Tancredo Neves	Presidente Tancredo Neves
Queimadas	Queimadas
Quijingue	Quijingue
Quixabeira	Quichabeira
Quixabeira	Quixabeira
Rafael Jambeiro	Rafael Jambeiro
Remanso	Remanso
Retirolândia	Retirolandia
Retirolândia	Retirolândia
Riachão das Neves	Riachao das Neves
Riachão das Neves	Riachão das Neves
Riachão do Jacuípe	Riachao do Jacuipe
Riachão do Jacuípe	Riachão do Jacuipe
Riachão do Jacuípe	Riachão do Jacuípe
Riacho de Santana	Riacho de Santana
Ribeira do Amparo	Ribeira do Amparo
Ribeira do Pombal	Ribeira do Pombal
Ribeirão do Largo	Rib do Largo
Ribeirão do Largo	Rib. do Largo
Ribeirão do Largo	Ribeirao do Largo
Ribeirão do Largo	Ribeirão do Largo
Rio de Contas	Rio de Contas
Rio do Antônio	Rio do Antonio
Rio do Antônio	Rio do Antônio
Rio do Pires	Rio do Pires
Rio Real	Rio Real
Rodelas	Rodelas
Ruy Barbosa	Rui Barbosa
Ruy Barbosa	Ruy Barbosa
Salinas da Margarida	Salina da Margarida

Nome dos Municípios	Variações dos nomes dos municípios
Salinas da Margarida	Salinas da Margarida
Salinas da Margarida	Salinas das Margaridas
Salinas da Margarida	Salinas Margarida
Salvador	Salvador
Santa Bárbara	Santa Barbara
Santa Bárbara	Santa Bárbara
Santa Bárbara	Sta Barbara
Santa Bárbara	Sta Bárbara
Santa Bárbara	Sta. Barbara
Santa Bárbara	Sta. Bárbara
Santa Brígida	Santa Brigida
Santa Brígida	Santa Brígida
Santa Brígida	Sta Brigida
Santa Brígida	Sta Brígida
Santa Brígida	Sta. Brigida
Santa Brígida	Sta. Brígida
Santa Cruz Cabrália	Santa Cruz Cabralia
Santa Cruz Cabrália	Santa Cruz Cabrália
Santa Cruz Cabrália	Santa Cruz de Cabralia
Santa Cruz Cabrália	Santa Cruz de Cabrália
Santa Cruz Cabrália	Sta Cruz Cabralia
Santa Cruz Cabrália	Sta Cruz Cabrália
Santa Cruz Cabrália	Sta Cruz de Cabralia
Santa Cruz Cabrália	Sta Cruz de Cabrália
Santa Cruz Cabrália	Sta. Cruz Cabralia
Santa Cruz Cabrália	Sta. Cruz Cabrália
Santa Cruz Cabrália	Sta. Cruz de Cabralia
Santa Cruz da Vitória	Santa Cruz da Vitoria
Santa Cruz da Vitória	Santa Cruz da Vitória
Santa Cruz da Vitória	Santa Cruz Vitoria

Nome dos Municípios	Variações dos nomes dos municípios
Santa Cruz da Vitória	Santa Cruz Vitória
Santa Cruz da Vitória	Sta Cruz da Vitoria
Santa Cruz da Vitória	Sta Cruz da Vitória
Santa Cruz da Vitória	Sta Cruz Vitoria
Santa Cruz da Vitória	Sta Cruz Vitória
Santa Cruz da Vitória	Sta. Cruz da Vitoria
Santa Cruz da Vitória	Sta. Cruz da Vitória
Santa Cruz da Vitória	Sta. Cruz Vitoria
Santa Cruz da Vitória	Sta. Cruz Vitória
Santa Inês	Santa Ines
Santa Inês	Santa Inês
Santa Luzia	Santa Lusia
Santa Luzia	Santa Luzia
Santa Maria da Vitória	Santa Maria da Vitoria
Santa Maria da Vitória	Santa Maria da Vitória
Santa Maria da Vitória	Santa Maria Vitoria
Santa Maria da Vitória	Santa Maria Vitória
Santa Maria da Vitória	Sta Maria da Vitoria
Santa Maria da Vitória	Sta Maria da Vitória
Santa Maria da Vitória	Sta Maria Vitoria
Santa Maria da Vitória	Sta Maria Vitória
Santa Maria da Vitória	Sta. Maria da Vitoria
Santa Maria da Vitória	Sta. Maria da Vitória
Santa Maria da Vitória	Sta. Maria Vitoria
Santa Maria da Vitória	Sta. Maria Vitória
Santa Rita de Cássia	Santa Rita de Cassia
Santa Rita de Cássia	Santa Rita de Cássia
Santa Rita de Cássia	Sta Rita de Cassia
Santa Rita de Cássia	Sta Rita de Cássia
Santa Rita de Cássia	Sta. Rita de Cassia
Santa Rita de Cássia	Sta. Rita de Cássia

Nome dos Municípios	Variações dos nomes dos municípios
Santa Terezinha	Santa Teresinha
Santa Terezinha	Santa Terezinha
Santa Terezinha	Sta Teresinha
Santa Terezinha	Sta Terezinha
Santa Terezinha	Sta. Teresinha
Santa Terezinha	Sta. Terezinha
Santaluz	Santa Luz
Santaluz	Santaluz
Santana	Santana
Santanópolis	Santanopolis
Santanópolis	Santanópolis
Santanópolis	Santanopólis
Santo Amaro	Santo Amaro
Santo Antônio de Jesus	Santo Antonio de Jesus
Santo Antônio de Jesus	Santo Antônio de Jesus
Santo Antônio de Jesus	Santo Antonio Jesus
Santo Antônio de Jesus	Santo Antônio Jesus
Santo Antônio de Jesus	Sto Antonio de Jesus
Santo Antônio de Jesus	Sto Antônio de Jesus
Santo Antônio de Jesus	Sto Antonio Jesus
Santo Antônio de Jesus	Sto Antônio Jesus
Santo Antônio de Jesus	Sto. Antonio de Jesus
Santo Antônio de Jesus	Sto. Antônio de Jesus
Santo Antônio de Jesus	Sto. Antônio de Jesus
Santo Antônio de Jesus	Sto. Antônio Jesus
Santo Estêvão	Santo Estevao
Santo Estêvão	Santo Estevão
Santo Estêvão	Santo Estêvão
São Desidério	Sao Desiderio
São Desidério	São Desidério
São Domingos	Sao Desiderio

Nome dos Municípios	Variações dos nomes dos municípios
São Domingos	Sao Domingos
São Domingos	São Domingos
São Felipe	Sao Felipe
São Felipe	São Felipe
São Félix	Sao Felix
São Félix	São Félix
São Félix	São Felix
São Félix do Coribe	Sao Felix do Coribe
São Félix do Coribe	São Félix do Coribe
São Félix do Coribe	São Felix do Coribe
São Francisco do Conde	Sao Francisco Conde
São Francisco do Conde	São Francisco Conde
São Francisco do Conde	Sao Francisco do Conde
São Francisco do Conde	São Francisco do Conde
São Gabriel	Sao Gabriel
São Gabriel	São Gabriel
São Gonçalo dos Campos	São Goncalo Campos
São Gonçalo dos Campos	São Goncalo Campos
São Gonçalo dos Campos	São Gonçalo Campos
São Gonçalo dos Campos	Sao Goncalo dos Campos
São Gonçalo dos Campos	São Goncalo dos Campos
São Gonçalo dos Campos	São Gonçalo dos Campos
São José da Vitória	Sao Jose da Vitoria
São José da Vitória	São Jose da Vitoria
São José da Vitória	São José da Vitoria
São José da Vitória	São José da Vitória
São José do Jacuípe	Sao Jose do Jacuipe
São José do Jacuípe	São Jose do Jacuipe
São José do Jacuípe	São José do Jacuípe
São Miguel das Matas	Sao Miguel das Matas
São Miguel das Matas	São Miguel das Matas

Nome dos Municípios	Variações dos nomes dos municípios
São Miguel das Matas	São Miguel Matas
São Sebastião do Passé	S. Sebastião Passé
São Sebastião do Passé	Sao Sebastiao do Passe
São Sebastião do Passé	Sao Sebastiao do Passé
São Sebastião do Passé	São Sebastião do Passé
Sapeaçu	Sapeacu
Sapeaçu	Sapeaçu
Sátiro Dias	Satiro Dias
Sátiro Dias	Sátiro Dias
Saubara	Saubara
Saúde	Saude
Saúde	Saúde
Seabra	Seabra
Sebastião Laranjeiras	Sebast. Laranjeiras
Sebastião Laranjeiras	Sebastiao Laranjeiras
Sebastião Laranjeiras	Sebastião Laranjeiras
Senhor do Bonfim	Senhor do Bonfim
Sento Sé	Sento Se
Sento Sé	Sento Sé
Serra do Ramalho	Serra do Ramalho
Serra Dourada	Serra Dourada
Serra Preta	Serra Preta
Serrinha	Serrinha
Serrolândia	Serrolandia
Serrolândia	Serrolândia
Simões Filho	Simoes Filho
Simões Filho	Simões Filho
Sítio do Mato	Sítio do Mato
Sítio do Mato	Sítio do Mato
Sítio do Quinto	Sítio do Quinto
Sítio do Quinto	Sítio do Quinto

Nome dos Municípios	Variações dos nomes dos municípios
Sobradinho	Sobradinho
Souto Soares	Souto Soares
Tabocas do Brejo Velho	Tabocas Brejo Velho
Tabocas do Brejo Velho	Tabocas do Brejo Velho
Tanhaçu	Tanhacu
Tanhaçu	Tanhaçu
Tanque Novo	Tanque Novo
Tanquinho	Tanquinho
Taperoá	Taperoa
Taperoá	Taperoá
Tapiramutá	Tapiramuta
Tapiramutá	Tapiramutá
Teixeira de Freitas	Teixeira de Freitas
Teixeira de Freitas	Teixeira Freitas
Teixeira de Freitas	Teixera de Freitas
Teixeira de Freitas	Teixera Freitas
Teixeira de Freitas	Texeira de Freitas
Teixeira de Freitas	Texeira Freitas
Teodoro Sampaio	Teodoro Sampaio
Teofilândia	Teofilandia
Teofilândia	Teofilândia
Teofilândia	Teoflandia
Teofilândia	Teofilândia
Teolândia	Teolandia
Teolândia	Teolândia
Terra Nova	Terra Nova
Tremedal	Tremedal
Tremedal	Tremendal
Tucano	Tucano
Uauá	Uaua
Uauá	Uauá

Nome dos Municípios	Variações dos nomes dos municípios
Ubaíra	Ubaira
Ubaíra	Ubaíra
Ubaitaba	Ubaitaba
Ubatã	Ubata
Ubatã	Ubatã
Ubatã	Ubatan
Uibaí	Uibai
Uibaí	Uibaí
Umburanas	Umburanas
Una	Una
Urandi	Urandi
Uruçuca	Urucuca
Uruçuca	Uruçuca
Utinga	Utinga
Valença	Valenca
Valença	Valença
Valente	Valente
Várzea da Roça	Varzea da Roca
Várzea da Roça	Varzea da Roça
Várzea da Roça	Várzea da Roça
Várzea da Roça	Várzea da Roca
Várzea da Roça	Varzea Roca
Várzea da Roça	Várzea Roça
Várzea do Poço	Varzea do Poco
Várzea do Poço	Varzea do Poço
Várzea do Poço	Várzea do Poco
Várzea do Poço	Várzea do Poço
Várzea do Poço	Varzea Poco
Várzea do Poço	Varzea Poço
Várzea do Poço	Várzea Poco
Várzea do Poço	Várzea Poço

Nome dos Municípios	Variações dos nomes dos municípios
Várzea Nova	Varzea Nova
Várzea Nova	Várzea Nova
Varzedo	Varzedo
Vera Cruz	Vera Crus
Vera Cruz	Vera Cruz
Vereda	Vereda
Vitória da Conquista	Vitoria Conquista
Vitória da Conquista	Vitória Conquista
Vitória da Conquista	Vitoria da Conquista
Vitória da Conquista	Vitória da Conquista
Wagner	Vagner
Wagner	Vágner
Wagner	Wagner
Wagner	Wágner
Wanderley	Vanderlei
Wanderley	Vanderley
Wanderley	Wanderlei
Wanderley	Wanderley
Wenceslau Guimarães	Wenceslau Guimaraes
Wenceslau Guimarães	Wenceslau Guimarães
Wenceslau Guimarães	Wenceslau Guimaraes
Wenceslau Guimarães	Wenceslau Guimarães
Xique-Xique	Xique Xique
Xique-Xique	Xiquexique
Xique-Xique	Xique-Xique

Nome dos Municípios	Variações dos nomes dos municípios
São Miguel das Matas	São Miguel Matas
São Sebastião do Passé	S. Sebastião Passé
São Sebastião do Passé	Sao Sebastiao do Passe
São Sebastião do Passé	Sao Sebastiao do Passé
São Sebastião do Passé	São Sebastião do Passé
Sapeaçu	Sapeacu
Sapeaçu	Sapeaçu
Sátiro Dias	Satiro Dias
Sátiro Dias	Sátiro Dias
Saubara	Saubara
Saúde	Saude
Saúde	Saúde
Seabra	Seabra
Sebastião Laranjeiras	Sebast. Laranjeiras
Sebastião Laranjeiras	Sebastiao Laranjeiras
Sebastião Laranjeiras	Sebastião Laranjeiras
Senhor do Bonfim	Senhor do Bonfim
Sento Sé	Sento Se
Sento Sé	Sento Sé
Serra do Ramalho	Serra do Ramalho
Serra Dourada	Serra Dourada
Serra Preta	Serra Preta
Serrinha	Serrinha
Serrolândia	Serrolandia
Serrolândia	Serrolândia
Simões Filho	Simoes Filho
Simões Filho	Simões Filho
Sítio do Mato	Sítio do Mato
Sítio do Mato	Sítio do Mato
Sítio do Quinto	Sítio do Quinto
Sítio do Quinto	Sítio do Quinto

Nome dos Municípios	Variações dos nomes dos municípios
Sobradinho	Sobradinho
Souto Soares	Souto Soares
Tabocas do Brejo Velho	Tabocas Brejo Velho
Tabocas do Brejo Velho	Tabocas do Brejo Velho
Tanhaçu	Tanhacu
Tanhaçu	Tanhaçu
Tanque Novo	Tanque Novo
Tanquinho	Tanquinho
Taperoá	Taperoa
Taperoá	Taperoá
Tapiramutá	Tapiramuta
Tapiramutá	Tapiramutá
Teixeira de Freitas	Teixeira de Freitas
Teixeira de Freitas	Teixeira Freitas
Teixeira de Freitas	Teixera de Freitas
Teixeira de Freitas	Teixera Freitas
Teixeira de Freitas	Texeira de Freitas
Teixeira de Freitas	Texeira Freitas
Teodoro Sampaio	Teodoro Sampaio
Teofilândia	Teofilandia
Teofilândia	Teofilândia
Teofilândia	Teoflandia
Teofilândia	Teofilândia
Teolândia	Teolandia
Teolândia	Teolândia
Terra Nova	Terra Nova
Tremedal	Tremedal
Tremedal	Tremendal
Tucano	Tucano
Uauá	Uaua
Uauá	Uauá

Nome dos Municípios	Variações dos nomes dos municípios
Ubaíra	Ubaira
Ubaíra	Ubaíra
Ubaitaba	Ubaitaba
Ubatã	Ubata
Ubatã	Ubatã
Ubatã	Ubatan
Uibaí	Uibai
Uibaí	Uibaí
Umburanas	Umburanas
Una	Una
Urandi	Urandi
Uruçuca	Urucuca
Uruçuca	Uruçuca
Utinga	Utinga
Valença	Valenca
Valença	Valença
Valente	Valente
Várzea da Roça	Varzea da Roca
Várzea da Roça	Varzea da Roça
Várzea da Roça	Várzea da Roça
Várzea da Roça	Várzea da Roca
Várzea da Roça	Varzea Roca
Várzea da Roça	Várzea Roça
Várzea do Poço	Varzea do Poco
Várzea do Poço	Varzea do Poço
Várzea do Poço	Várzea do Poco
Várzea do Poço	Várzea do Poço
Várzea do Poço	Varzea Poco
Várzea do Poço	Varzea Poço
Várzea do Poço	Várzea Poco
Várzea do Poço	Várzea Poço

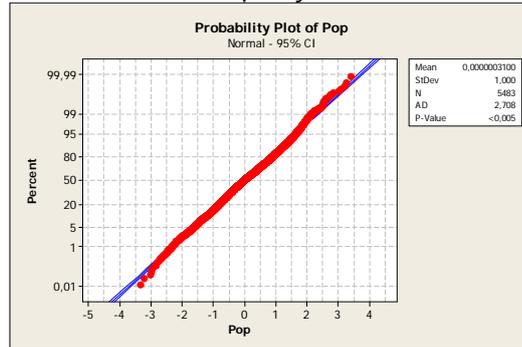
Nome dos Municípios	Variações dos nomes dos municípios
Várzea Nova	Varzea Nova
Várzea Nova	Várzea Nova
Varzedo	Varzedo
Vera Cruz	Vera Crus
Vera Cruz	Vera Cruz
Vereda	Vereda
Vitória da Conquista	Vitoria Conquista
Vitória da Conquista	Vitória Conquista
Vitória da Conquista	Vitoria da Conquista
Vitória da Conquista	Vitória da Conquista
Wagner	Vagner
Wagner	Vágner
Wagner	Wagner
Wagner	Wágner
Wanderley	Vanderlei
Wanderley	Vanderley
Wanderley	Wanderlei
Wanderley	Wanderley
Wenceslau Guimarães	Wenceslau Guimaraes
Wenceslau Guimarães	Wenceslau Guimarães
Wenceslau Guimarães	Wenceslau Guimaraes
Wenceslau Guimarães	Wenceslau Guimarães
Xique-Xique	Xique Xique
Xique-Xique	Xiquexique
Xique-Xique	Xique-Xique

APENDICE F - Padronizações (Transformação de Box&Cox)

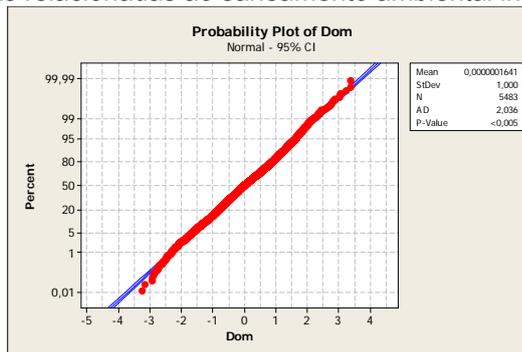
Este apêndice apresenta os testes de normalidade para os dados padronizados com transformação de Box&Cox.

Observa-se que, mesmo após a padronização e a transformação, a normalidade não foi aceita. Isto provavelmente se deve ao fato dos dados originais apresentarem grande assimetria.

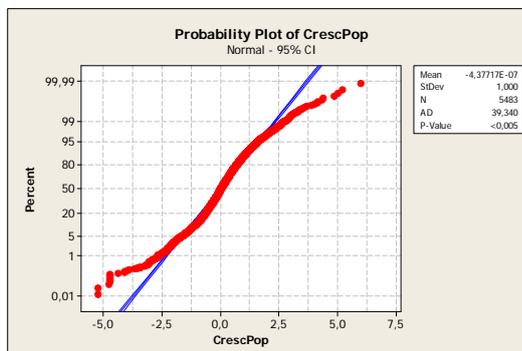
População



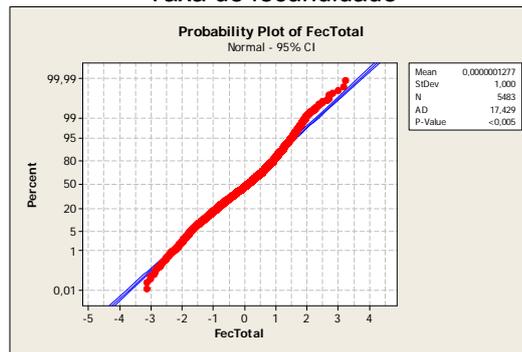
Doenças relacionadas ao saneamento ambiental inadequado



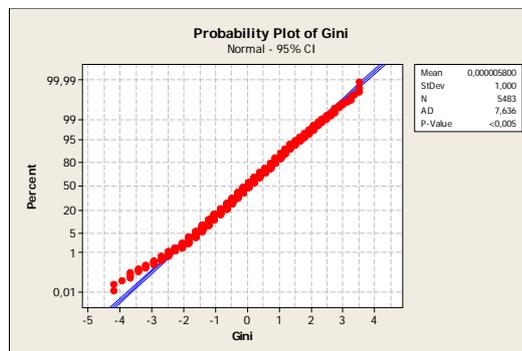
Taxa de crescimento da população



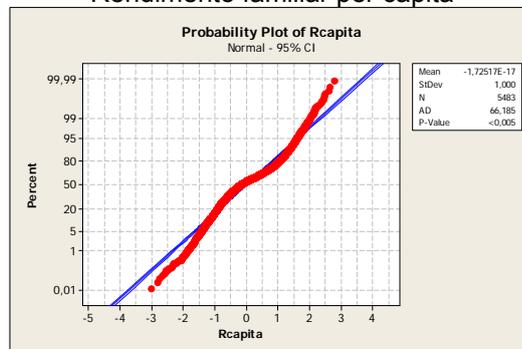
Taxa de fecundidade



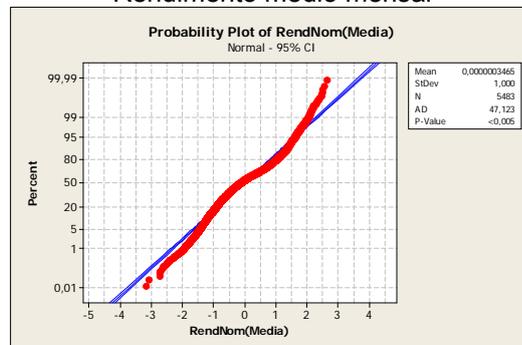
Índice de Gini da distribuição do rendimento



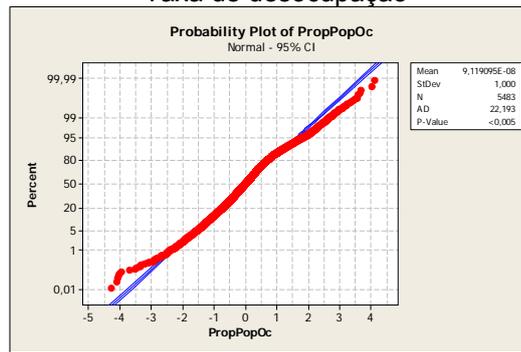
Rendimento familiar per capita



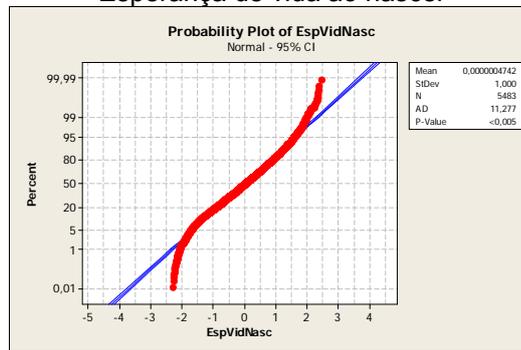
Rendimento médio mensal



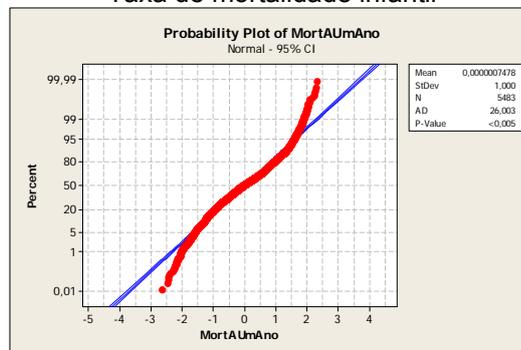
Taxa de desocupação



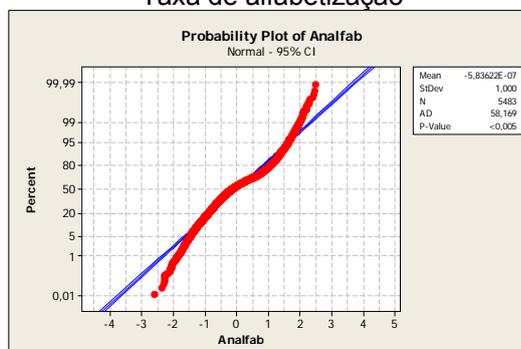
Esperança de vida ao nascer



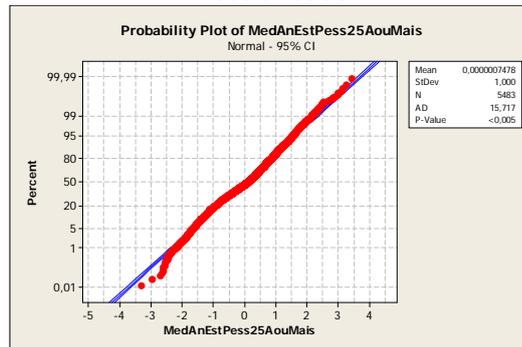
Taxa de mortalidade infantil



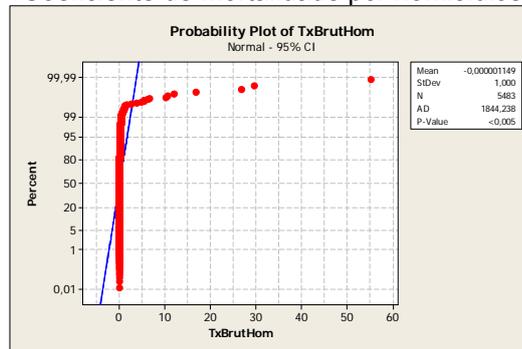
Taxa de alfabetização



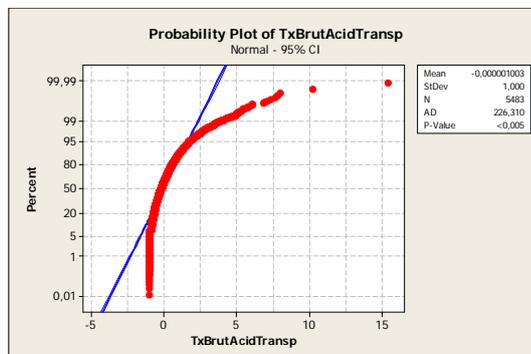
Escolaridade



Coefficiente de mortalidade por homicídios



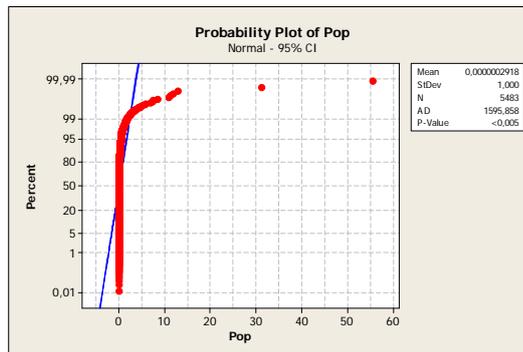
Coefficiente de mortalidade por acidentes de transporte



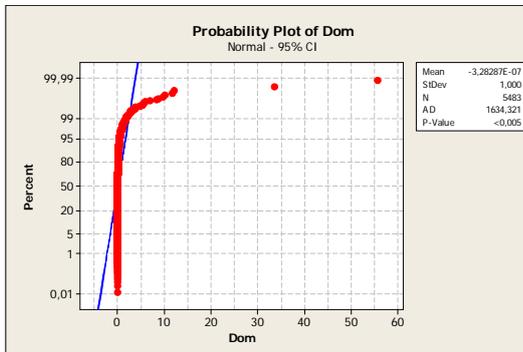
APENDICE G - Padronizações (Dados originais)

Este apêndice apresenta os testes de normalidade para os dados originais padronizados. Observa-se que, mesmo após a padronização, a normalidade não foi aceita. Isto provavelmente se deve ao fato dos dados originais apresentarem grande assimetria.

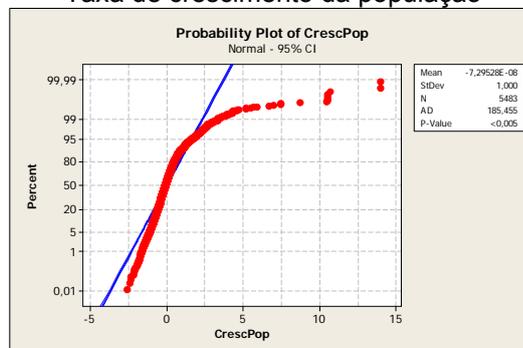
População



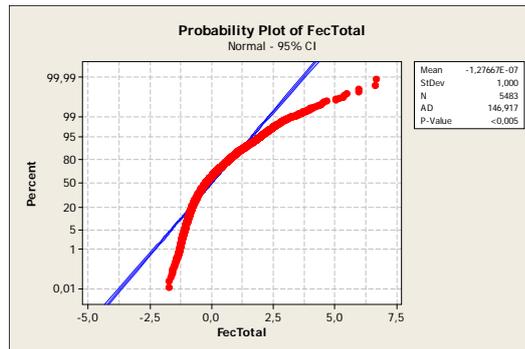
Doenças relacionadas ao saneamento ambiental inadequado



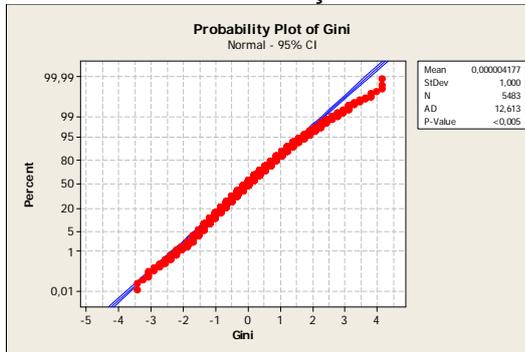
Taxa de crescimento da população



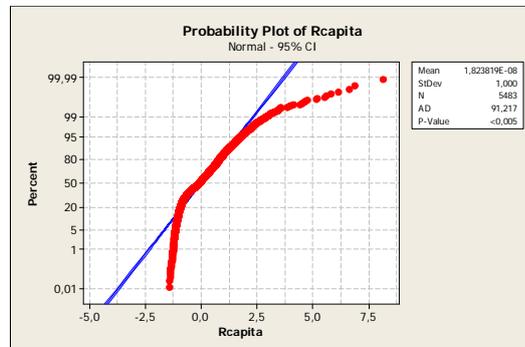
Taxa de fecundidade



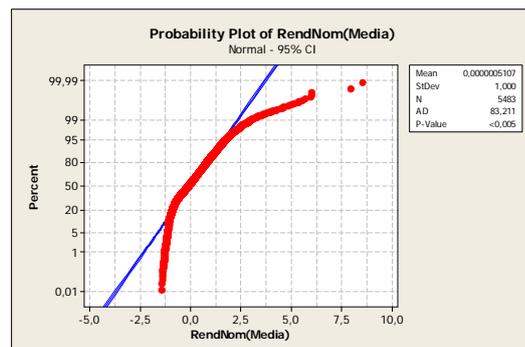
Índice de Gini da distribuição do rendimento



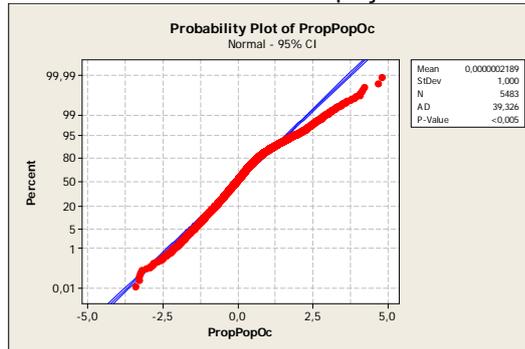
Rendimento familiar per capita



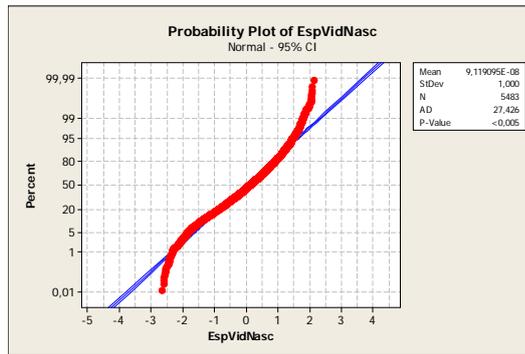
RendNom(Media)



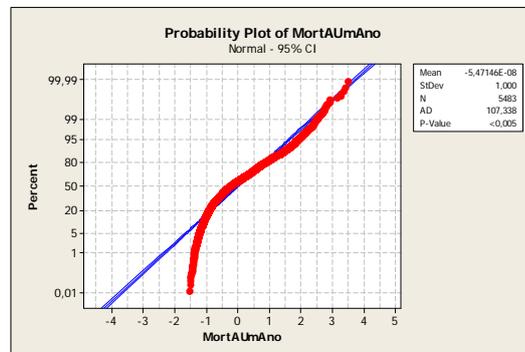
Taxa de desocupação



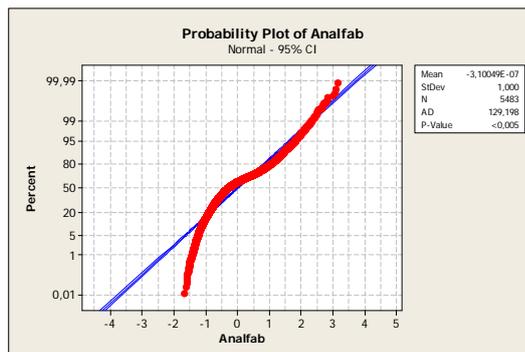
Esperança de vida ao nascer



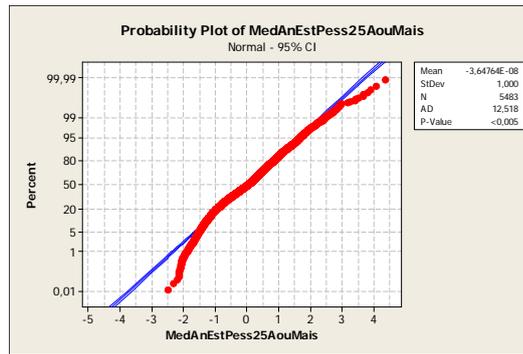
Taxa de mortalidade infantil



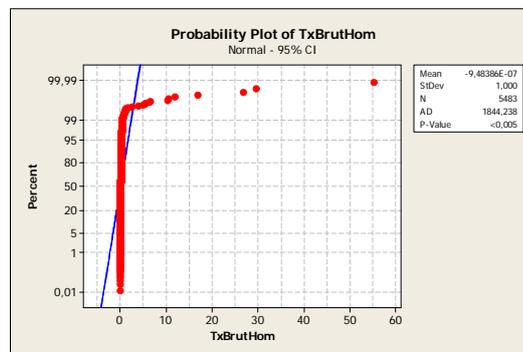
Taxa de alfabetização



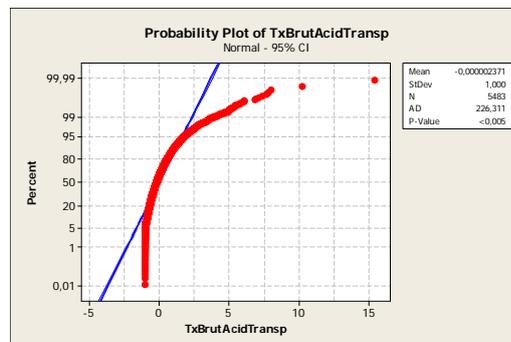
Escolaridade



Coeficiente de mortalidade por homicídios



Coeficiente de mortalidade por acidentes de transporte

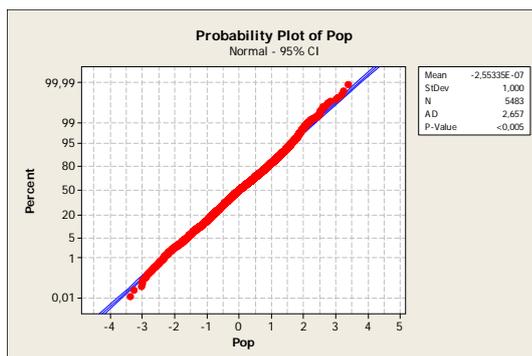


APENDICE H - Padronizações (Transformação x^λ)

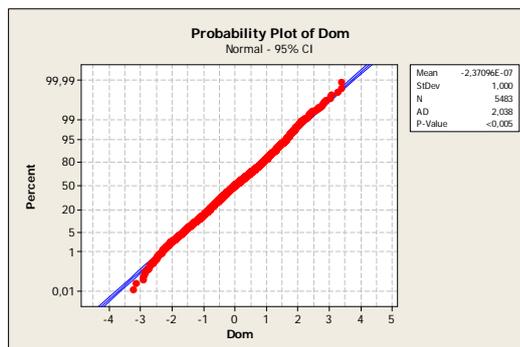
Este apêndice apresenta os testes de normalidade para os dados padronizados com transformação X^λ , em que λ é obtido pela transformação de Box&Cox.

Observa-se que, mesmo após a padronização e a transformação, a normalidade não foi aceita. Isto provavelmente se deve ao fato dos dados originais apresentarem grande assimetria.

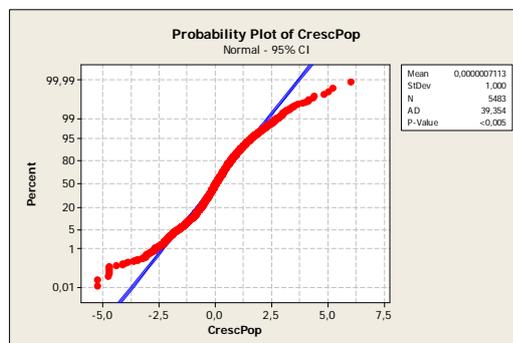
População - B&C=-0,23



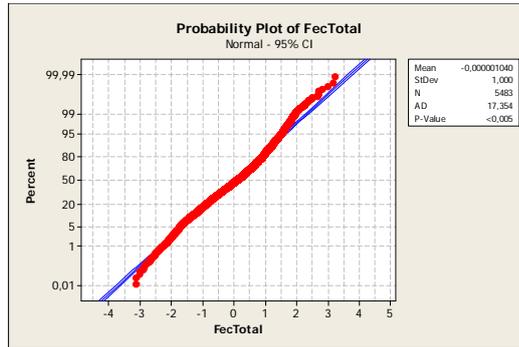
Doenças relacionadas ao saneamento ambiental inadequado - B&C=-0,26



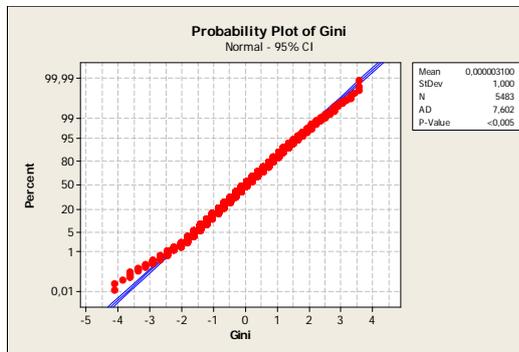
Taxa de crescimento da população - B&C=-0,65



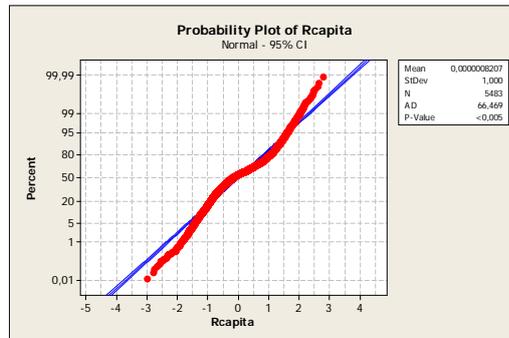
Taxa de fecundidade - B&C=-0,85



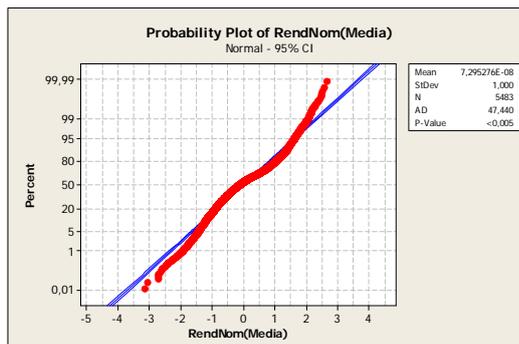
Índice de Gini da distribuição do rendimento - B&C=0,11



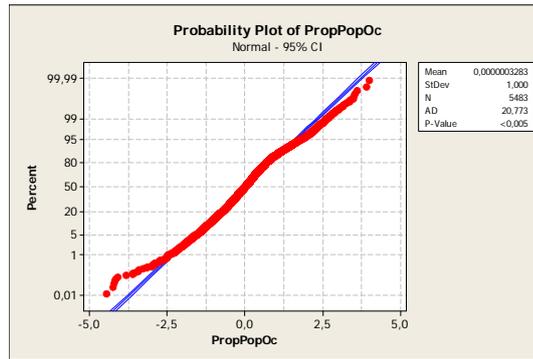
Rendimento familiar per capita - B&C=-0,09



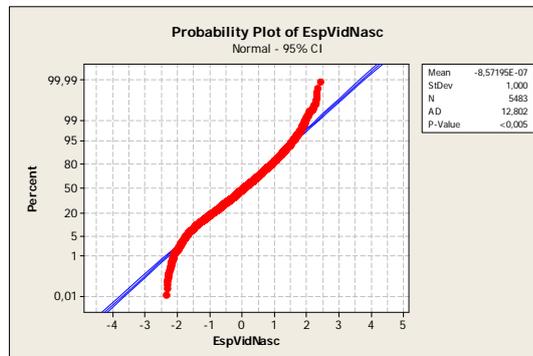
Rendimento médio mensal - B&C=-0,14



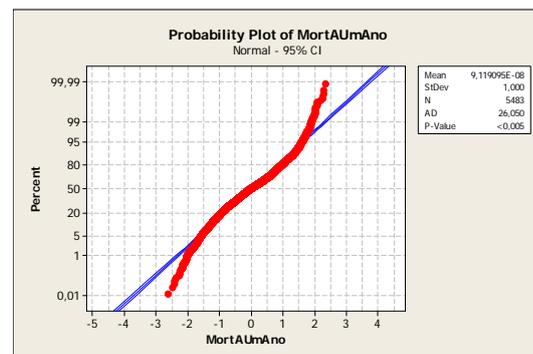
Taxa de desocupação - B&C=0,42



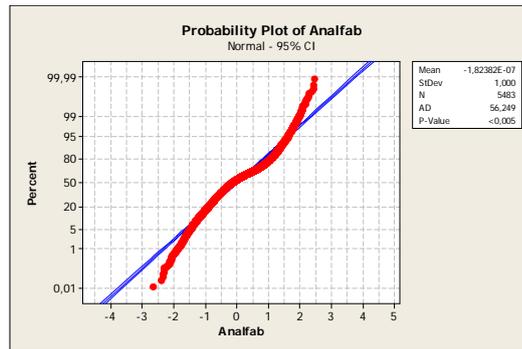
Esperança de vida ao nascer - B&C=2,64



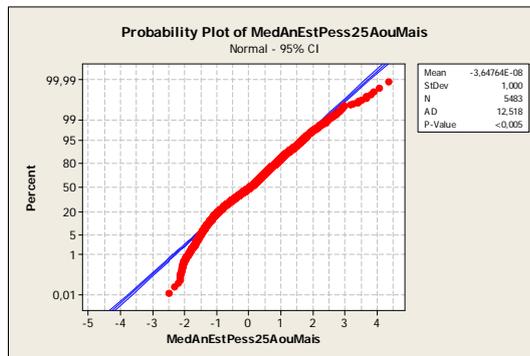
Taxa de mortalidade infantil - B&C=0,17



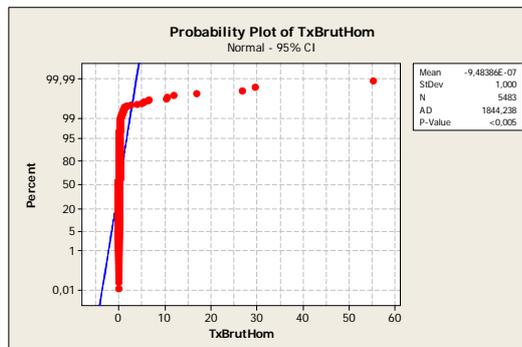
Taxa de alfabetização - B&C=0,48

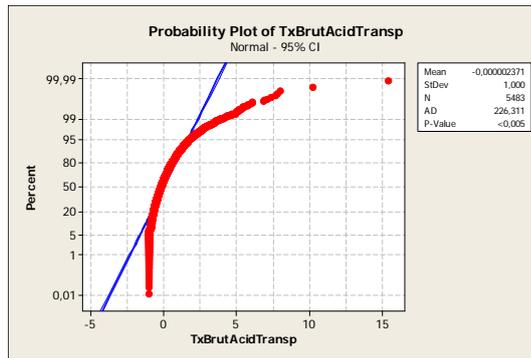


Escolaridade - B&C=0,46



Coefficiente de mortalidade por homicídios - B&C=?



Coeficiente de mortalidade por acidentes de transporte - B&C=?

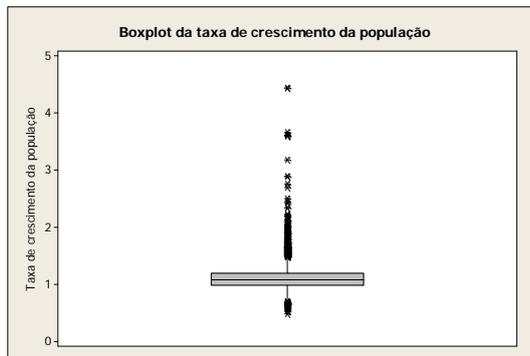


Figura 1: Boxplot da taxa de crescimento da população

A Figura 1 indica que a distribuição da taxa de crescimento da população é assimétrica.

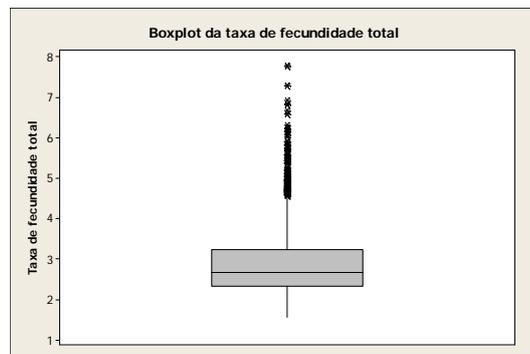


Figura 2: Boxplot da taxa de fecundidade total

Aparentemente a distribuição da taxa de fecundidade total é assimétrica (Figura 2).

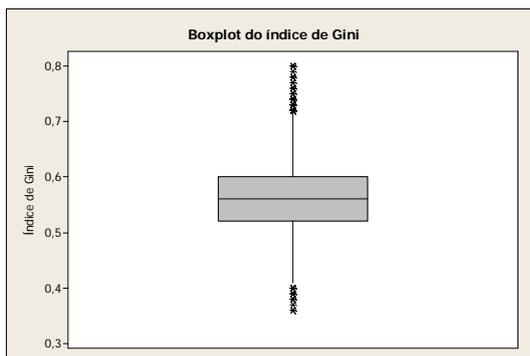


Figura 3: Boxplot do índice de Gini

Figura 3 indica que o índice de Gini apresenta valores discrepantes.

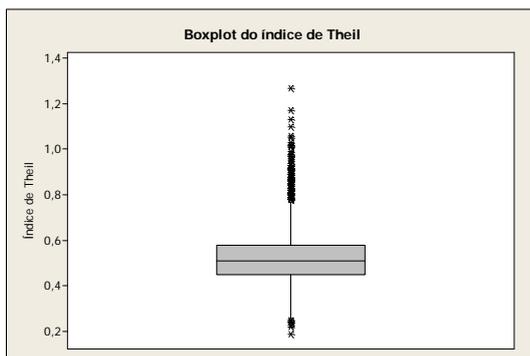


Figura 4: Boxplot do índice de Theil

A Figura 4 indica que a distribuição do índice de Theil é levemente assimétrica.

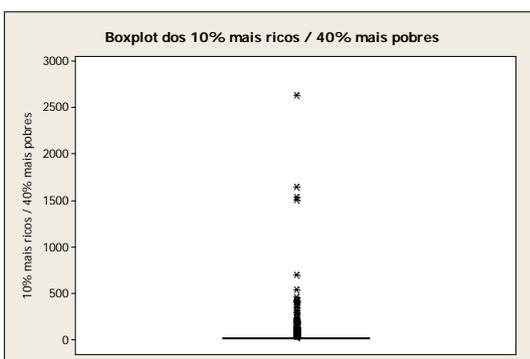


Figura 5: Boxplot dos 10% mais ricos / 40% mais pobres

Observa-se que a distribuição da taxa de 10% mais ricos / 40% mais pobres é assimétrica (Figura 5).

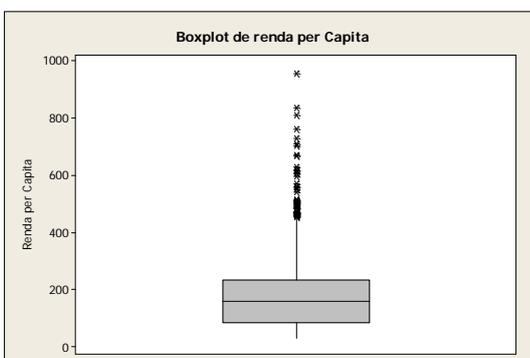


Figura 6: Boxplot da renda per Capita

A Figura 6 indica que a distribuição da renda per Capita é assimétrica.

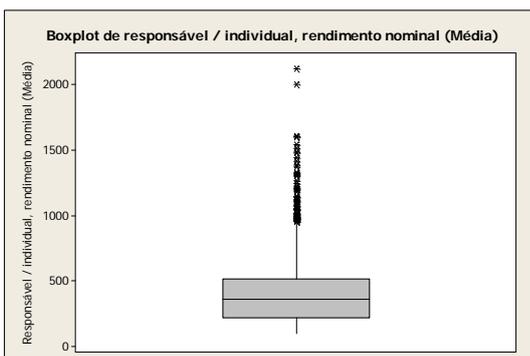


Figura 7: Boxplot de responsável / individual,

rendimento nominal (Média)

Observa-se que a distribuição da taxa média de responsável / individual, rendimento nominal é assimétrica (Figura 7).

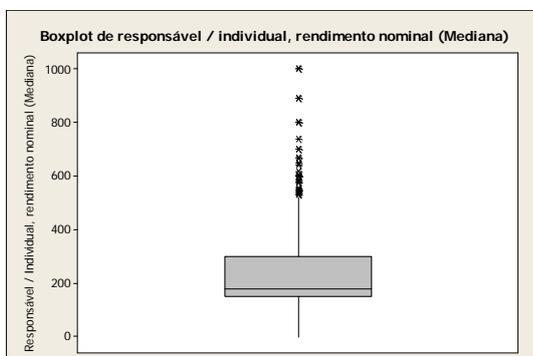


Figura 8: Boxplot de responsável / individual, rendimento nominal (Mediana)

A Figura 8 indica que a distribuição da taxa da mediana de Responsável / Individual, rendimento nominal é assimétrica.

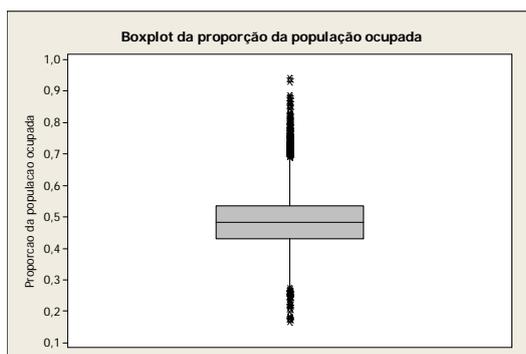


Figura 9: Boxplot da proporção da população ocupada

Aparentemente a proporção da população ocupada apresenta valores discrepantes (Figura 9).

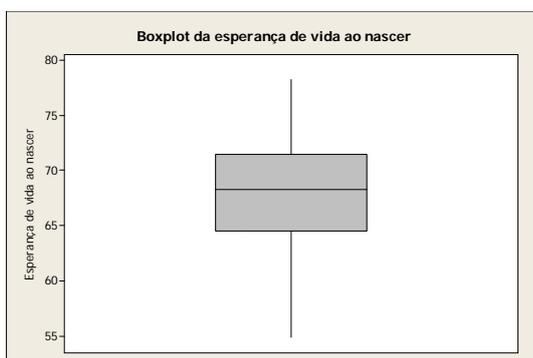


Figura 10: Boxplot da esperança de vida ao nascer

A Figura 10 indica que, aparentemente, a distribuição da esperança de vida ao nascer é simétrica.

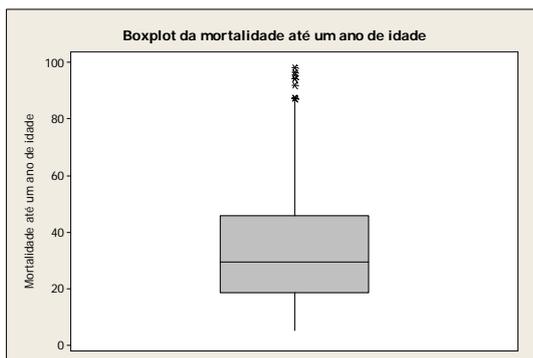


Figura 11: Boxplot da mortalidade até um ano de idade

A Figura 11 indica que a distribuição da mortalidade até um ano de idade é levemente assimétrica.

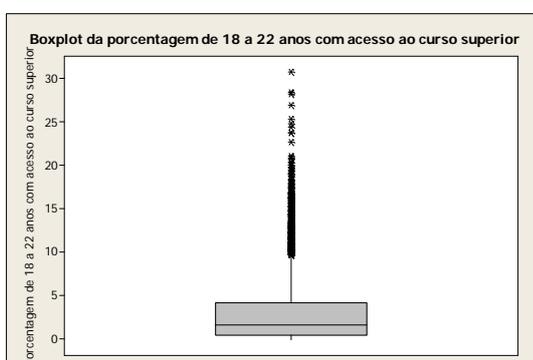


Figura 12: Boxplot da porcentagem de 18 a 22 anos Com acesso ao curso superior

A distribuição da porcentagem de pessoas de 18 a 22 anos com acesso ao curso superior é assimétrica (Figura 12).

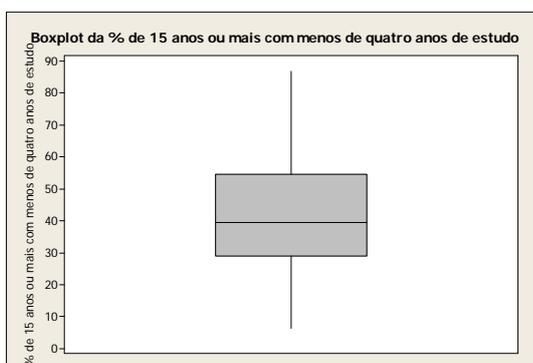


Figura 13: Boxplot da porcentagem 15 anos ou mais com menos de quatro anos de estudo

A Figura 13 indica que a distribuição da porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo é simétrica.

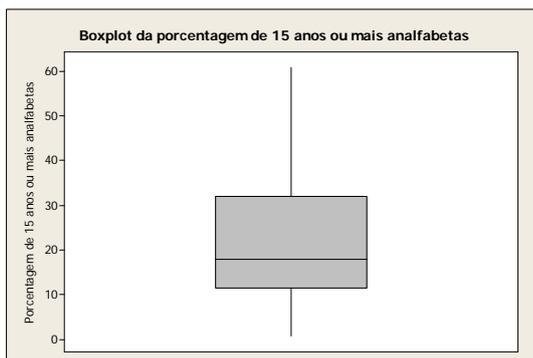


Figura 14: Boxplot da percentagem de 15 anos ou mais analfabetas

A distribuição da percentagem de pessoas de 15 anos ou mais analfabetas é levemente assimétrica (Figura 14).

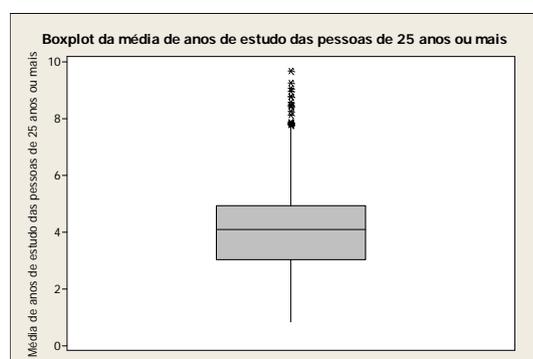


Figura 15: Boxplot da média de anos de estudo das pessoas de 25 anos ou mais

A Figura 15 indica que a média de anos de estudo das pessoas de 25 anos ou mais é levemente assimétrica.

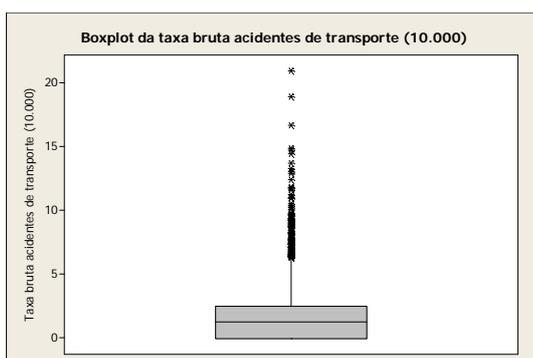


Figura 16: Boxplot da taxa bruta acidentes de transporte (10.000)

A distribuição da taxa bruta de acidentes de transporte (por 10.000 pessoas) é assimétrica (Figura 16).

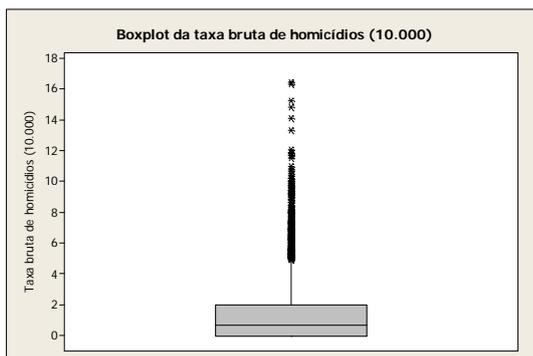


Figura 17: Boxplot da taxa bruta de homicídios (10.000)

A Figura 17 mostra que a distribuição da taxa bruta de homicídios (por 10.000 pessoas) é assimétrica.

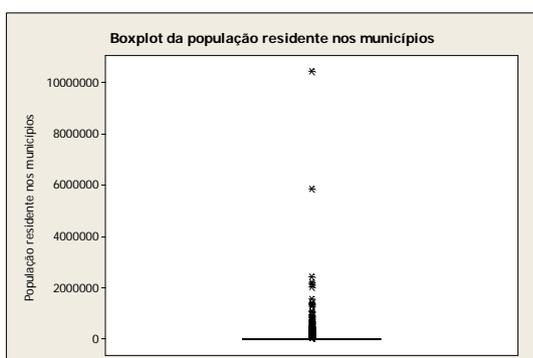


Figura 18: Boxplot da população residente por município

A distribuição da população residente nos municípios é assimétrica (Figura 18).

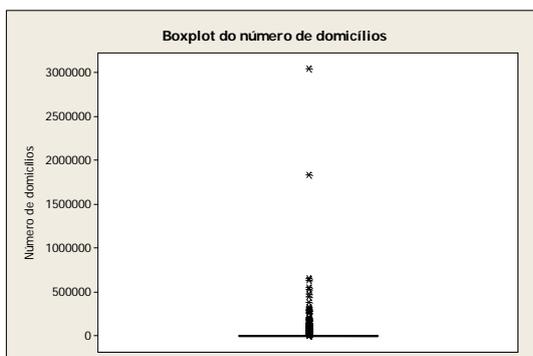


Figura 19: Boxplot do número de domicílios por município

A Figura 19 mostra que a distribuição do número de domicílios por município é assimétrica.

Dinâmica populacional

Taxa de crescimento da população *versus* taxa de fecundidade total

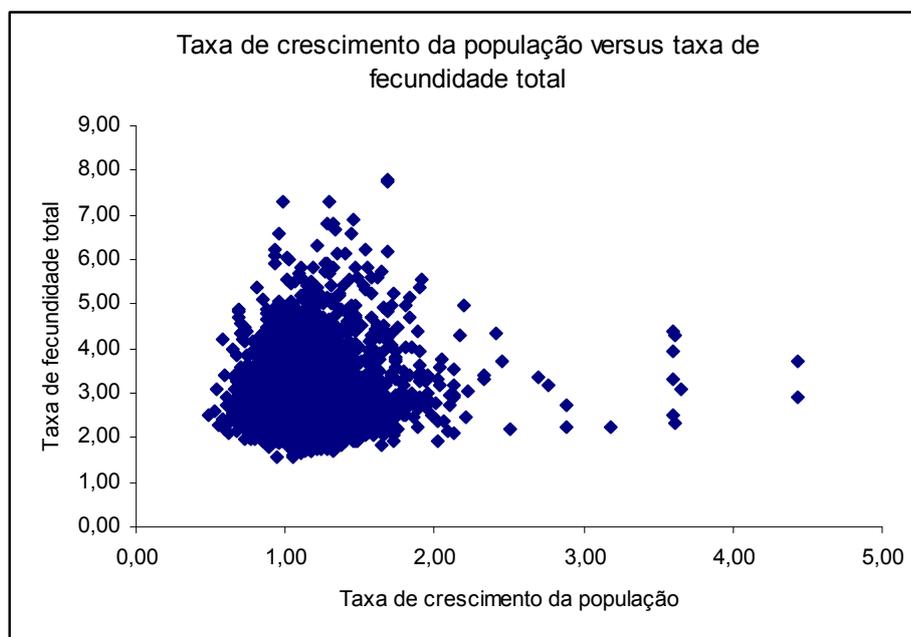


Figura 1: Diagrama de dispersão da taxa de crescimento da população *versus* taxa de fecundidade total

Analisando a Figura 1, observa-se que aparentemente não existe relação entre taxa de crescimento da população e taxa de fecundidade total.

O coeficiente de correlação de Spearman é $r_s = 0,07414459$, comprovando que não existe relação entre a taxa de crescimento da população e a taxa de fecundidade total.

Índice de Gini versus índice de Theil

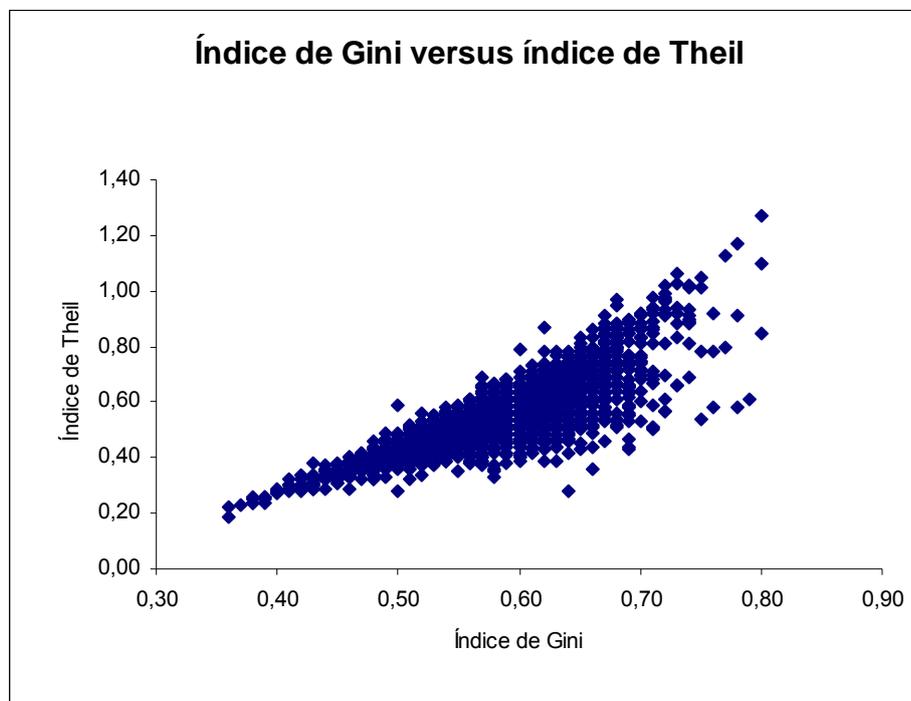


Figura 2: Diagrama de dispersão do índice de Gini versus o índice de Theil

A Figura 2 indica que existe associação entre o Índice de Gini e o Índice de Theil.

Índice de Gini *versus* taxa de 10% mais ricos / 40% mais pobres

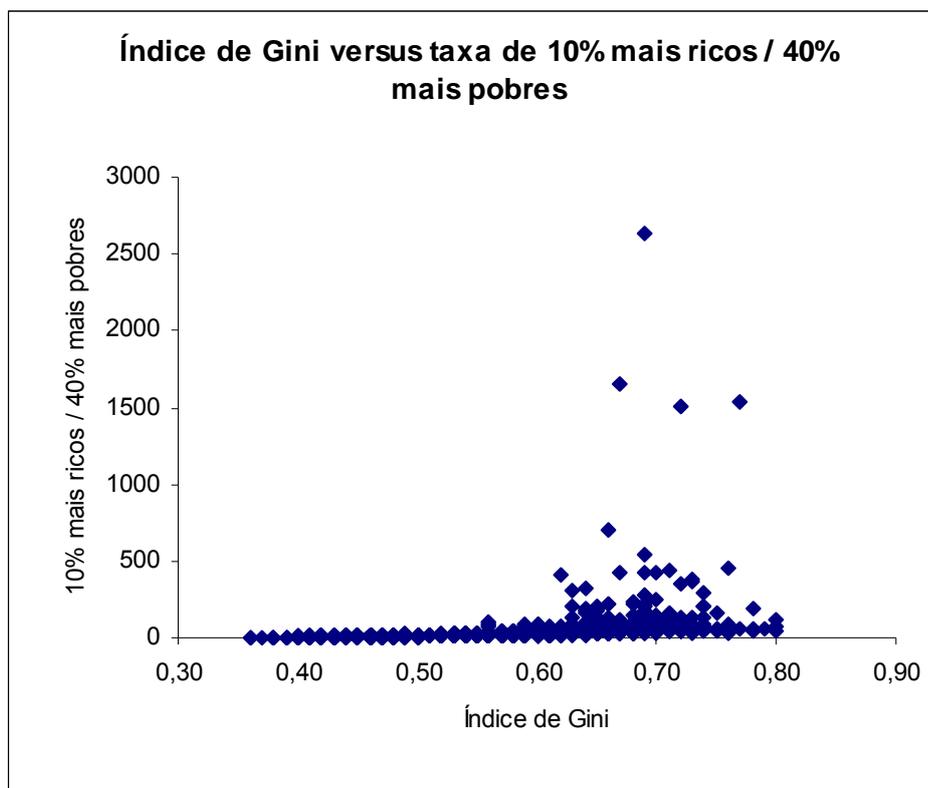


Figura 3: Diagrama de dispersão do índice de Gini *versus* a taxa dos 10% mais ricos / 40% mais pobres

Analisando a Figura 3, observa-se que aparentemente não existe relação entre o índice de Gini e a taxa dos 10% mais ricos / 40% mais pobres

Índice de Gini *versus* a renda per Capita

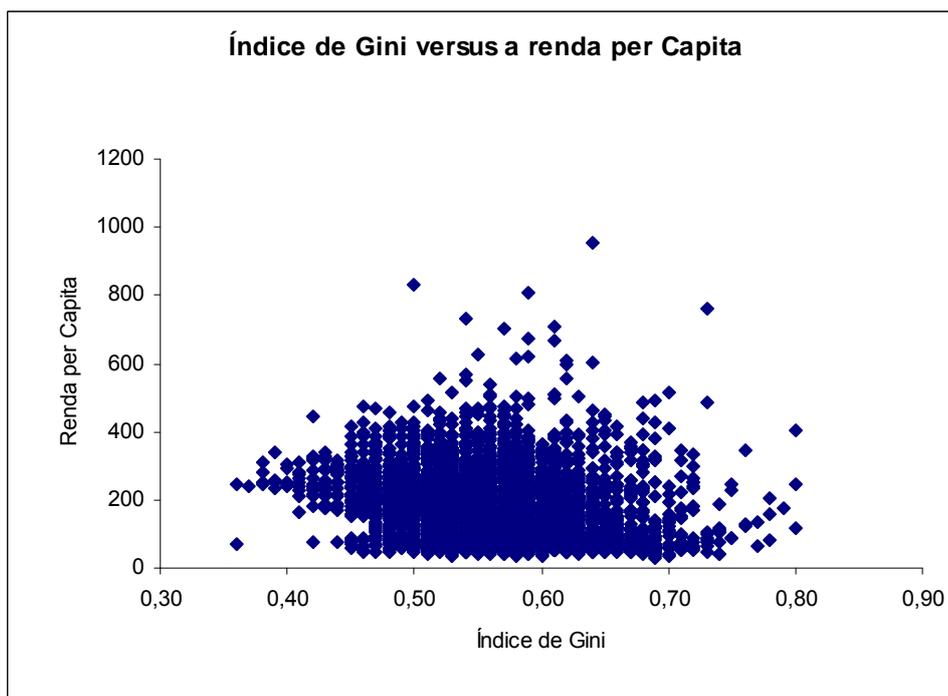


Figura 4: Diagrama de dispersão do índice de Gini *versus* a renda per Capita

Índice de Gini versus a taxa de responsável / individual, rendimento nominal (Média)

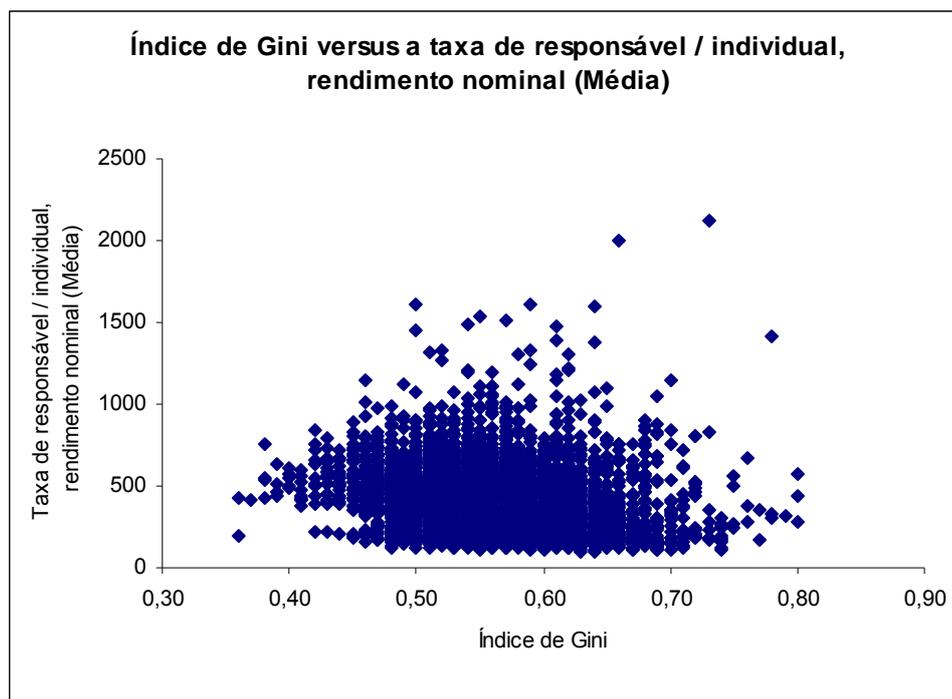


Figura 5: Diagrama de dispersão do índice de Gini versus a taxa de responsável / individual, rendimento nominal (Média)

Índice de Gini versus a taxa de responsável / individual, rendimento nominal (Mediana)

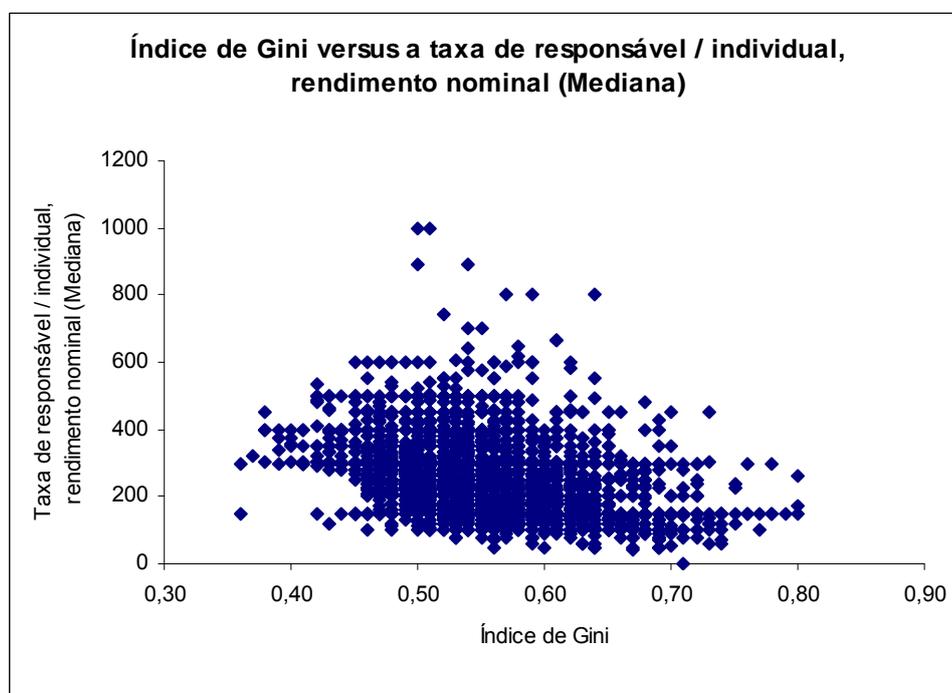
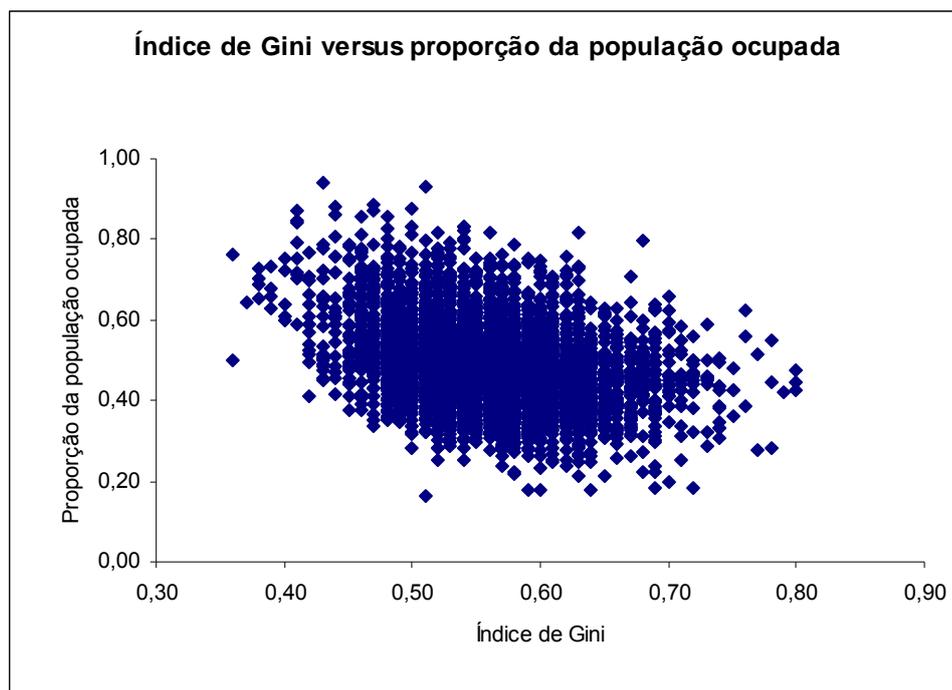
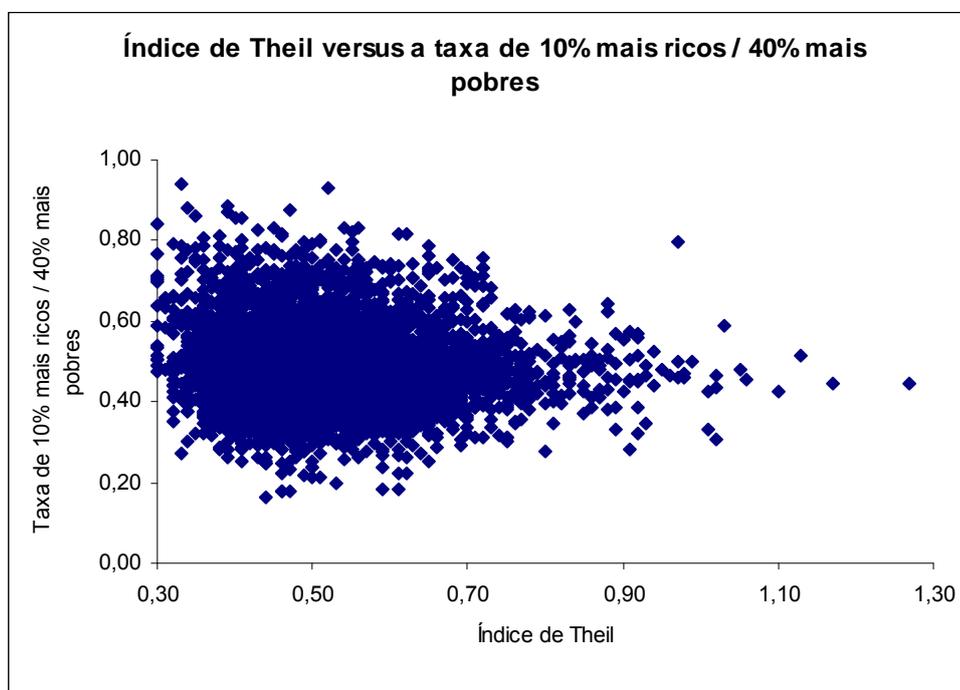


Figura 6: Diagrama de dispersão do índice de Gini versus a taxa de responsável / individual, rendimento nominal (Mediana)

Índice de Gini versus a proporção da população ocupadaFigura 7: Diagrama de dispersão do índice de Gini *versus* a proporção da população ocupada**Índice de Theil versus 10% mais ricos / 40% mais pobres**Figura 8: Diagrama de dispersão do índice de Theil *versus* a taxa de 10% mais ricos / 40% mais pobres

Esperança de vida ao nascer *versus* mortalidade até um ano de idade

Esperança de vida ao nascer

Tabela 10: Esperança de vida ao nascer

Média	67,7760
Desvio Padrão	4,8460
Coeficiente de Variação	7,1500
Mínimo	54,9000
1º Quartil	64,5600
Mediana	68,2800
3º Quartil	71,4500
Máximo	78,1800
Assimetria.	-0,37
Curtose	-0,53

A média da esperança de vida ao nascer é 67,78, em que 50% da esperança de vida ao nascer está entre 64,56 e 71,45 (Tabela 10).

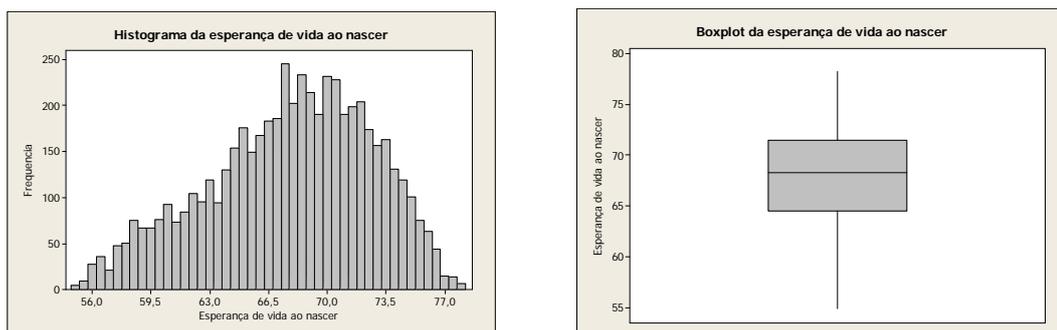


Figura 10: Histograma (a) e boxplot (b) da esperança de vida ao nascer

A Figura 10 mostra que a esperança de vida ao nascer é levemente assimétrica à esquerda.

Mortalidade até um ano de idade

Tabela 11: Mortalidade até um ano de idade

Média	33,9650
Desvio Padrão	18,3820
Coeficiente de Variação	54,1200
Mínimo	5,3800
1º Quartil	18,6200
Mediana	29,3800
3º Quartil	45,9200
Máximo	98,1200
Assimetria.	0,72
Curtose	-0,34

A média da mortalidade até um ano de idade é 33,965, com um coeficiente de variação de 54,12%, variando de 5,38 à 98,12 (Tabela 11).

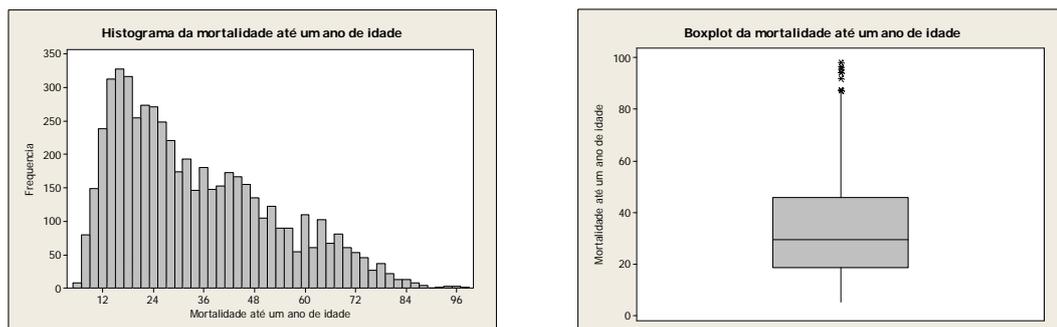


Figura 11: Histograma (a) e boxplot (b) da mortalidade até um ano de idade

Aparentemente a mortalidade até um ano de idade é assimétrica à direita e apresenta valores discrepantes (Figura 11).

Porcentagem 18 a 22 anos com acesso ao curso superior versus porcentagem 15 anos ou mais com menos de quatro anos de estudo

O coeficiente de correlação de Spearman é $r_s = -0,737$, comprovando que existe relação negativa entre a porcentagem de 18 a 22 anos com acesso ao curso superior e a porcentagem de 15 anos ou mais com menos de quatro anos de estudo.

Porcentagem 18 a 22 anos com acesso ao curso superior versus porcentagem 15 anos ou mais analfabetas

O coeficiente de correlação de Spearman é $r_s = -0,695$, comprovando que existe relação negativa entre a porcentagem de 18 a 22 anos com acesso ao curso superior e a porcentagem de 15 anos ou mais analfabetas.

Porcentagem 15 anos ou mais com menos de quatro anos de estudo versus porcentagem de 15 anos ou mais analfabetas

O coeficiente de correlação de Spearman é $r_s = 0,967$, comprovando que existe forte relação entre a porcentagem 15 anos ou mais com menos de quatro anos de estudo e a porcentagem de 15 anos ou mais analfabetas.

Socialização (Educação)

Porcentagem 18 a 22 anos com acesso ao curso superior

Tabela 12: Porcentagem de pessoas de 18 a 22 anos com acesso ao curso superior

Média	3,0130
Desvio Padrão	3,6570
Coefficiente de Variação	121,3700
Mínimo	0,0000
1º Quartil	0,5100
Mediana	1,6400
3º Quartil	4,2000
Máximo	30,7400
Assimetria.	2,11
Curtose	5,76

A Tabela 12 mostra que a média da porcentagem pessoas de 18 a 22 anos com acesso ao curso superior é 3,01, com um coeficiente de variação de 121,37%. Observa-se ainda que esta porcentagem varia de 0,00 à 30,74.

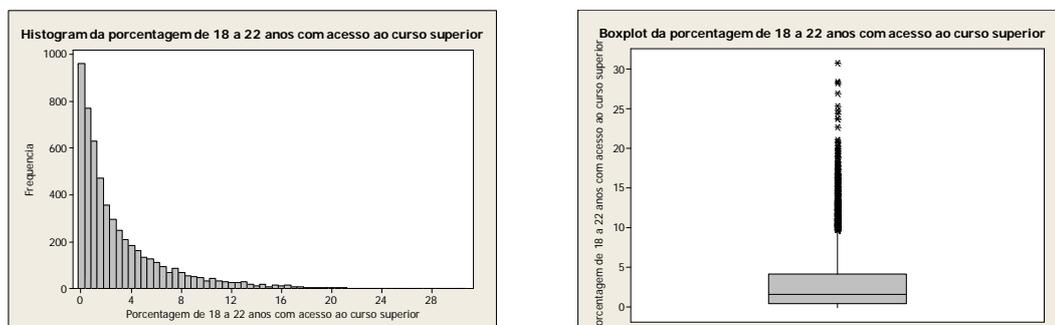


Figura 12: Histograma (a) e boxplot (b) da porcentagem de 18 a 22 anos com acesso ao curso superior

A Figura 12 mostra que a porcentagem pessoas de 18 a 22 anos com acesso ao curso superior é assimétrica à direita e apresenta valores discrepantes.

Porcentagem 15 anos ou mais com menos de quatro anos de estudo

Tabela 13: Porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo

Média	41,3290
Desvio Padrão	15,5020
Coefficiente de Variação	37,5100
Mínimo	6,2400
1º Quartil	29,0200
Mediana	39,2100
3º Quartil	54,4100
Máximo	86,5800
Assimetria.	0,23
Curtose	-0,87

A Tabela 13 mostra que a média da porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo é 41,33, com um coeficiente de variação de 37,51%, variando de 6,24 à 86,58.

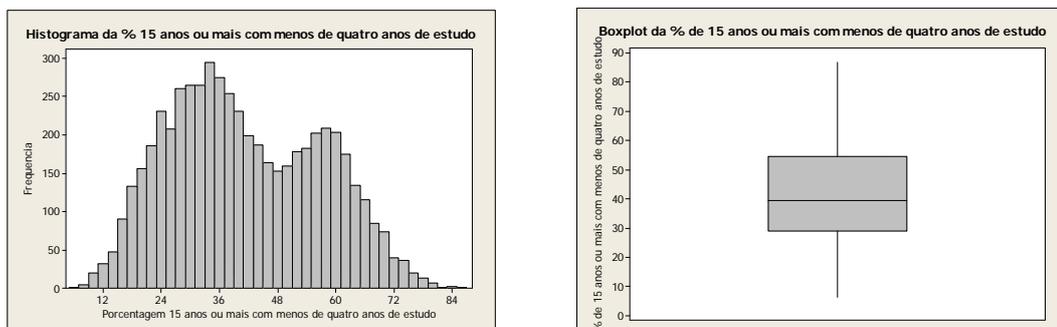


Figura 13: Histograma (a) e boxplot (b) da porcentagem de 15 anos ou mais com menos de quatro anos de estudo

A Figura 13 indica que a porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo tem uma grande concentração ao redor de 35% e de 60% e aparentemente é simétrica.

Porcentagem 15 anos ou mais analfabetas

Tabela 14: Porcentagem de pessoas de 15 anos ou mais analfabetas

Média	21,7000
Desvio Padrão	12,4230
Coefficiente de Variação	57,2500
Mínimo	0,9100
1º Quartil	11,6300
Mediana	17,9100
3º Quartil	31,9900
Máximo	60,6600
Assimetria.	0,58
Curtose	-0,75

A média da porcentagem de pessoas de 15 anos ou mais analfabetas é 21,70, variando de 0,91 à 60,66 (Tabela 14).

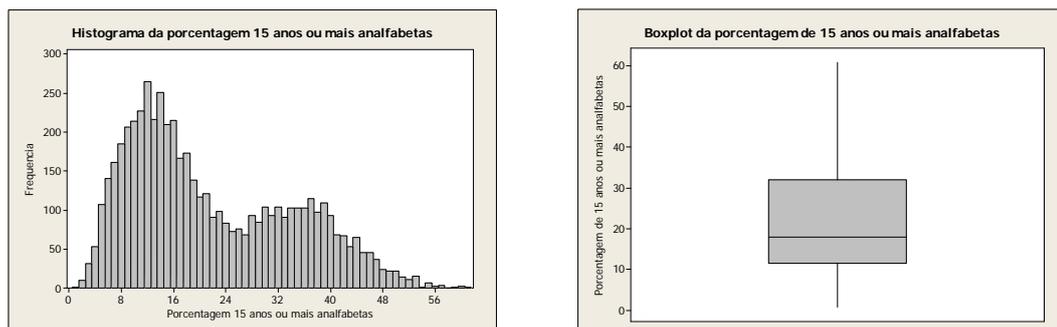


Figura 14: Histograma (a) e boxplot (b) da porcentagem de 15 anos ou mais analfabetas

Aparentemente a porcentagem de pessoas de 15 anos ou mais analfabetas é levemente assimétrica à direita (Figura 14).

Média de anos de estudo das pessoas de 25 anos ou mais

Tabela 15: Média de anos de estudo das pessoas de 25 anos ou mais

Média	4,0471
Desvio Padrão	1,2849
Coefficiente de Variação	31,7500
Mínimo	0,8100
1º Quartil	3,0200
Mediana	4,0800
3º Quartil	4,9200
Máximo	9,6500
Assimetria.	0,27
Curtose	-0,23

A Tabela 15 mostra que a média das médias de anos de estudo das pessoas de 25 anos ou mais é 4,08, com um coeficiente de variação de 31,75%, variando de 3,02 à 9,65.

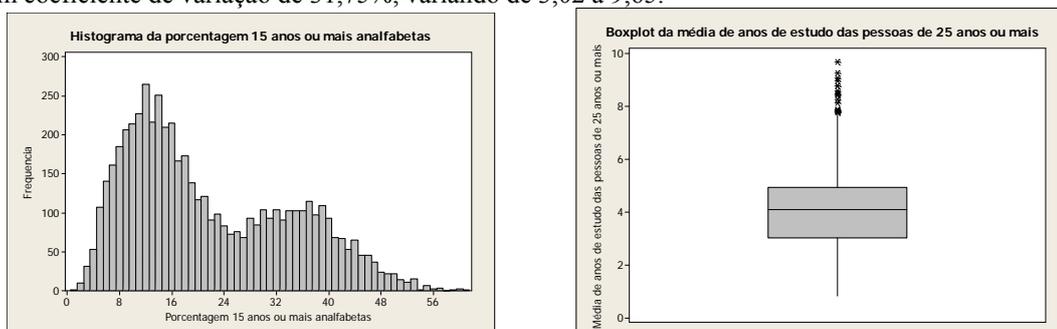


Figura 14: Histograma (a) e boxplot (b) da porcentagem de 15 anos ou mais analfabetas

Aparentemente a porcentagem de pessoas de 15 anos ou mais analfabetas é assimétrica à direita (Figura 14).

Taxa bruta de acidentes de transporte (10.000) versus taxa bruta de homicídios (10.000)

O coeficiente de correlação de Spearman é $r_s=0,236$, mostrando que existe fraca relação entre a taxa bruta de acidentes de transporte (10.000) e a taxa bruta de homicídios (10.000).

Taxa Bruta Acidentes de transporte (10.000)

Tabela 16: Taxa bruta acidentes de transporte (10.000)

Média	1,6680
Desvio Padrão	1,8957
Coefficiente de Variação	113,6500
Mínimo	0,0000
1º Quartil	0,0000
Mediana	1,2712
3º Quartil	2,5323
Máximo	20,8768
Assimetria.	2,05
Curtose	8,24

Observa-se que a média da taxa bruta acidentes de transporte (por 10.000 pessoas) é 2,67, com um coeficiente de variação de 113,65%. Observa-se ainda que a taxa bruta acidentes de transporte varia de 0,00 à 20,88, com um coeficiente de assimetria de 2,05 (Tabela 16).

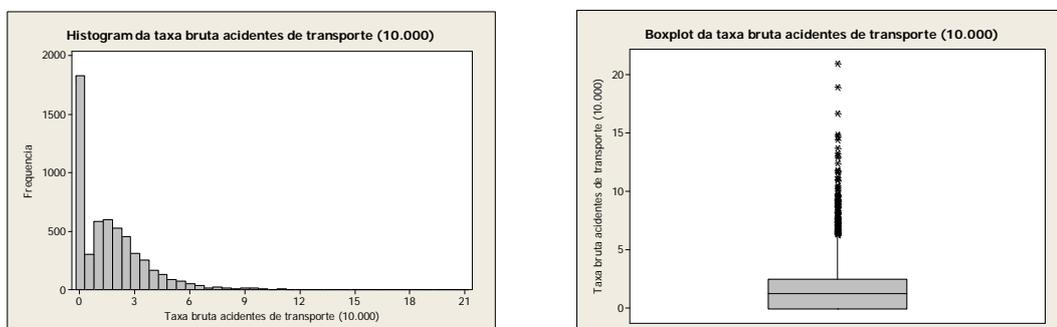


Figura 16: Histograma (a) e boxplot (b) da taxa bruta de acidentes de transporte (10.000)

Observa-se que a taxa bruta de acidentes de transporte (por 10.000 pessoas) é assimétrica à direita apresentando pontos discrepantes (Figura 16).

Taxa Bruta de Homicídios (10.000)

Tabela 17: Taxa bruta de homicídios (10.000)

Média	1,3220
Desvio Padrão	1,8430
Coefficiente de Variação	139,4200
Mínimo	0,0000
1° Quartil	0,0000
Mediana	0,6793
3° Quartil	1,9806
Máximo	16,3934
Assimetria.	2,27
Curtose	7,55

A Tabela 17 mostra que a taxa bruta de homicídios (por 10.000 habitantes) varia de 0 a 16,39, em que a taxa média é 1,32 com um coeficiente de variação de 139,42. Observa-se ainda que o coeficiente de assimetria é 2,27 e 50% dos municípios apresentam a taxa bruta de homicídios menor ou igual a 0,68.

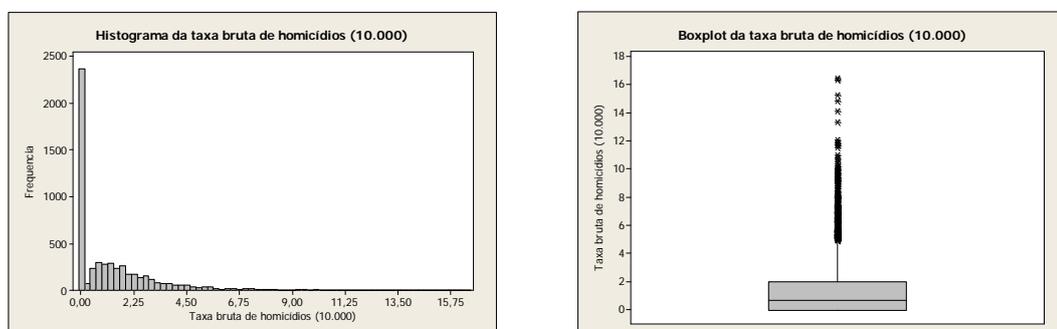


Figura 17: Histograma (a) e boxplot (b) da taxa bruta de homicídios (10.000)

A Figura 17 indica que a taxa bruta de homicídios (por 10.000 pessoas) é assimétrica à direita, apresentando uma grande concentração de taxas iguais a zero, apresentando pontos discrepantes.

APENDICE O - População residente por município *versus* número de domicílios por município

O coeficiente de correlação de Spearman é $r_s=0,991$, mostrando que existe forte relação entre a população residente nos municípios e o número de domicílios por municípios.

População residente por município

Tabela 18: População residente por município

Média	30923,0000
Desvio Padrão	187153,0000
Coeficiente de Variação	605,2300
Mínimo	795,0000
1º Quartil	5154,0000
Mediana	10425,0000
3º Quartil	21451,0000
Máximo	10434252,0000
Assimetria.	38,70
Curtose	1930,11

A Tabela 18 mostra que em média existem 30923 pessoas por município, com um desvio padrão de 187153 e um coeficiente de variação de 605,23%. Observa-se ainda que a população residente por município varia de 795 à 10434252.

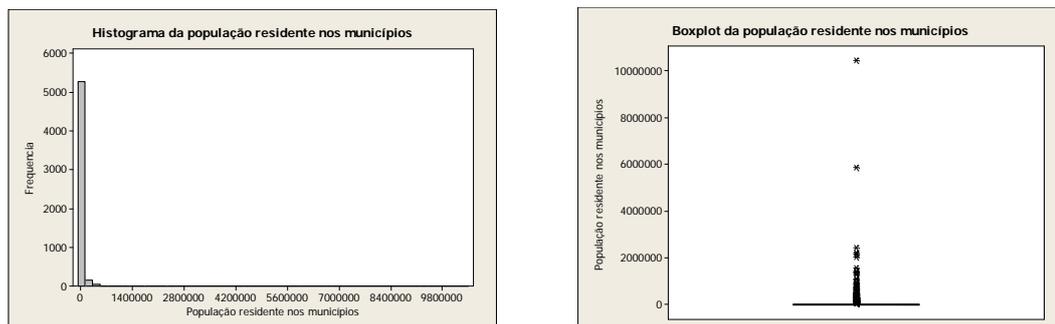


Figura 18: Histograma (a) e boxplot (b) da população residente nos municípios

Observa-se que a concentração da população residente nos municípios é assimétrica à direita, apresentando valores discrepantes (Figura 18).

Número de domicílios por município

Tabela 19: Número de domicílios por município

Média	8291,0000
Desvio Padrão	54503,0000
Coeficiente de Variação	657,4000
Mínimo	227,0000
1º Quartil	1329,0000
Mediana	2647,0000
3º Quartil	5457,0000
Máximo	3039104,0000
Assimetria.	39,82
Curtose	1989,10

Em média existem 8291 domicílios por município, em que o número de domicílios varia de 227 à 3039104, com um desvio padrão de 54.503. Observa-se ainda que o coeficiente de assimetria é 39,82 (Tabela 19).

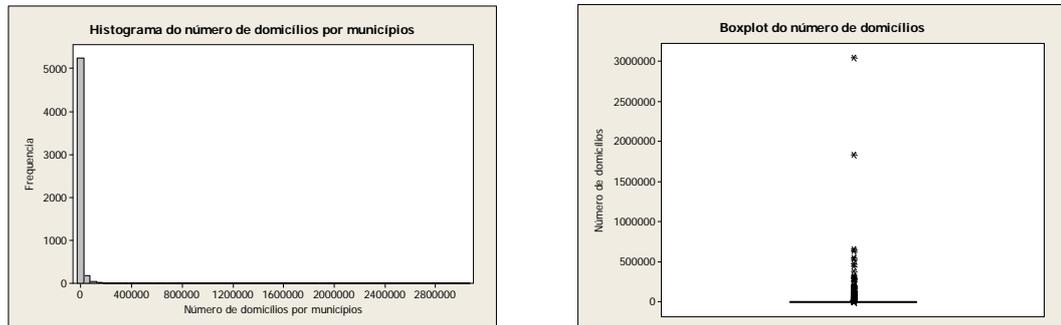


Figura 19: Histograma (a) e boxplot (b) do número de Domicílios por municípios

A Figura 19 indica que o número de domicílios por município é assimétrica à direita, apresentando valores discrepantes.

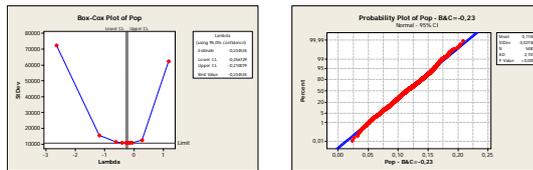
APENDICE P - Transformações de Box&Cox

Este arquivo mostra os testes de normalidade para os dados com transformação de Box&Cox.

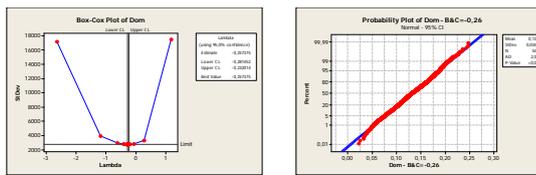
Observa-se que, mesmo após a transformação, a normalidade não foi aceita. Isto provavelmente se deve ao fato dos dados originais apresentarem grande assimetria.

Obs: Note que logo após o nome das variáveis está “B&C=”. O valor que aparece logo após o “=” é o “ λ ” encontrado na transformação de Box&Cox. Existem três casos em que aparece “?” após o “=”. Isto se deve ao fato de não existir um “ λ ” para transformar os dados.

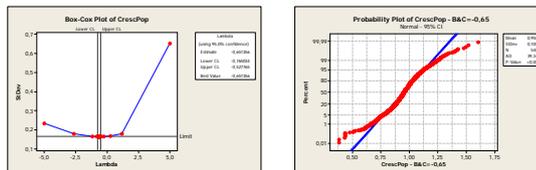
População



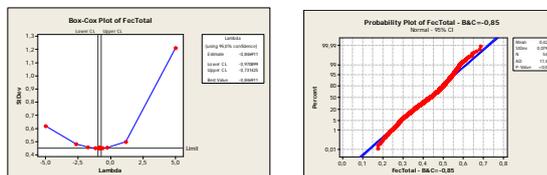
Doenças relacionadas ao saneamento ambiental inadequado



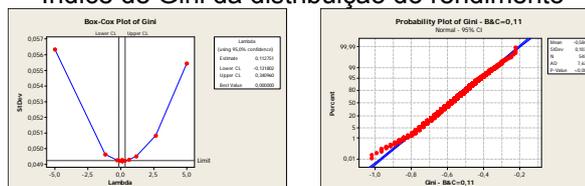
Taxa de crescimento da população



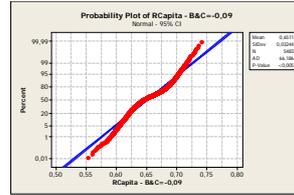
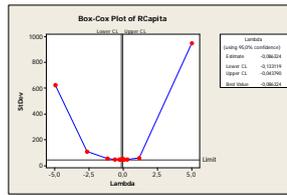
Taxa de fecundidade



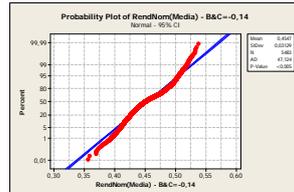
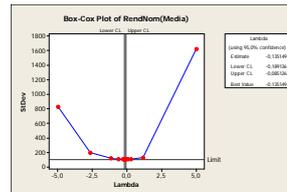
Índice de Gini da distribuição do rendimento



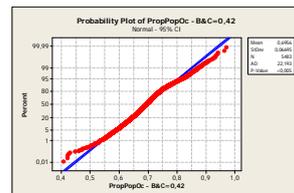
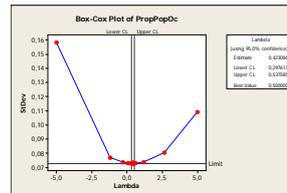
Rendimento familiar per capita



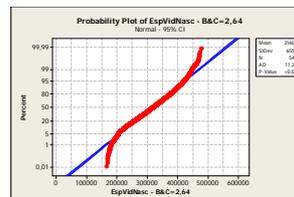
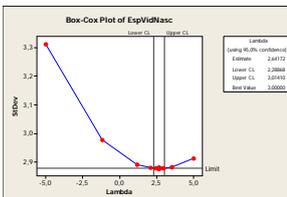
Rendimento médio mensal



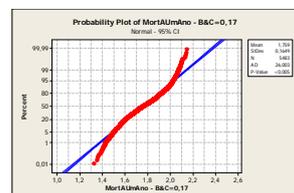
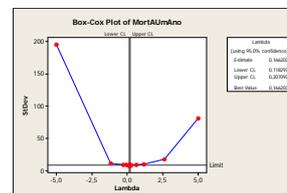
Taxa de desocupação



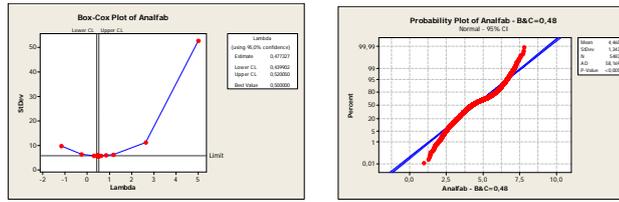
Esperança de vida ao nascer



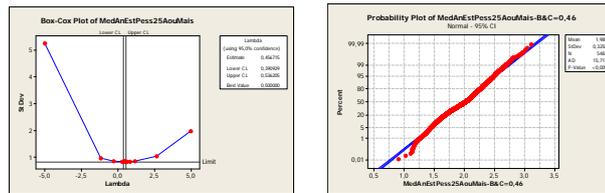
Taxa de mortalidade infantil



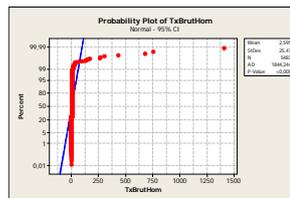
Taxa de alfabetização



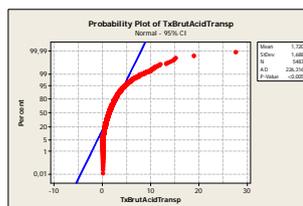
Escolaridade



Coefficiente de mortalidade por homicídios



Coefficiente de mortalidade por acidentes de transporte



Dinâmica Populacional

Taxa de crescimento da população

Tabela 1: Taxa de crescimento da população

Média	1,1120
Desvio Padrão	0,2377
Coeficiente de Variação	21,3800
Mínimo	0,4840
1º Quartil	0,9889
Mediana	1,0814
3º Quartil	1,1876
Máximo	4,4396
Assimetria	3,75
Curtose	33,78

A Tabela 1 mostra que a média da taxa de crescimento da população é 1,1120, variando de 0,4840 à 4,4396. Observa-se ainda que os dados são assimétricos à direita e que o coeficiente de variação é 21,38%.

Taxa de fecundidade total

Tabela 2: Taxa de fecundidade total

Média	2,8598
Desvio Padrão	0,7394
Coeficiente de Variação	25,8600
Mínimo	1,5600
1º Quartil	2,3200
Mediana	2,6700
3º Quartil	3,2200
Máximo	7,7900
Assimetria.	1,47
Curtose	3,20

A média da taxa de fecundidade total é 2,8598, com um coeficiente de variação de 25,86% e 50% da taxa de fecundidade está entre 2,32 e 3,22 (Tabela 2).

Sistema de Produção (Trabalho e rendimento)

Índice de Gini

Tabela 3: Índice de Gini

Média	0,5601
Desvio Padrão	0,0579
Coeficiente de Variação	10,3300
Mínimo	0,3600
1º Quartil	0,5200
Mediana	0,5600
3º Quartil	0,6000
Máximo	0,8000
Assimetria.	0,30
Curtose	0,44

Analisando a Tabela 3, observa-se que a média do índice de Gini é 0,5601, em que 50% dos índices está entre 0,56 e 0,80.

Índice de Theil

Tabela 4: Índice de Theil

Média	0,5214
Desvio Padrão	0,1091
Coefficiente de Variação	20,9200
Mínimo	0,1900
1º Quartil	0,4500
Mediana	0,5100
3º Quartil	0,5800
Máximo	1,2700
Assimetria.	0,93
Curtose	2,30

A Tabela 4 mostra que a média do índice de Theil é 0,5214, com um coeficiente de variação de 20,92% e variando de 0,19 a 1,27.

10% mais ricos / 40% mais pobres

Tabela 5: 10% mais ricos / 40% mais pobres

Média	24,5500
Desvio Padrão	57,5000
Coefficiente de Variação	234,1800
Mínimo	4,3300
1º Quartil	13,7000
Mediana	17,9300
3º Quartil	24,4400
Máximo	2631,8300
Assimetria.	28,76
Curtose	1062,69

A média da taxa dos 10% mais ricos / 40% mais pobres é 24,55, com um coeficiente de variação de 234,18%. Observa-se ainda que 50% da taxa dos 10% mais ricos / 40% mais pobres são menores ou iguais a 17,93, variando de 4,33 à 2631,83, com um coeficiente de assimetria de 28,76% (Tabela 5).

Renda per Capita

Tabela 6: Renda per Capita

Média	171,3300
Desvio Padrão	96,3100
Coefficiente de Variação	56,2100
Mínimo	31,9300
1º Quartil	86,9200
Mediana	159,5700
3º Quartil	233,2200
Máximo	954,6500
Assimetria.	1,15
Curtose	2,80

A Tabela 6 mostra que a média da renda per Capita por município é 171,33, com um coeficiente de variação de 56,21%, variando de 31,93 à 954,65.

Responsável / individual, rendimento nominal (Média)

Tabela 7: Responsável / individual, rendimento nominal (Média)

Média	393,5500
Desvio Padrão	202,2500
Coefficiente de Variação	51,3900
Mínimo	99,0200
1º Quartil	223,1000
Mediana	365,5600
3º Quartil	515,1400
Máximo	2119,9600
Assimetria.	1,28
Curtose	3,73

A Tabela 7 mostra que a média das médias de responsável / individual, rendimento nominal é 393,55, variando de 99,02 a 2119,96, em que 50% das médias de responsável / individual, rendimento nominal está entre 223,10 e 515,14.

Responsável / individual, rendimento nominal (Mediana)

Tabela 8: Responsável / individual, rendimento nominal (Mediana)

Média	222,0200
Desvio Padrão	99,7100
Coefficiente de Variação	44,9100
Mínimo	0,0000
1º Quartil	151,0000
Mediana	180,0000
3º Quartil	300,0000
Máximo	1000,0000
Assimetria.	1,51
Curtose	3,62

A média da mediana da taxa de responsável / individual, rendimento nominal é 222,02, com coeficiente de variação de 44,91%, variando de 0,00 à 1.000,00 (Tabela 8).

Proporção da população ocupada

Tabela 9: Proporção da população ocupada

Média	0,4884
Desvio Padrão	0,0943
Coefficiente de Variação	19,3200
Mínimo	0,1655
1º Quartil	0,4302
Mediana	0,4818
3º Quartil	0,5344
Máximo	0,9424
Assimetria.	0,62
Curtose	1,30

A Tabela 9 mostra que a média da proporção da população ocupada é 0,4884, com coeficiente de variação de 19,32%. Observa-se ainda que 50% da proporção da população ocupada está entre 0,43 e 0,53.

Longevidade (Saúde)

Esperança de vida ao nascer

Tabela 10: Esperança de vida ao nascer

Média	67,7760
Desvio Padrão	4,8460
Coefficiente de Variação	7,1500
Mínimo	54,9000
1º Quartil	64,5600
Mediana	68,2800
3º Quartil	71,4500
Máximo	78,1800
Assimetria.	-0,37
Curtose	-0,53

A média da esperança de vida ao nascer é 67,78, em que 50% da esperança de vida ao nascer está entre 64,56 e 71,45 (Tabela 10).

Mortalidade até um ano de idade

Tabela 11: Mortalidade até um ano de idade

Média	33,9650
Desvio Padrão	18,3820
Coefficiente de Variação	54,1200
Mínimo	5,3800
1º Quartil	18,6200
Mediana	29,3800
3º Quartil	45,9200
Máximo	98,1200
Assimetria.	0,72
Curtose	-0,34

A média da mortalidade até um ano de idade é 33,965, com um coeficiente de variação de 54,12%, variando de 5,38 à 98,12 (Tabela 11).

Socialização (Educação)

Porcentagem 18 a 22 anos com acesso ao curso superior

Tabela 12: Porcentagem de pessoas de 18 a 22 anos com acesso ao curso superior

Média	3,0130
Desvio Padrão	3,6570
Coefficiente de Variação	121,3700
Mínimo	0,0000
1º Quartil	0,5100
Mediana	1,6400
3º Quartil	4,2000
Máximo	30,7400
Assimetria.	2,11
Curtose	5,76

A Tabela 12 mostra que a média da porcentagem pessoas de 18 a 22 anos com acesso ao curso superior é 3,01, com um coeficiente de variação de 121,37%. Observa-se ainda que esta porcentagem varia de 0,00 à 30,74.

Porcentagem 15 anos ou mais com menos de quatro anos de estudo

Tabela 13: Porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo

Média	41,3290
Desvio Padrão	15,5020
Coefficiente de Variação	37,5100
Mínimo	6,2400
1º Quartil	29,0200
Mediana	39,2100
3º Quartil	54,4100
Máximo	86,5800
Assimetria.	0,23
Curtose	-0,87

A Tabela 13 mostra que a média da porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo é 41,33, com um coeficiente de variação de 37,51%, variando de 6,24 à 86,58.

Porcentagem 15 anos ou mais analfabetas

Tabela 14: Porcentagem de pessoas de 15 anos ou mais analfabetas

Média	21,7000
Desvio Padrão	12,4230
Coefficiente de Variação	57,2500
Mínimo	0,9100
1º Quartil	11,6300
Mediana	17,9100
3º Quartil	31,9900
Máximo	60,6600
Assimetria.	0,58
Curtose	-0,75

A média da porcentagem de pessoas de 15 anos ou mais analfabetas é 21,70, variando de 0,91 à 60,66 (Tabela 14).

Média de anos de estudo das pessoas de 25 anos ou mais

Tabela 15: Média de anos de estudo das pessoas de 25 anos ou mais

Média	4,0471
Desvio Padrão	1,2849
Coefficiente de Variação	31,7500
Mínimo	0,8100
1º Quartil	3,0200
Mediana	4,0800
3º Quartil	4,9200
Máximo	9,6500
Assimetria.	0,27
Curtose	-0,23

A Tabela 15 mostra que a média das médias de anos de estudo das pessoas de 25 anos ou mais é 4,08, com um coeficiente de variação de 31,75%, variando de 3,02 à 9,65.

Violência (Segurança)

Taxa Bruta Acidentes de transporte (10.000)

Tabela 16: Taxa bruta acidentes de transporte (10.000)

Média	1,6680
Desvio Padrão	1,8957
Coefficiente de Variação	113,6500
Mínimo	0,0000
1º Quartil	0,0000
Mediana	1,2712
3º Quartil	2,5323
Máximo	20,8768
Assimetria.	2,05
Curtose	8,24

Observa-se que a média da taxa bruta acidentes de transporte (por 10.000 pessoas) é 2,67, com um coeficiente de variação de 113,65%. Observa-se ainda que a taxa bruta acidentes de transporte varia de 0,00 à 20,88, com um coeficiente de assimetria de 2,05 (Tabela 16).

Taxa Bruta de Homicídios (10.000)

Tabela 17: Taxa bruta de homicídios (10.000)

Média	1,3220
Desvio Padrão	1,8430
Coefficiente de Variação	139,4200
Mínimo	0,0000
1º Quartil	0,0000
Mediana	0,6793
3º Quartil	1,9806
Máximo	16,3934
Assimetria.	2,27
Curtose	7,55

A Tabela 17 mostra que a taxa bruta de homicídios (por 10.000 habitantes) varia de 0 a 16,39, em que a taxa média é 1,32 com um coeficiente de variação de 139,42. Observa-se ainda que o coeficiente de assimetria é 2,27 e 50% dos municípios apresentam a taxa bruta de homicídios menor ou igual a 0,68.

População residente por município

Tabela 18: População residente por município

Média	30923,0000
Desvio Padrão	187153,0000
Coefficiente de Variação	605,2300
Mínimo	795,0000
1º Quartil	5154,0000
Mediana	10425,0000
3º Quartil	21451,0000
Máximo	10434252,0000
Assimetria.	38,70
Curtose	1930,11

A Tabela 18 mostra que em média existem 30923 pessoas por município, com um desvio padrão de 187153 e um coeficiente de variação de 605,23%. Observa-se ainda que a população residente por município varia de 795 à 10434252.

Número de domicílios por município

Tabela 19: Número de domicílios por município

Média	8291,0000
Desvio Padrão	54503,0000
Coefficiente de Variação	657,4000
Mínimo	227,0000
1° Quartil	1329,0000
Mediana	2647,0000
3° Quartil	5457,0000
Máximo	3039104,0000
Assimetria.	39,82
Curtose	1989,10

Em média existem 8291 domicílios por município, em que o número de domicílios varia de 227 à 3039104, com um desvio padrão de 54.503. Observa-se ainda que o coeficiente de assimetria é 39,82 (Tabela 19).

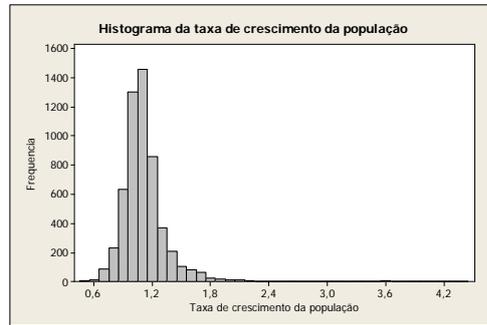


Figura 1: Histograma da taxa de crescimento da população

A Figura 1 indica que a taxa de crescimento da população é assimétrica à direita.

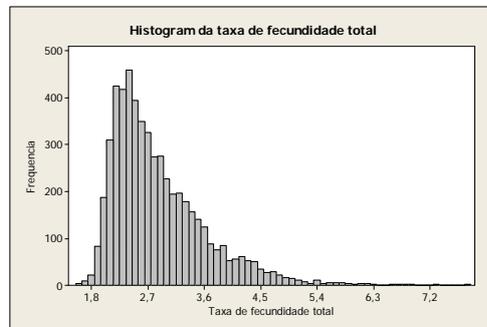


Figura 2: Histograma da taxa de fecundidade total

Aparentemente a taxa de fecundidade total é assimétrica à direita (Figura 2).

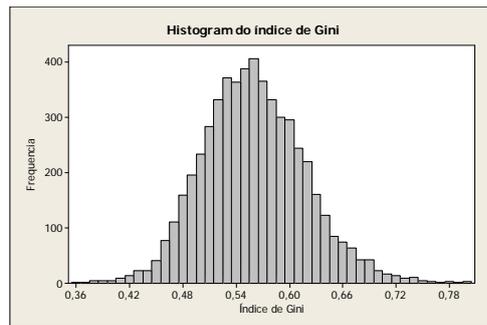


Figura 3: Histograma do índice de Gini

Analisando a Figura 3, observa-se que aparentemente o índice de Gini é simétrico.

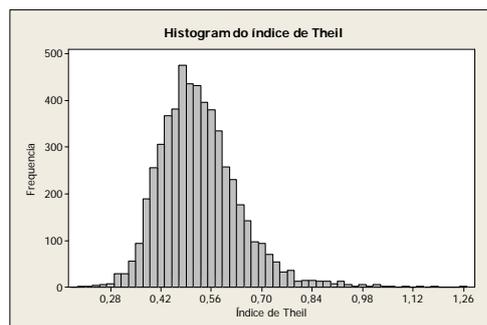


Figura 4: Histograma do índice de Theil

A Figura 4 indica que o índice de Theil é levemente assimétrico à direita.

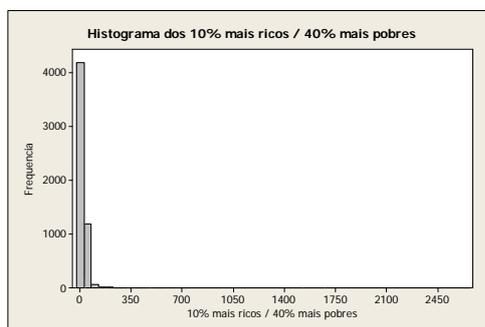


Figura 5: Histograma dos 10% mais ricos / 40% mais pobres

A Figura 5 mostra que a taxa dos 10% mais ricos / 40% mais pobres é assimétrica à direita.

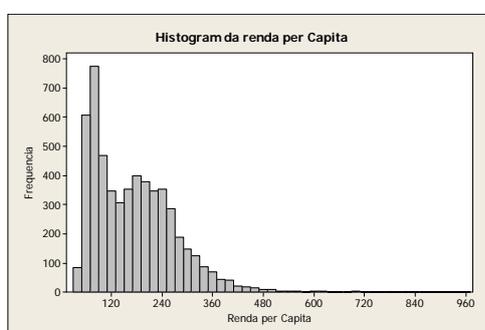


Figura 6: Histograma da renda per Capita

A renda per Capita é assimétrica à direita (Figura 6).

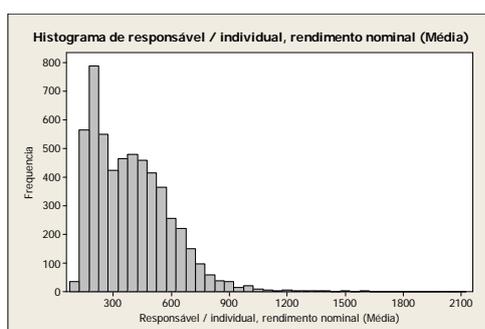


Figura 7: Histograma de responsável / individual, rendimento nominal (Média)

A Figura 7 indica que a média da taxa do responsável / individual, rendimento nominal é assimétrica à direita.

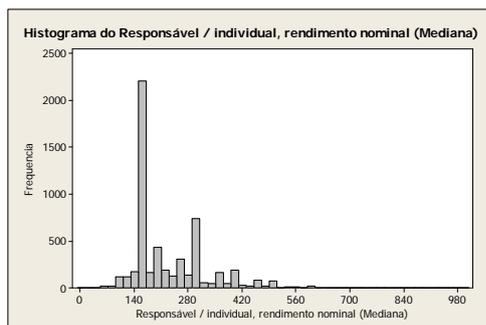


Figura 8: Histograma de responsável / individual, rendimento nominal (Mediana)

A taxa da mediana de responsável / individual, rendimento nominal é levemente assimétrica à direita (Figura 8).

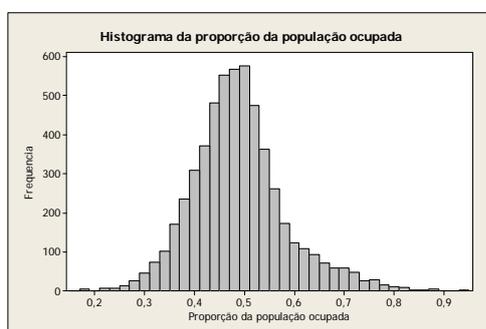


Figura 9: Histograma da proporção da população ocupada

Aparentemente a proporção da população ocupada é simétrica (Figura 9).

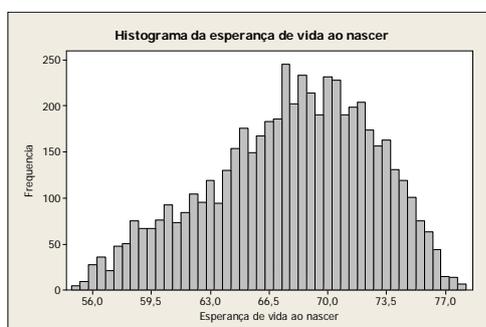


Figura 10: Histograma da esperança de vida ao nascer

A Figura 10 mostra que a esperança de vida ao nascer é levemente assimétrica à esquerda.

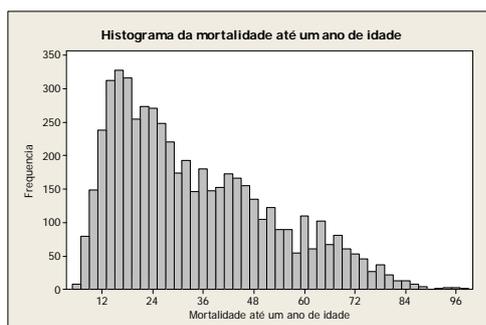


Figura 11: Histograma da mortalidade até um ano de idade

Aparentemente a mortalidade até um ano de idade é assimétrica à direita (Figura 11).

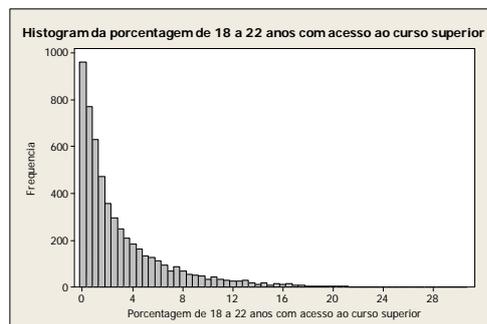


Figura 12: Histograma da porcentagem de 18 a 22 anos com acesso ao curso superior

A Figura 12 mostra que a porcentagem pessoas de 18 a 22 anos com acesso ao curso superior é assimétrica à direita.

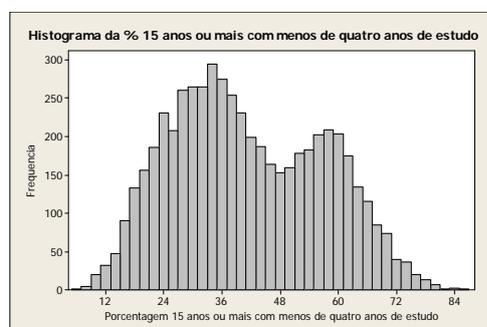


Figura 13: Histograma da porcentagem de 15 anos ou mais com menos de quatro anos de estudo

A Figura 13 indica que a porcentagem de pessoas de 15 anos ou mais com menos de quatro anos de estudo tem uma grande concentração ao redor de 35% e de 60%.

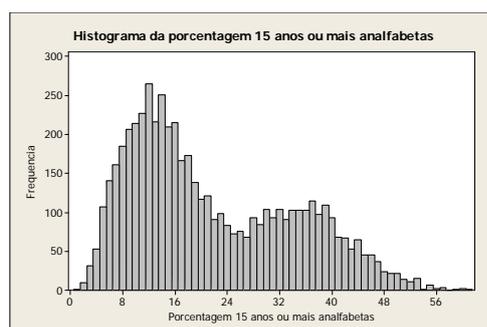


Figura 14: Histograma da porcentagem de 15 anos ou mais analfabetas

Aparentemente a porcentagem de pessoas de 15 anos ou mais analfabetas é assimétrica à direita (Figura 14).

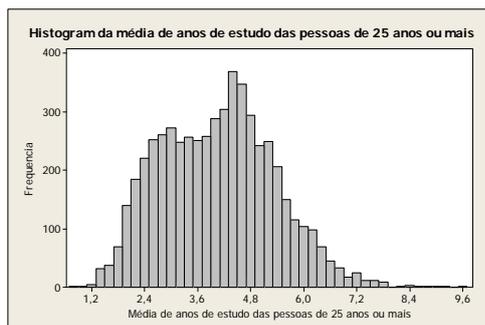


Figura 15: Histograma da média de anos de estudo das pessoas de 25 anos ou mais

A Figura 15 indica que a média de anos de estudo das pessoas de 25 anos ou mais está concentrada entre dois e cinco anos.

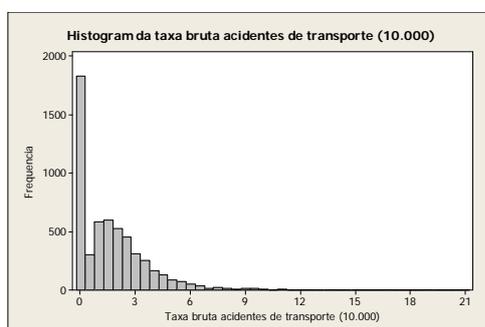


Figura 16: Histograma da taxa bruta de acidentes de transporte (10.000)

Observa-se que a taxa bruta de acidentes de transporte (por 10.000 pessoas) é assimétrica à direita (Figura 16).

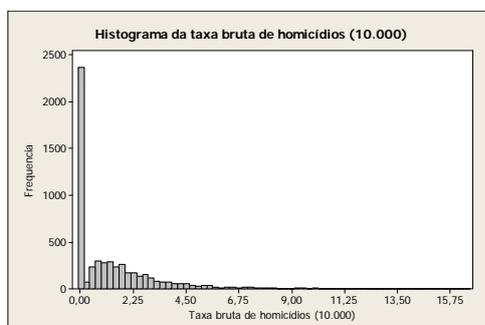


Figura 17: Histograma da taxa bruta de homicídios (10.000)

A Figura 17 indica que a taxa bruta de homicídios (por 10.000 pessoas) é assimétrica à direita, apresentando uma grande concentração de taxas iguais a zero.

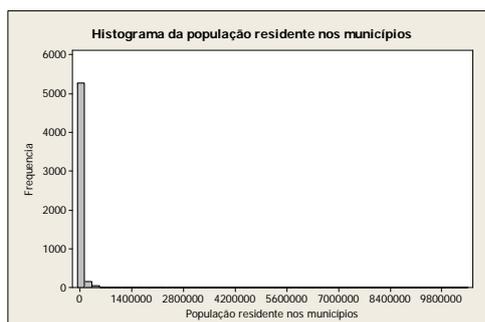


Figura 18: Histograma da população residente nos municípios

18). Observa-se que a concentração da população residente nos municípios é assimétrica à direita (Figura

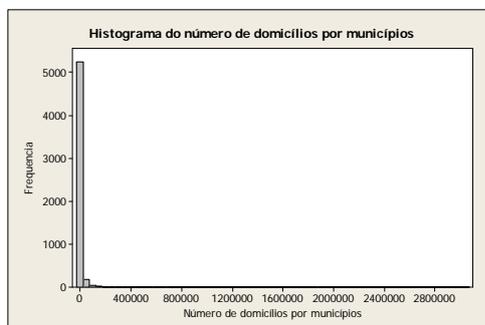


Figura 19: Histograma do número de Domicílios por municípios

A Figura 19 indica que o número de domicílios por município é assimétrica à direita.

VARIÁVEIS EXÓGENAS

População_Resident e Domicílios, rs=**0,9914**

RENDA

Índice de Gini e Índice de Theil rs=**0,8789**

Índice de Gini e DEZ_Ricos_40_Pobres rs=**0,9476**

Índice de Theil e DEZ_Ricos_40_Pobres rs=**0,8045**

Renda per Capita e Responsável, rendimento nominal (Média) rs=**0,9613**

Renda per Capita e Responsável, rendimento nominal (Mediana) rs=**0,9290**

Responsável, rendimento nominal (Média) e Responsável, rendimento nominal (Mediana) rs=**0,9279**

SAÚDE

Esperança de vida ao nascer e Mortalidade até um ano de idade rs=**-0,9518**

EDUCAÇÃO

Dezoito a Vinte e dois anos com acesso ao curso superior e Quinze anos ou mais com menos de quatro anos de estudo rs=**-0,7365**

Dezoito a Vinte e dois anos com acesso ao curso superior e Quinze anos ou mais analfabetas rs=**-0,6945**

Dezoito a Vinte e dois anos com acesso ao curso superior e Média de anos de estudo das pessoas de 25 anos ou mais rs=**0,7731**

Quinze anos ou mais com menos de quatro anos de estudo e Quinze anos ou mais analfabetas rs=**0,9668**

Quinze anos ou mais com menos de quatro anos de estudo e Média de anos de estudo das pessoas de 25 anos ou mais rs=**-0,9583**

Quinze anos ou mais analfabetas e Média de anos de estudo das pessoas de 25 anos ou mais rs=**-0,9212**

Dados Completos sem correlação superior à 0,5

DINÂMICA POPULACIONAL

Taxa de Crescimento da População e Taxa de fecundidade total rs=0,0741

RENDA

Índice de Gini e Renda per Capita rs=-0,2678

Índice de Gini e Responsável, rendimento nominal (Média) rs=-0,2242

Índice de Gini e Responsável, rendimento nominal (Mediana)rs=-0,4004

Índice de Gini e Proporcão da População Ocupada rs=-0,3203

Índice de Theil e Renda per Capita rs=-0,0446

Índice de Theil e Responsável, rendimento nominal (Média) rs=-0,0071

Índice de Theil e Responsável, rendimento nominal (Mediana) rs=-0,2039

Índice de Theil e Proporcão da População Ocupada rs=-0,1092

DEZ_Ricos_40_Pobres e Renda per Capita rs=-0,2972

DEZ_Ricos_40_Pobres e Responsável, rendimento nominal (Média) rs=-0,2640

DEZ_Ricos_40_Pobres e Responsável, rendimento nominal (Mediana) rs=-0,4204

DEZ_Ricos_40_Pobres e Proporcão da População Ocupada rs=-0,3342

Renda per Capita e Proporcão da População Ocupada rs=0,4873

Responsável, rendimento nominal (Média) e e Proporcão da População Ocupada rs=0,4789

Responsável, rendimento nominal (Mediana) e Proporcão da População Ocupada rs=0,4864

SAÚDE

Esperança de vida ao nascer e Tx_Bruta_Desnutrição (10.000) rs=-0,0239

Esperança de vida ao nascer e Tx_Bruta_Aids (10.000) rs=0,4463

Esperança de vida ao nascer e Tx_Bruta_Dengue (10.000) rs=-0,1769

Esperança de vida ao nascer e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=-0,2096

Mortalidade até um ano de idade e Tx_Bruta_Desnutrição (10.000) rs=-0,0127

Mortalidade até um ano de idade e Tx_Bruta_Aids (10.000) rs=-0,4908

Mortalidade até um ano de idade e Tx_Bruta_Dengue (10.000) rs=0,1762

Mortalidade até um ano de idade e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=0,2252

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Aids (10.000) rs=0,0780

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Dengue (10.000) rs=0,0637

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=0,0109

Tx_Bruta_Aids (10.000) e Tx_Bruta_Dengue (10.000) rs=0,0074

Tx_Bruta_Aids (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=-0,1480

Tx_Bruta_Dengue (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=0,1902

SEGURANÇA

Tx Bruta de Homicídios (10.000) e Tx Bruta Acidentes de transporte (10.000) rs=0,2520

APENDICE T - Dados sem as observações com correlação superior à 0,5

VARIÁVEIS EXÓGENAS

População_Resident e Domicílios, rs=**0,9915**

RENDA

Índice de Gini e Índice de Theil rs=**0,8728**

Índice de Gini e DEZ_Ricos_40_Pobres rs=**0,9484**

Índice de Theil e DEZ_Ricos_40_Pobres rs=**0,7931**

Renda per Capita e Responsável, rendimento nominal (Média) rs=**0,9628**

Renda per Capita e Responsável, rendimento nominal (Mediana) rs=**0,9289**

Responsável, rendimento nominal (Média) e Responsável, rendimento nominal (Mediana) rs=**0,9289**

SAÚDE

Esperança de vida ao nascer e Mortalidade até um ano de idade rs=**-0,9529**

Mortalidade até um ano de idade e Tx_Bruta_Aids (10.000) rs=**-0,5025**

EDUCAÇÃO

Dezoito a Vinte e dois anos com acesso ao curso superior e Quinze anos ou mais com menos de quatro anos de estudo rs=**-0,7453**

Dezoito a Vinte e dois anos com acesso ao curso superior e Quinze anos ou mais analfabetas rs=**-0,6983**

Dezoito a Vinte e dois anos com acesso ao curso superior e Média de anos de estudo das pessoas de 25 anos ou mais rs=**0,7760**

Quinze anos ou mais com menos de quatro anos de estudo e Quinze anos ou mais analfabetas rs=**0,9667**

Quinze anos ou mais com menos de quatro anos de estudo e Média de anos de estudo das pessoas de 25 anos ou mais rs=**-0,9653**

Quinze anos ou mais analfabetas e Média de anos de estudo das pessoas de 25 anos ou mais rs=**-0,9266**

Dados sem as observações que tiveram 3 zeros com correlação inferior à 0,5

DINÂMICA POPULACIONAL

Taxa de Crescimento da População e Taxa de fecundidade total rs=0,0735

RENDA

Índice de Gini e Renda per Capita rs=-0,2428

Índice de Gini e Responsável, rendimento nominal (Média) rs=-0,2028

Índice de Gini e Responsável, rendimento nominal (Mediana) rs=-0,3794

Índice de Gini e Proporcao da Populacao Ocupada rs=-0,2850

Índice de Theil e Renda per Capita rs=-0,0156

Índice de Theil e Responsável, rendimento nominal (Média) rs=0,0162

Índice de Theil e Responsável, rendimento nominal (Mediana) rs=-0,1814

Índice de Theil e Proporcao da Populacao Ocupada rs=-0,0775

DEZ_Ricos_40_Pobres e Renda per Capita rs=-0,2904

DEZ_Ricos_40_Pobres e Responsável, rendimento nominal (Média) rs=-0,2621

DEZ_Ricos_40_Pobres e Responsável, rendimento nominal (Mediana) rs=-0,4149

DEZ_Ricos_40_Pobres e Proporcao da Populacao Ocupada rs=-0,3081

Renda per Capita e Proporcao da Populacao Ocupada rs=0,4725

Responsável, rendimento nominal (Média) e e Proporcao da Populacao Ocupada rs=0,4644

Responsável, rendimento nominal (Mediana) e Proporcao da Populacao Ocupada rs=0,4673

SAÚDE

Esperança de vida ao nascer e Tx_Bruta_Desnutrição (10.000) rs=0,0167

Esperança de vida ao nascer e Tx_Bruta_Aids (10.000) rs=0,4627

Esperança de vida ao nascer e Tx_Bruta_Dengue (10.000) rs=-0,1268

Esperança de vida ao nascer e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=-0,1926

Mortalidade até um ano de idade e Tx_Bruta_Desnutrição (10.000) rs=-0,0489

Mortalidade até um ano de idade e Tx_Bruta_Dengue (10.000) rs=0,1377

Mortalidade até um ano de idade e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) rs=0,2117

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Aids (10.000) rs=0,0907

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Dengue (10.000) $rs=-0,1409$

Tx_Bruta_Desnutrição (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) $rs=-0,1229$

Tx_Bruta_Aids (10.000) e Tx_Bruta_Dengue (10.000) $rs=0,0228$

Tx_Bruta_Aids (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) $rs=-0,1692$

Tx_Bruta_Dengue (10.000) e Tx_Bruta_Leishmaniose Tegumentar Americana (10.000) $rs=0,0447$

SEGURANÇA

Tx Bruta de Homicídios (10.000) e Tx Bruta Acidentes de transporte (10.000) $rs=0,2598$

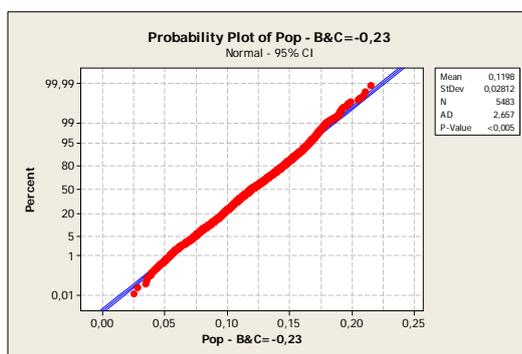
APENDICE U - Dados transformados x^λ

Este arquivo mostra os testes de normalidade para os dados transformados X^λ , em que λ é obtido pela transformação de Box&Cox.

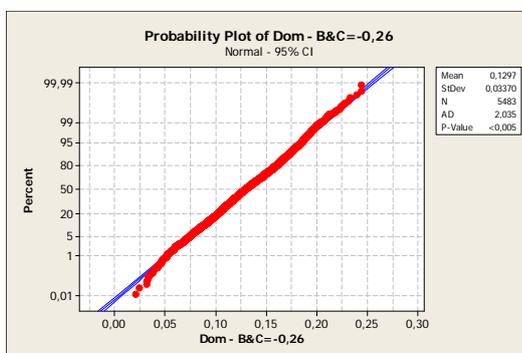
Observa-se que, mesmo após a transformação, a normalidade não foi aceita. Isto provavelmente se deve ao fato dos dados originais apresentarem grande assimetria.

Obs: Note que logo após o nome das variáveis está "B&C=". O valor que aparece logo após o "=" é o " λ " encontrado na transformação de Box&Cox. Existem três casos em que aparece "?" após o "=". Isto se deve ao fato de não existir um " λ " para transformar os dados.

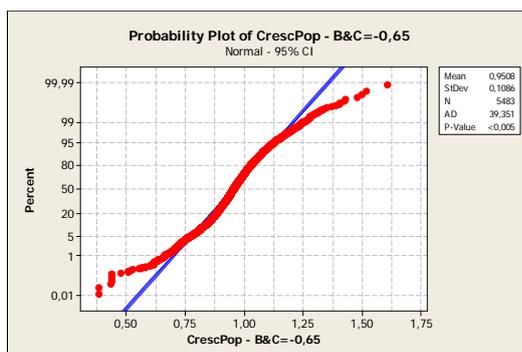
População - B&C=-0,23



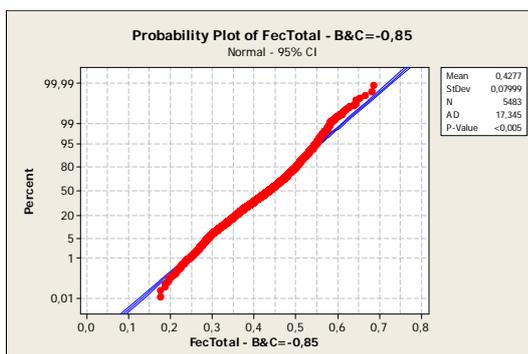
Doenças relacionadas ao saneamento ambiental inadequado - B&C=-0,26



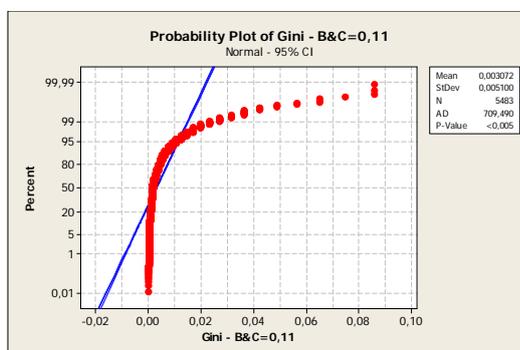
Taxa de crescimento da população - B&C=-0,65



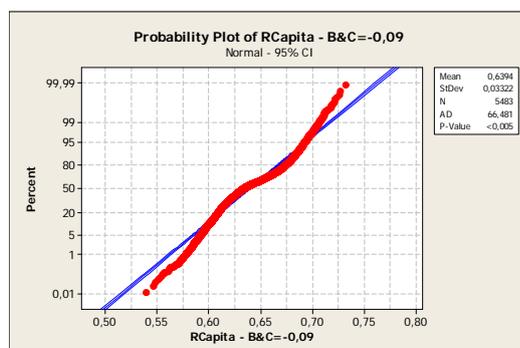
Taxa de fecundidade - B&C=-0,85



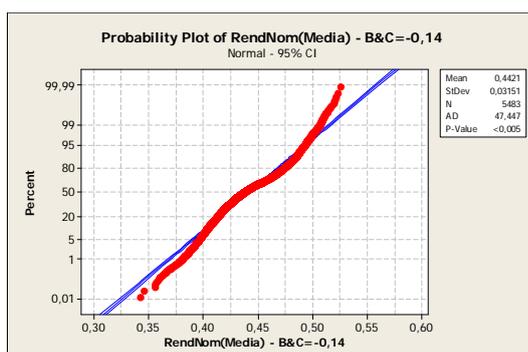
Índice de Gini da distribuição do rendimento - B&C=0,11



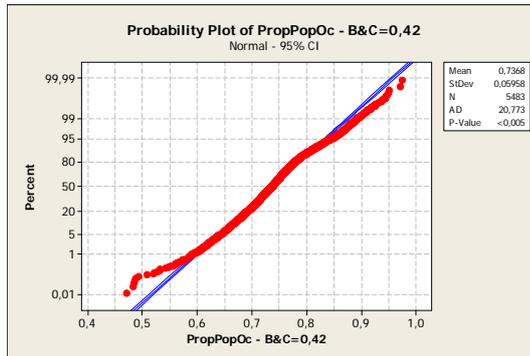
Rendimento familiar per capita - B&C=-0,09



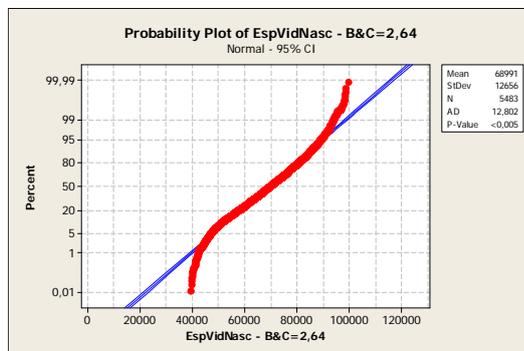
Rendimento médio mensal - B&C=-0,14



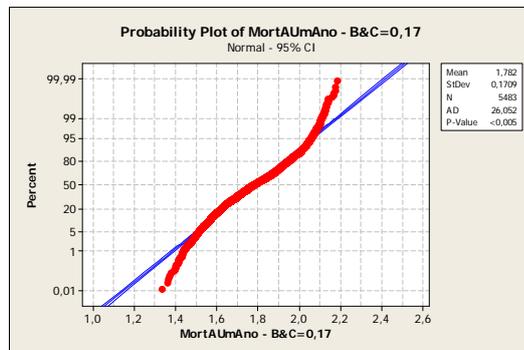
Taxa de desocupação - B&C=0,42



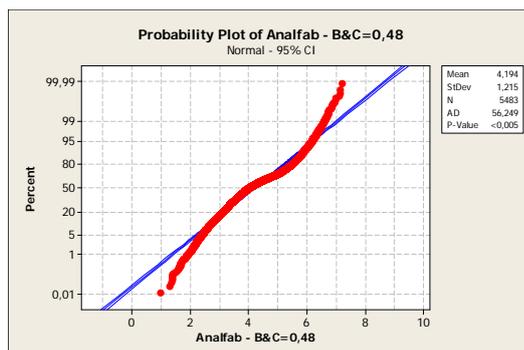
Esperança de vida ao nascer - B&C=2,64



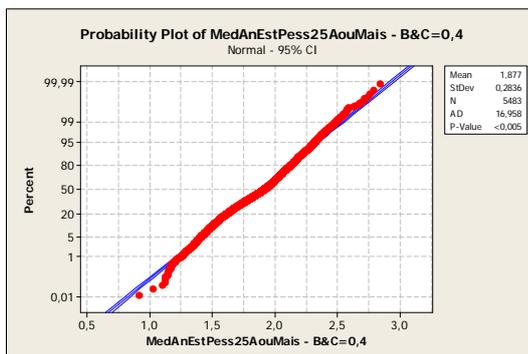
Taxa de mortalidade infantil - B&C=0,17



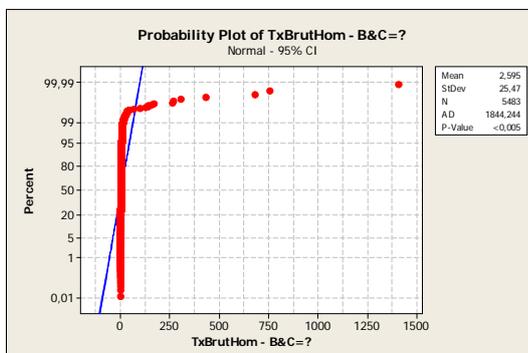
Taxa de alfabetização - B&C=0,48



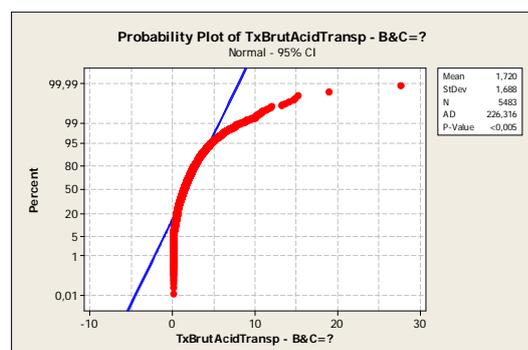
Escolaridade - B&C=0,46



Coefficiente de mortalidade por homicídios - B&C=?



Coefficiente de mortalidade por acidentes de transporte - B&C=?



Este apêndice trata da aplicação de modelos de equações estruturais com o objetivo de explicar o desenvolvimento social por meio de seus respectivos indicadores, conceitos e construtos associados ao termo teórico. Para isto, foram utilizados quinze indicadores sociais de 5.483 municípios brasileiros referentes ao ano 2000. Concluí-se que o modelo hipotético de influências representado por diagrama circular não pôde ser ajustado, por meio do método de Modelagem de Equações Estruturais (MEE) devido a problemas de convergência. Alterações nesse modelo foram propostas de acordo com os dados disponíveis e, em seguida, foi submetido a uma verificação a partir dos resultados do Índice de Desenvolvimento Humano (IDH) como função das variáveis latentes, sistema produtivo, longevidade e sócio-demográfico usando a modelagem de equações estruturais.

1 Desenvolvimento Social: Modelagem de Equações Estruturais e fenômenos complexos

Com o objetivo de encontrar um modelo de equações estruturais que explique o desenvolvimento social no Brasil a partir da definição de variáveis não observáveis ou latentes, foram selecionados cinco construtos e seus respectivos conceitos e indicadores.

- Dinâmica Populacional:
 - Taxa de Crescimento da População
 - Taxa de Fecundidade total

- Renda e Trabalho:
 - Índice de Gini
 - Índice de Theil
 - Razão entre os 10% mais ricos e 40% mais pobres
 - Renda per Capita
 - Rendimento nominal médio do responsável no domicílio
 - Rendimento nominal mediano do responsável no domicílio
 - Proporção da População Ocupada

- Saúde:
 - Esperança de vida ao nascer
 - Mortalidade até um ano de idade
 - Taxa Bruta de Aids
 - Taxa Bruta de Dengue

- Educação
 - Proporção de indivíduos entre 18 e 22 anos com acesso ao curso superior

Proporção de indivíduos entre 15 anos ou mais com menos de 4 anos de estudo

Proporção de analfabetos com 15 anos ou mais

Média de anos de estudo das pessoas de 25 anos ou mais

➤ Segurança

Taxa Bruta de Homicídios

Taxa Bruta de Acidentes de transporte

➤ Desenvolvimento Humano:

IDHM

IDHM-Renda

IDHM-Longevidade

IDHM-Educação

Os indicadores e variáveis utilizadas nesse apêndice foram os seguintes: Taxa de Crescimento da População; Taxa de fecundidade total; Índice de Gini; Renda per Capita; Rendimento Nominal Médio do Responsável no Domicílio; Proporção da População Ocupada; Esperança de vida ao nascer; Mortalidade até um ano de idade; Taxa Bruta de AIDS (10.000); Proporção de Analfabetos com 15 anos ou mais; Média de anos de Estudo das Pessoas de 25 anos ou mais; IDHM (1991, 2000); IDHM-Renda (1991, 2000); IDHM-Longevidade (1991, 2000); IDHM-Educação (1991, 2000). Como a unidade de observação é o município, o número total de municípios com informações para essas variáveis foi de 5.483. Com isso foram identificados quatro construtos, que são definidos na Tabela 1.

Tabela 1: Descrição dos construtos utilizados no estudo.

Sistema Produtivo	Longevidade	Sócio Demográfico	Desenv. Humano
Índice de Gini	Esperança de vida ao nascer	Taxa de Crescimento da População	IDHM (1991, 2000)
Renda per Capita	Mortalidade até um ano de idade	Taxa de fecundidade total	IDHM-Renda (1991, 2000)
Rendimento nominal médio do responsável no domicílio	Taxa Bruta de Aids (10.000)	Proporção de analfabetos com 15 anos ou mais	IDHM-Longevidade (1991, 2000)
Proporção da População Ocupada	-	Média de anos de estudo	IDHM-Educação (1991, 2000)

Nota: a primeira linha contém a denominação dos construtos e as colunas a partir da segunda linha representam os indicadores que compõem os construtos.

2 Modelos de Equações Estruturais

A *MEE* é uma importante técnica estatística que possibilita a análise de fenômenos complexos de diversas áreas do conhecimento. Segundo *Klem e Maruyana (1995)*, dois resultados essenciais desta técnica são esperados: primeiro, uma estimativa da magnitude dos efeitos estabelecida entre as variáveis é obtida. Estas estimativas são condicionadas ao fato da especificação do modelo estar correta. Segundo, é possível testar se o modelo é consistente com os dados observados, apesar de não se poder afirmar ainda assim que o modelo esteja correto, mas unicamente que ele é plausível (*Farias e Santos, 2000*).

Esse procedimento permite que se estude a causalidade entre as variáveis de interesse em estudo, partindo de uma justificativa teórica sólida (*Farias e Santos, 2000*). Nos MEEs as variáveis podem ser classificadas como: variáveis endógenas e variáveis exógenas. As primeiras são aquelas que sofrem efeitos exercidos por outras variáveis, enquanto as segundas são aquelas que não sofrem influência de outras (*Farias e Santos, 2000; Loelin, 1998*).

No modelo de equações estruturais, também há variáveis que não podem ser observadas diretamente, chamadas de variáveis latentes ou construtos. Essas variáveis são medidas a partir de variáveis observadas. As relações entre essas variáveis são determinadas a priori pelos estudiosos da área na qual a pesquisa será executada, isso reforçando o caráter confirmatório dessa técnica. Há dois tipos de modelos na *MEE*: o de medição ou mensuração e o estrutural. O primeiro descreve as relações entre as variáveis latentes e seus indicadores observados. O segundo define a relação entre as variáveis latentes exógenas e endógenas no modelo.

O pressuposto básico para a aplicação do primeiro é o de normalidade multivariada da distribuição conjunta das variáveis envolvidas no estudo. Os *softwares* mais comuns utilizados para a estimação dos parâmetros são o *LISREL* e o *AMOS*. Por meio dos mesmos é possível realizar as análises a partir do diagrama construído por eles, sem a necessidade de indicar as equações. Neste trabalho se fará uso do *LISREL*, usando o método de máxima verossimilhança para estimação dos parâmetros.

De acordo com *Pilati e Laros (2007)*, após a definição dos aspectos relativos à estimação dos parâmetros do modelo de equações estruturais, é necessário que o pesquisador se posicione a respeito dos índices de ajustes que serão utilizados para teste do modelo. Esse aspecto da *MEE* é importante e implicou no desenvolvimento de várias estratégias de avaliação da adequação do ajuste do modelo aos dados.

Segundo *Hox e Bechger (1998)*, o índice clássico de ajuste de um modelo é o qui-quadrado (χ^2). Para esses autores o grande problema desse índice é seu alto grau de sensibilidade ao tamanho das amostras, o que torna o índice significativo, o que não é desejado, pois indica a rejeição do modelo. O contrário também é verdadeiro, pois se a amostra é reduzida, o índice será não-significativo, o que pode levar o pesquisador a aceitar um modelo quando outros concorrentes são mais verossímeis. Por isso índices de adequação de modelos de equações estruturais precisam ser avaliados simultaneamente.

Os índices de adequação são classificados em cinco categorias (*Pilati e Laros, 2007; Ullman, 2007*). A primeira é chamada de índices comparativos usando a lógica de comparação da distribuição qui-quadrado de modelos independentes com qui-quadrado do modelo em teste. São exemplos desses índices o *NFI (Normed Fit Index)*, o *NNFI (Non-Normed Fit Index)*, o *IFI (Incremental Fit Index)* e o *CFI (Comparative Fit Index)*.

Um índice peculiar nessa categoria é o *RMSEA (Root Mean Square Error of Approximation)* que testa o ajuste próximo do modelo por meio da comparação entre o modelo em teste e um modelo saturado com o mesmo

conjunto de dados. A literatura aponta que valores entre 0,90 e 0,95 são indicadores de ajuste suficiente e valores acima de 0,95 são considerados bons ajustes. Em relação ao *RMSEA*, espera-se um valor menor do que 0,08 (Pilati e Laros, 2007; Hancock & Freeman, 2001; MacCallum, Browne & Sugawara, 1996) considerado como adequado, dado o poder do teste. Outra característica relevante do *RMSEA* é a possibilidade de estimação de intervalos de confiança do índice, o que auxilia no julgamento do ajuste. A segunda categoria, denominada de índice de adequação absoluto, tem apenas um representante, o *MFI*. Este índice é considerado absoluto porque não compara o modelo em teste com qualquer outro possível modelo (Pilati e Laros, 2007).

A terceira categoria de índices são aqueles relativos à proporção de variância explicada. São exemplos dessa categoria de índices: o *GFI* (*Goodness of Fit Index*) e o *AGFI* (*Adjusted Goodness of Fit Index*). Esses índices baseiam-se no cálculo da proporção da variância explicada por meio da estimação. No caso destes, também se espera valores próximos a 1,0 para a indicação de ajuste adequado do modelo. Pode-se aceitar um modelo com valores de 0,90 (Pilati e Laros, 2007; Hox e Bechger, 1998).

A quarta categoria de índices diz respeito ao grau de parcimônia do modelo testado. Fazem parte dessa categoria o *PGFI* (*Parcimony Goodness of Fit Index*), o *AIC* (*Akaike Information Criterion*) e o *CAIC* (*Consistent Akaike Information Criterion*). Esses índices dão informações sobre a adequação do modelo por meio de uma relação entre o número de parâmetros estimados e o número de pontos de dados na matriz de covariâncias (Pilati e Laros, 2007).

A última categoria de índices é a dos índices de ajuste baseados em resíduos. O principal representante é o *RMR* (*Root Mean Square Residual*) e sua variação *SRMR* (*Standardized Root Mean Square Residual*). Ambos avaliam a distância entre os elementos preditos por meio da estimação dos parâmetros do modelo e os dados observados na matriz de covariâncias. Nesse último caso, esperam-se valores próximos de zero como indicadores de ajuste do modelo.

A Figura 1 apresenta o modelo de mensuração, no qual são estimados os valores das cargas fatoriais de cada variável latente, além das correlações entre os construtos. Pode ser verificado que as variáveis, Renda per Capita e Rendimento nominal médio do responsável no domicílio apresentam as maiores contribuições no construto Sistema Produtivo (-1.00 e -0.95, respectivamente). Já no construto Longevidade, as maiores contribuições foram das variáveis Esperança de vida e Mortalidade ata um ano, 0.95 e -0.99, respectivamente. Em contrapartida, no construto Sócio Demográfico as variáveis, proporção de analfabetos com 15 anos ou mais e Média de anos de estudo das pessoas de 25 anos ou mais tiveram maior contribuição nesse construto.

Além disso, pode-se observar que os construtos sistema produtivo e Sócio Demográfico apresentam maior correlação, apesar de todas as variáveis latentes serem altamente correlacionadas. Observa-se, por exemplo, ainda que o erro associado à variável índice de Gini é 0.95 e o erro associado à renda per capita é 0.00. O nível descritivo associado à estatística de teste χ^2 aponta para a rejeição da hipótese de que o modelo teórico está bem ajustado ($p=0,000$). Além disso, o índice de ajuste *RMSEA* aponta para a mesma conclusão, uma vez que o *RMSEA* igual a 0,175 foi superior a 0.08. Os resultados de *NFI*, *NNFI*, *IFI* e *CFI* apresentam valores entre 0.9 e 0.95, indicando que o ajuste foi suficiente, em oposição aos resultados do *RMSEA* e χ^2 .

Como já foi mencionado a estatística χ^2 é muito sensível quando o número de observações é muito grande, como é o caso neste trabalho. O que se pode dizer observando esses índices conjuntamente é que o

modelo de mensuração das variáveis parece estar adequado, apesar dos valores *RMSEA* e χ^2 estarem apontando para uma conclusão contrária a esta. A partir desta análise, partiu-se para o modelo estrutural.

Como o interesse é verificar as relações dos construtos sistema produtivo e longevidade com o construto sócio demográfico, foi feito um modelo estrutural para ilustração dessas relações, o qual pode ser visualizado na Figura 2. Os resultados mostraram que o construto sistema produtivo exerce um maior efeito (-0,68) no sócio demográfico quando comparado com o efeito do construto longevidade (0,29).

O diagrama apresentado na Figura 3 ilustra as relações apenas com respeito ao IDHM (global), pois não foi possível ajustar modelos para as variáveis IDHM (RENDA), IDHM (EDUCAÇÃO) e o IDHM (LONGEVIDADE), devido aos problemas de convergência.

Figura 1: Modelo de Mensuração (AFC).

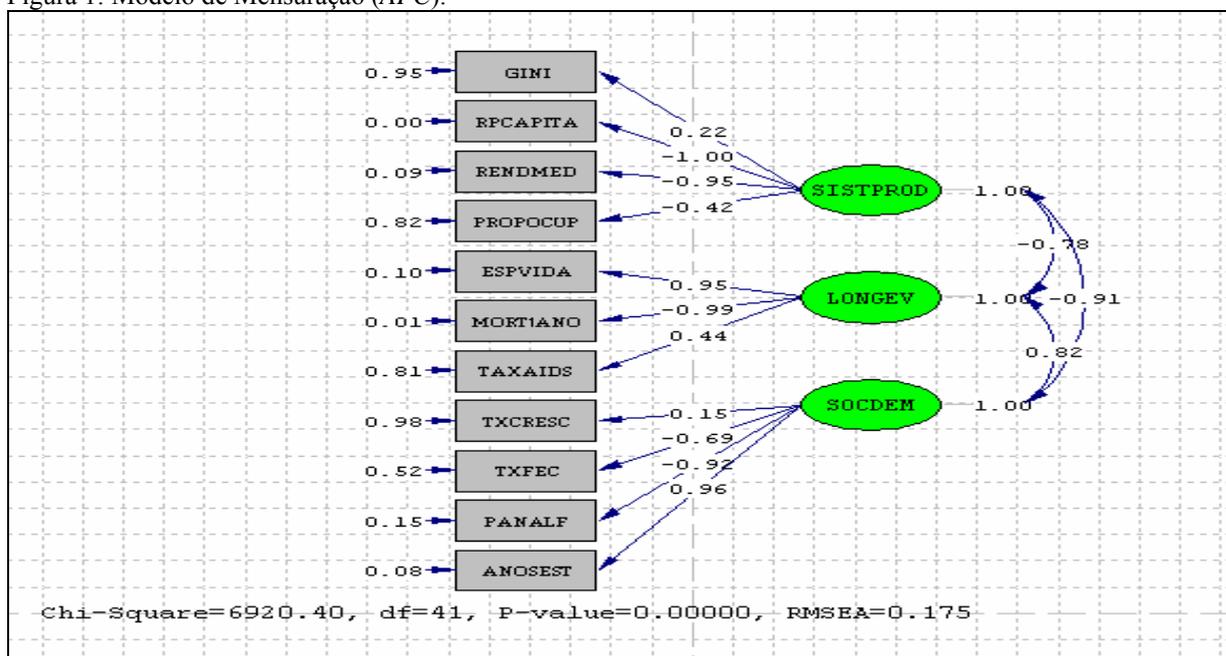
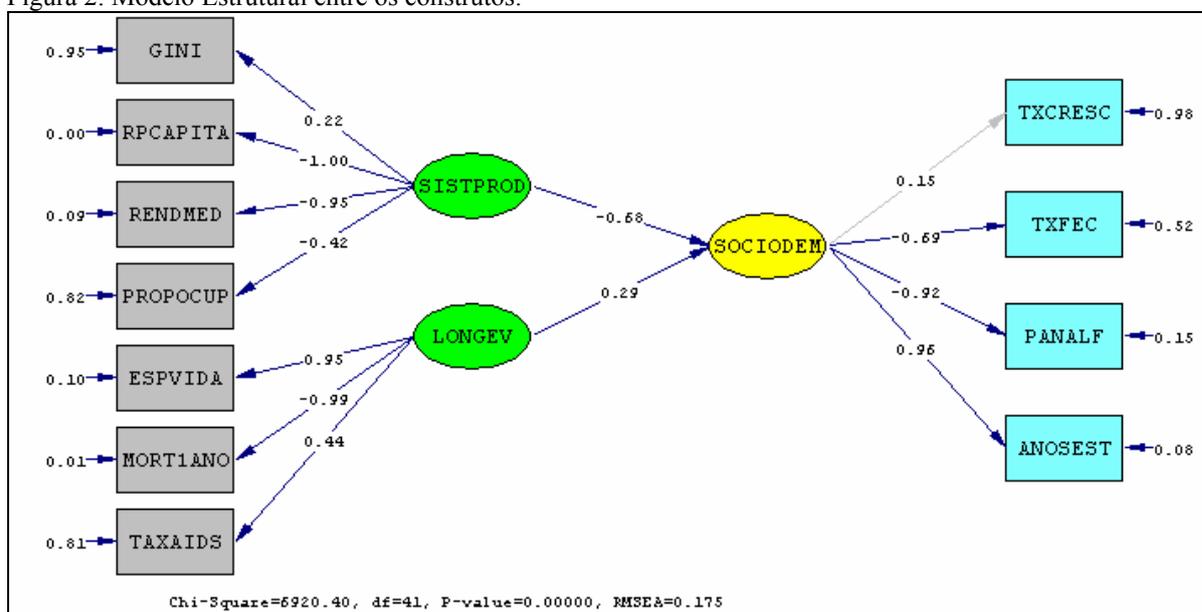


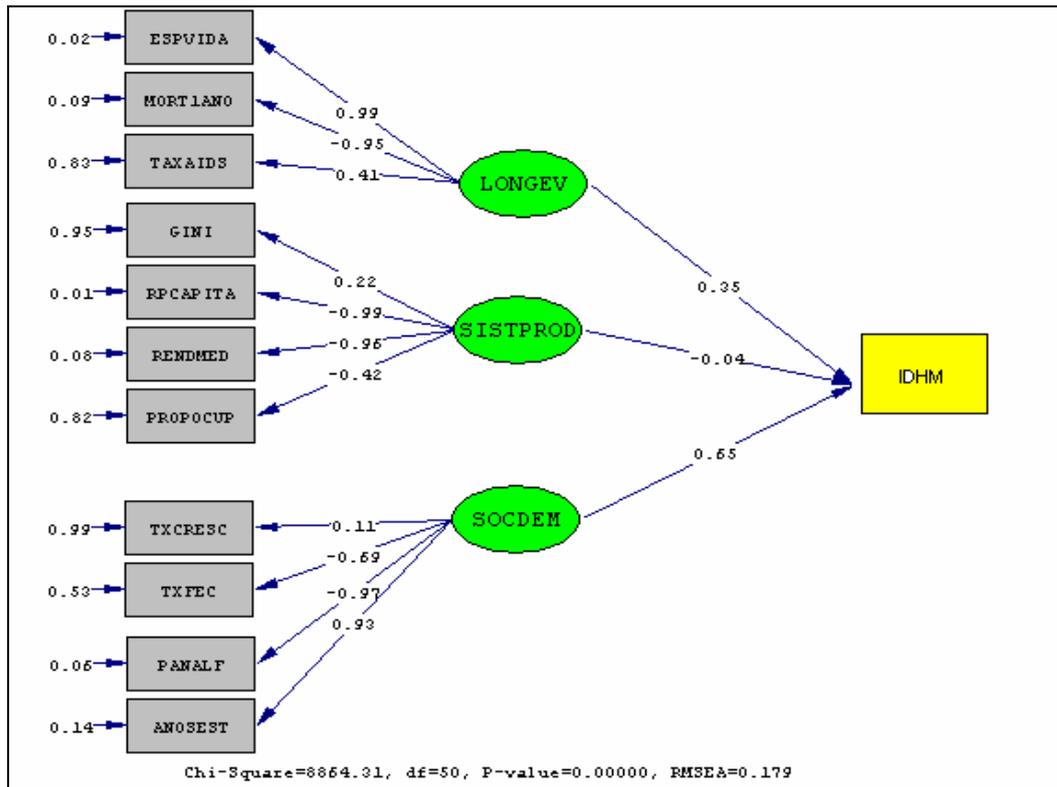
Figura 2: Modelo Estrutural entre os construtos.



Conforme a Figura 3, o construto que apresentou maior efeito sobre o IDHM foi o sócio demográfico (0.65), enquanto que o Sistema Produtivo teve menor efeito no IDHM (-0.04). O construto Longevidade teve efeito positivo (0.35) no IDHM. A única variável endógena do modelo apresentado na Figura 3 é o IDHM.

Os índices de bondade de ajuste, segundo *LISREL* resultados encontrados com a utilização do *LISREL* indicaram um χ^2 igual a 8864.31 (DF= 50, p= 0.000), *RMSEA* igual a 0.179 e os valores de *NFI*, *NNFI*, *IFI* e *CFI* ente 0.90 e 0.93. Apesar das divergências, pode-se considerar que o modelo está adequado, conforme discussão apresentada anteriormente.

Figura 3: Modelo Estrutural com IDHM.



DIAGNÓSTICOS

Figura 4: Diagnóstico do modelo de regressão contendo como variável resposta o DHM.

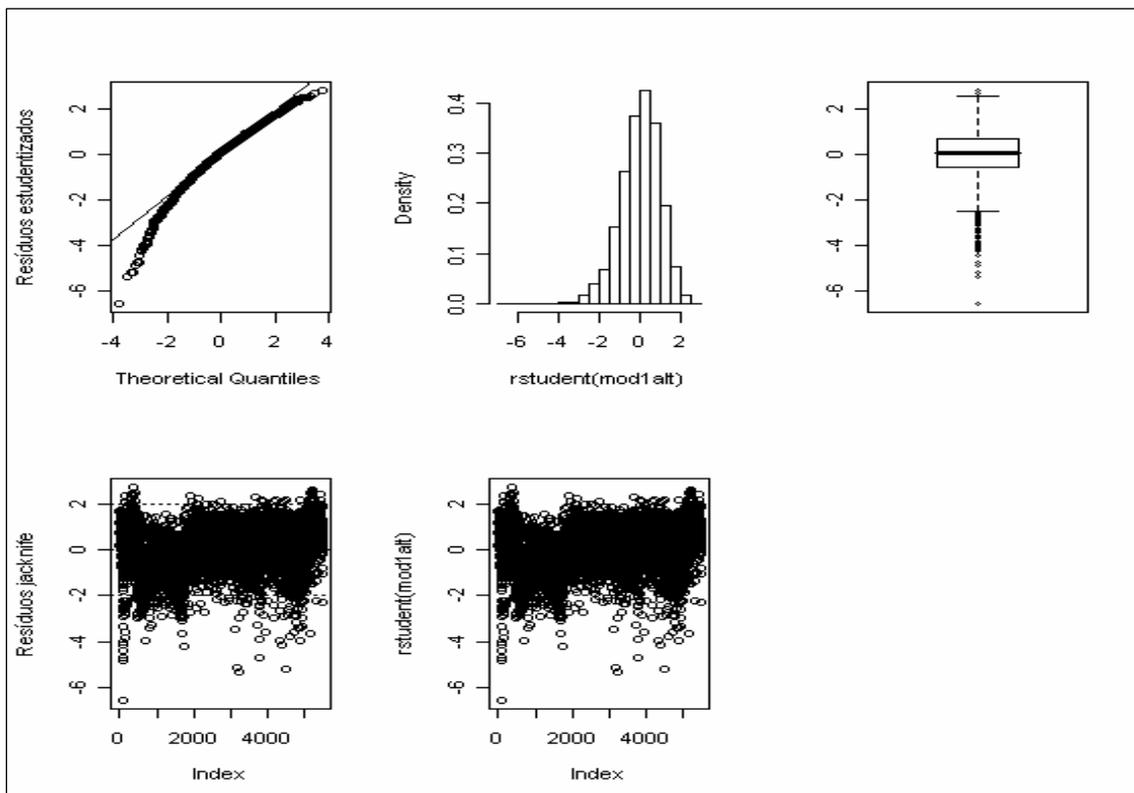


Figura 5: Diagnóstico do modelo de regressão contendo como variável resposta o IDHM (REND).

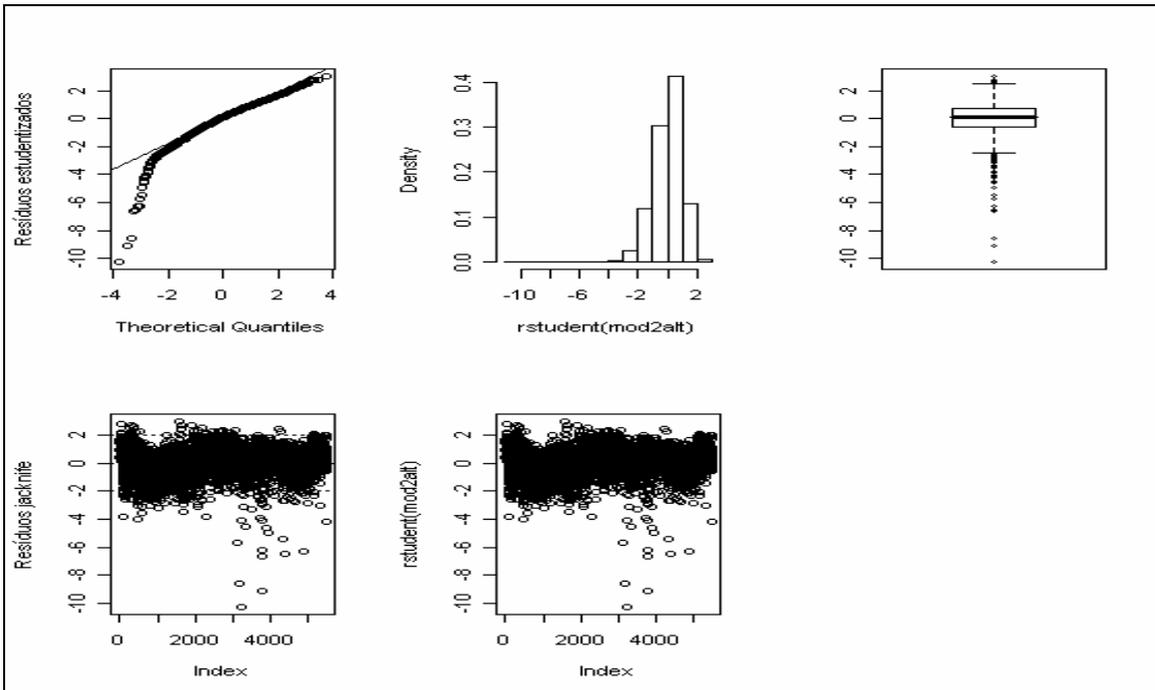


Figura 6: Diagnóstico do modelo de regressão contendo como variável resposta o IDHM (LONGEVIDADE).

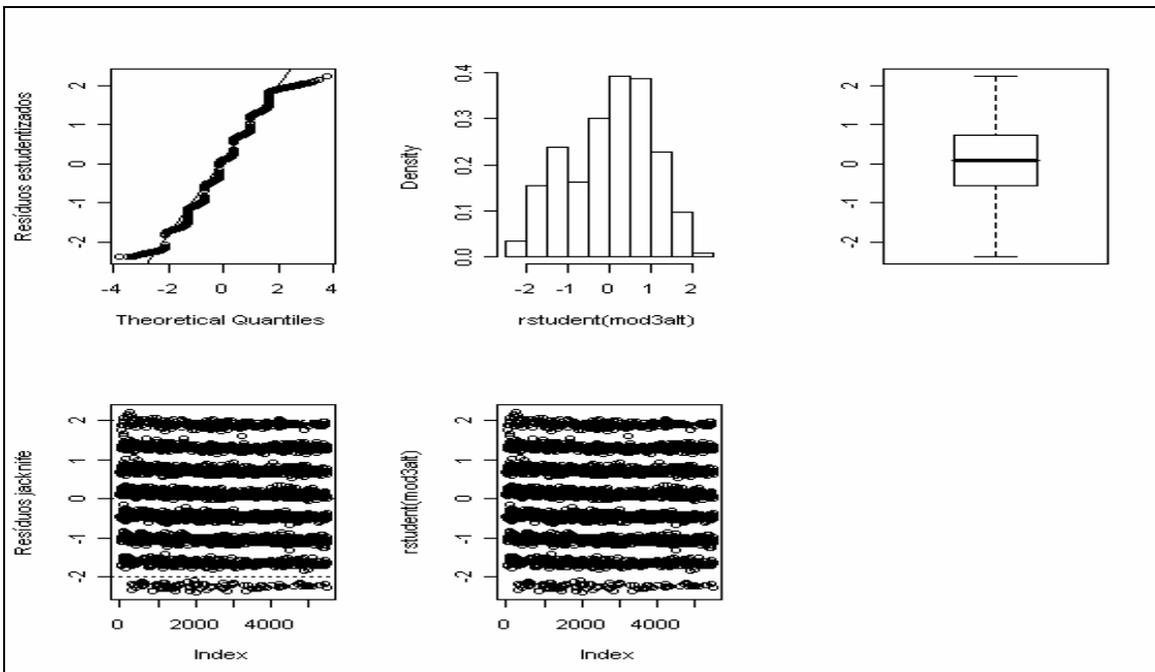
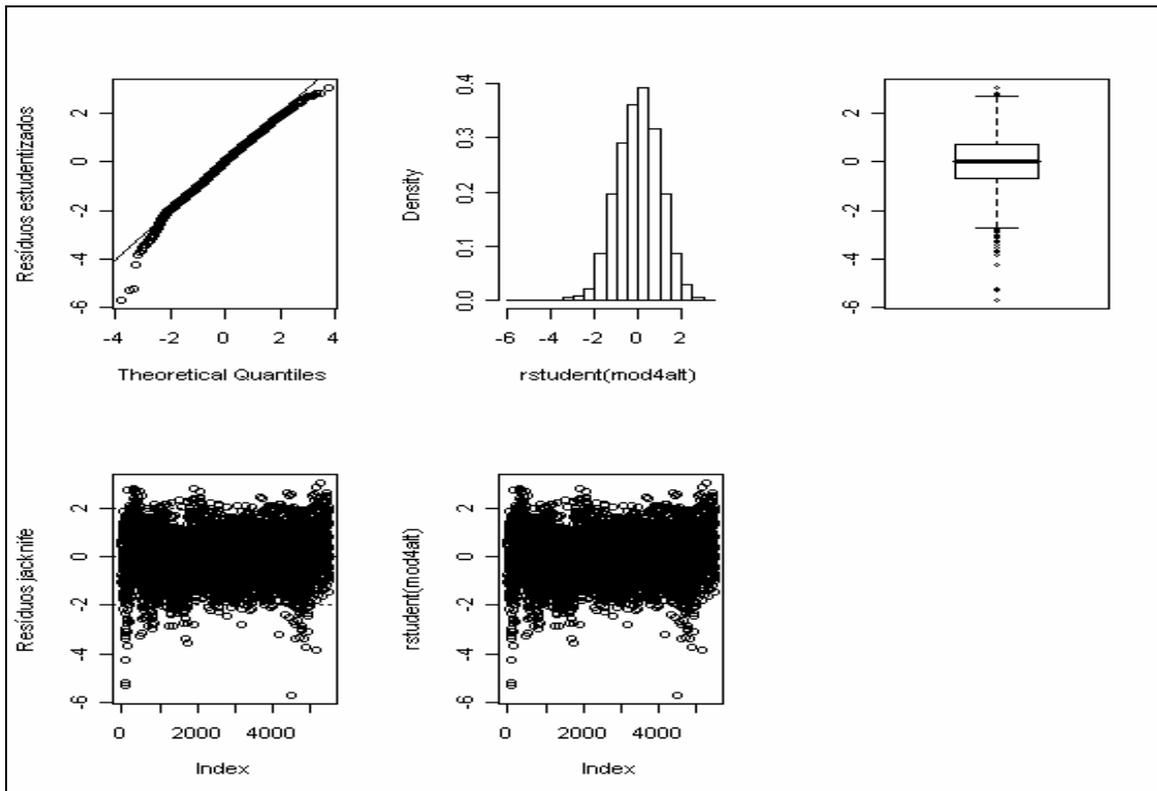


Figura 7: Diagnóstico do modelo de regressão contendo como variável resposta o IDHM (EDUCAÇÃO).



APÊNDICE X – Modelo circular do desenvolvimento social brasileiro

Este apêndice oferece a possibilidade de, com o emprego da função complementar Visualização do KDD, levantar subsídios para conjecturar sobre o modelo conceitual circular comum aos Índices e Sistemas de Indicadores de Desenvolvimento Social através das relações identificadas entre os 15 indicadores de Desenvolvimento Social selecionados entre 46 das mais conhecidas ferramentas nacionais que contemplam a dimensão Social do Desenvolvimento, já estudadas no capítulo 4 desta tese.

Como foi apresentado no Capítulo 4, as relações entre os indicadores podem ser de caráter teórico (aquelas presentes nos trabalhos que discutem os processos sociais envolvidos no fenômeno do desenvolvimento social), como também de natureza prática, tais como os trabalhos que recorrem a pesquisas empíricas para analisar os dados sobre as transformações sociais em certa localidade. A abordagem adotada nesta etapa do trabalho, permitiu selecionar um conjunto de autores que corroboram as relações entre os indicadores selecionados e contribuem com a formulação das hipóteses para subsidiar a construção do primeiro esboço do modelo conceitual entre os indicadores de desenvolvimento social. O Quadro 7.1 apresenta as principais associações, entre os indicadores, identificadas nos Índices e Sistemas de Indicadores de Desenvolvimento analisados neste capítulo.

Como se sabe, essas associações selecionadas para sustentação das relações entre conceitos envolvidos no fenômeno social apresentam-se de modo simultâneo em realidades empíricas de Desenvolvimento Social. Para auxiliar a compreensão da multidimensionalidade e da complexidade que marcam esse fenômeno, as principais correlações ora descritas encontram-se ilustradas na Figura 7.1. Tal diagrama provê uma visão da forma como se articulam os fatores.

A discussão conceitual sobre Sistemas de Indicadores e o referencial teórico sobre os conceitos ou fatores associados ao Desenvolvimento Social fornece subsídios suficientes para que se possa construir um modelo conceitual ou uma representação gráfica que evidencie a forma como se articulam aqueles fatores no conjunto de indicadores estudados. Tal representação consiste em uma elaboração hipotética que reflete as proposições encontradas nos documentos analisados.

Quadro 7.1 – Matriz das principais referências bibliográficas selecionadas para identificar as relações hipotéticas entre os principais indicadores de desenvolvimento social

Relação	Autores
9-Desnutrição → 8-Mortalidade Infantil	França et al (2001); Hartz (1996); Gomes; Santo (1997); Szwarcwald; Castilho (1995)
10-Doenças relacionadas saneamento → 8-Mortalidade Infantil	França et al (2001); Gomes; Santo (1997); Szwarcwald; Castilho (1995)
15-Mortalidade acidentes de transporte → 7-Esperança de vida ao nascer	Camargo (2002)
10-Concentração de Renda → 14-Homicídios	Teixeira (2004)
4-Desocupação → 14-Homicídios	Teixeira (2004)
15-Mortalidade acidentes de transporte → 7-Esperança de vida ao nascer	Aidar (2002); Maia (2000); Nogueira (2004); Teixeira (2004)
3-Desigualdade Distribuição de Renda → 14-Homicídio	Gawryszewskil; Costa (2005); Barata; Ribeiro (2000); Lima et al (2002)
2-Fecundidade → 1-Crescimento da População	Guillard (1855); Gould; Kolb (1964)
13-Escolaridade → 6-Rendimento Médio	Segnini (2000); Reis e Schwartzman (2003: 17)
13-Escolaridade → 4-Desocupação	Severino (2005); Segnini (2000); Reis e Schwartzman (2003: 17); Sen (1999; 90; 2003)
4-Desocupação → 6-Rendimento Médio	Castel (1998); Ferreira (2000); Guimarães (2002b: 104-105); Pochmann(2001: 111)

Fonte: Elaboração do autor

Nesta etapa do estudo, para observar a presença de relações circulares e a estrutura dessas relações por meio da aplicação do método KDD, foi empregada inicialmente a função complementar Visualização. A Visualização, como foi visto, é um poderoso recurso de análise de dados, sendo muitas vezes suficiente para obter as respostas necessárias.

A técnica empregada nesta seção consiste de representar graficamente (diagrama de caminho), o modelo empírico circular e o sistema de relações entre os indicadores de Desenvolvimento Social mais relevantes, já identificados no Capítulo 4 desta tese.

Como resultado é selecionado o modelo que parece refletir as discussões teóricas sobre Desenvolvimento Social encontrada nos 46 Índices e Sistemas de Indicadores estudados. A utilização do KDD permite que sejam analisados os efeitos diretos, indiretos e totais que cada um dos aspectos incluídos no modelo exerce sobre a representação do Desenvolvimento Social.

Diante da multiplicidade de relações simultâneas que podem ser observadas entre os indicadores e da complexidade da estrutura resultante, optou-se por apresentar todas as relações de causalidade observadas⁹⁵ no conjunto de indicadores selecionados. Como se observa no Quadro 7.1 abaixo, todos os indicadores selecionados apresentam ao menos um tipo de relação causal com os demais indicadores.

⁹⁵ A letra “X” foi utilizada para indicar a presença da relação de causalidade observada entre os indicadores apresentados.

Diante do desafio de representar o Desenvolvimento Social por um diagrama de caminho do tipo circular, seria natural pensar em empregar a técnica de Modelagem de Equações Estruturais (MEE). Entretanto, para representar o modelo empírico circular por meio da MEE, é necessário elaborar previamente um sistema de proposições teóricas inter-relacionadas, referentes aos construtos associados ao termo teórico, dos quais podem ser deduzidas hipóteses testáveis. Como foi possível constatar no Capítulo 4, falta de reflexões teóricas explícitas em todos os documentos de divulgação encontrados, o que dificulta a compreensão dos construtos e conceitos que os Índices e Sistemas de Indicadores se propõem a medir.

Quadro 7.1 – Estrutura de relações e indicadores identificados nos 46 índices e sistemas de indicadores nacionais – Brasil, 2009

INDICADOR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 - Taxa de crescimento da população		X			X	X					X	X	X		
2 - Taxa de fecundidade	X				X	X						X	X		
3 - Índice de Gini da distribuição do rendimento				X	X	X								X	
4 - Taxa de desocupação			X		X	X		X			X	X	X	X	
5 - Rendimento familiar per capita	X	X		X		X	X	X	X	X	X	X	X	X	X
6 - Rendimento médio mensal	X	X		X	X		X	X	X	X	X	X	X	X	X
7 - Esperança de vida ao nascer					X	X		X	X					X	X
8 - Taxa de mortalidade infantil					X	X	X		X	X	X	X	X		
9 - Prevalência de desnutrição total				X	X	X	X	X			X	X	X		
10 - Doenças relacionadas ao saneamento ambiental inadequado					X	X		X			X	X	X		
11 - Taxa de escolarização	X			X	X	X		X	X	X		X	X	X	X
12 - Taxa de alfabetização	X	X		X	X	X		X	X	X	X		X	X	X
13 - Escolaridade	X	X		X	X	X		X	X	X	X	X		X	X
14 - Coeficiente de mortalidade por homicídios			X	X	X	X	X				X	X	X		
15 - Coeficiente de mortalidade por acidentes de transporte					X	X	X				X	X	X		

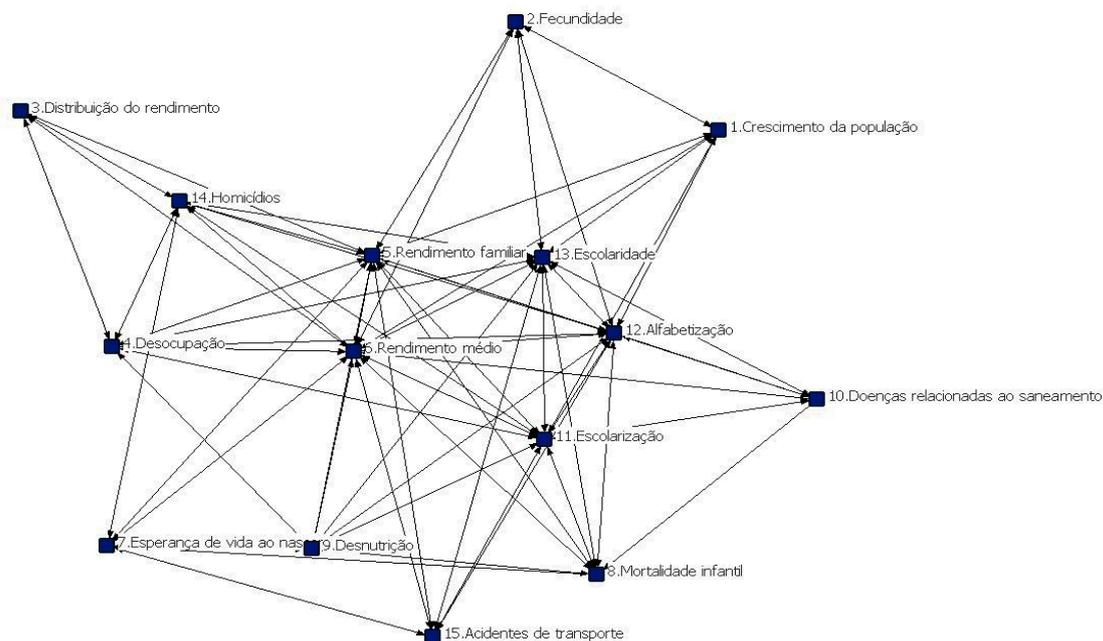
Fonte: Elaboração própria

A Modelagem de Equações Estruturais também baseia-se nas suposições de que suas variáveis devem ser contínuas e normalmente distribuídas. Os 15 indicadores selecionados para representar o termo teórico Desenvolvimento Social, armazenados na jazida de dados, foram submetidos a transformações⁹⁶ e, em nenhuma das transformações realizadas foi possível aceitar o pressuposto de normalidade multivariada da distribuição conjunta dos indicadores/variáveis envolvidos no estudo. Portanto, não foi possível recorrer a Modelagem de Equações Estruturais para elaborar o modelo empírico circular desejado.

A partir dos conceitos levantados com o estudo das principais propostas disponíveis na literatura foi construído um novo modelo conceitual das relações entre os indicadores (Figuras 7.2, 7.3 e 7.4).

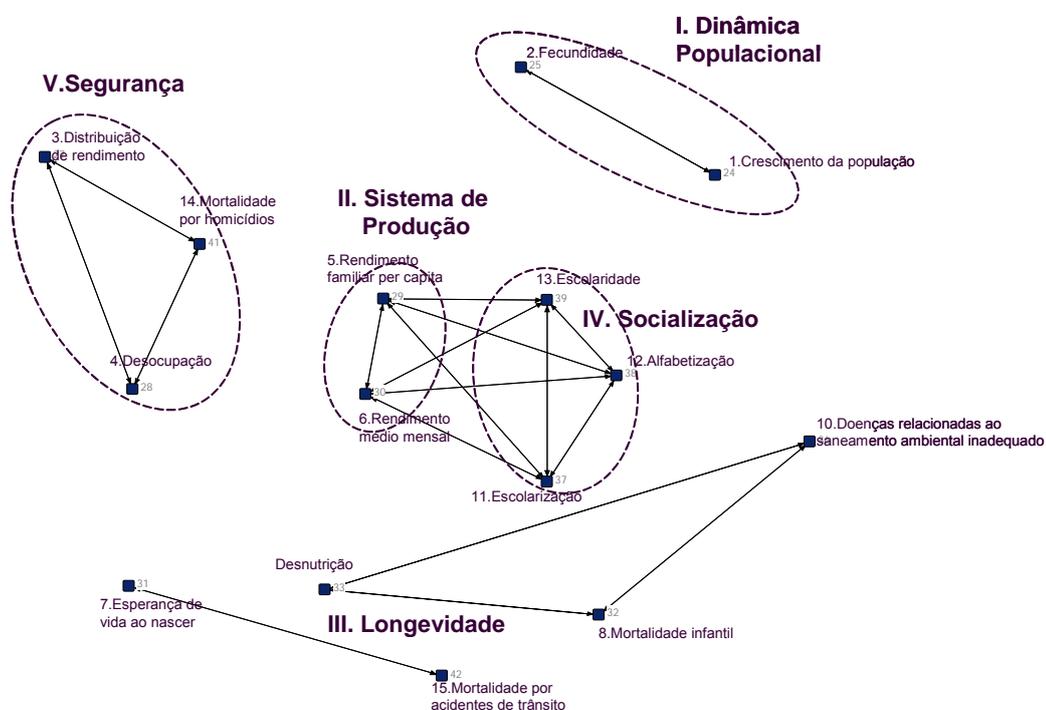
⁹⁶ Os testes de normalidade para os dados transformados $X\lambda$, em que λ é obtido pela transformação de Box&Cox estão disponíveis no Apêndice desta tese.

FIGURA 7.2 – Modelo Conceitual Circular **simplificado** das principais relações entre os indicadores de Desenvolvimento Social



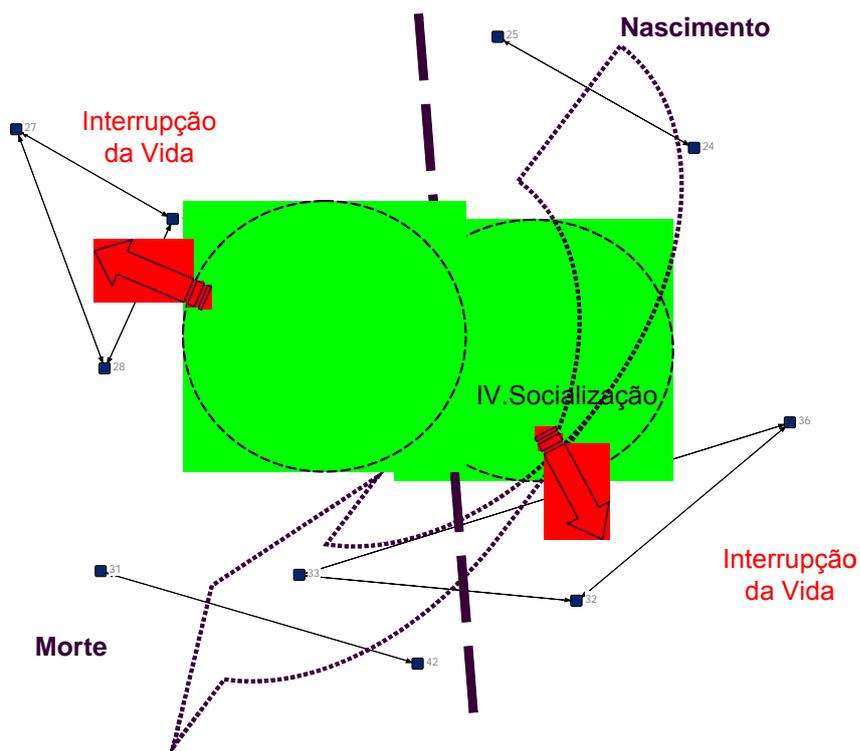
Fonte: Elaboração própria

Figura 7.3 – Modelo conceitual circular **simplificado** das principais relações entre os Construtos de Desenvolvimento Social



Fonte: Elaboração própria

Figura 7.4 – Modelo conceitual **simplificado** das principais relações entre os indicadores, conceitos e construtos do Desenvolvimento Social



Fonte: Elaboração própria

As Figuras 7.2, 7.3 e 7.4 fornecem uma representação pictórica simplificada das relações resultantes das interligações entre os indicadores selecionados para este estudo, obtida por meio do emprego função complementar Visualização do KDD. O modelo empírico circular apresentado permite a avaliação da articulação entre os fatores associados, o que propicia uma melhor compreensão do caráter relacional e complexo, composto por diversos fatores do Desenvolvimento Social que se inter-relacionam e se reforçam mutuamente.

A construção abstrata representada na figura 7.4, elaborada a partir das definições e convenções terminológicas sobre o conceito de Desenvolvimento Social ao longo desta tese, pode representar a realidade de uma maneira muito aproximada através do diagrama de caminho utilizado para representar o modelo empírico circular do Desenvolvimento Social.

Através da análise da estrutura interna do corpo de relações que constituem o fenômeno Desenvolvimento Social, este Capítulo oferece importantes possibilidades para as teorias da sociologia e de outras áreas do conhecimento.

É inquestionável a relevância dessa forma de representação para os estudos sobre a mensuração do Desenvolvimento Social na atualidade. A necessidade de que a realidade social seja apreendida e representada de maneira apropriada coloca-se imperativa para que se possam desenvolver ações efetivas voltadas à promoção do bem-estar humano e ambiental com potencial de sucesso.