

ESTE EXEMPLAR CORRESPONDE A REDAÇÃO FINAL DA
TESE DEFENDIDA POR Pedro Gerber
Machado..... E APROVADA
PELA COMISSÃO JULGADORA EM 30 / 07 / 2012
Walter
ORIENTADOR

**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
COMISSÃO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
PLANEJAMENTO DE SISTEMAS ENERGÉTICOS**

Pedro Gerber Machado

**Análise de Indicadores Socioeconômicos em
Municípios Canavieiros: estudos de caso em
São Paulo e Alagoas**

Pedro Gerber Machado

Análise de Indicadores Socioeconômicos em Municípios Canavieiros: estudos de caso em São Paulo e Alagoas

Dissertação apresentada ao Curso de Mestrado da Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas, como requisito para a obtenção do título de Mestre em Planejamento de Sistemas Energéticos.

Orientador: Arnaldo Cesar da Silva Walter

Campinas
2012

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Machado, Pedro Gerber

M18a Análise de indicadores socioeconômicos em municípios
canavieiros: estudos de caso em São Paulo e Alagoas / Pedro
Gerber Machado. --Campinas, SP: [s.n.], 2012.

Orientador: Arnaldo Cesar da Silva Walter.

Dissertação de Mestrado - Universidade Estadual de
Campinas, Faculdade de Engenharia Mecânica.

1. Indicadores econômicos. 2. Indicadores sociais. 3.
Etanol. 4. Cana-de-açúcar. I. Walter, Arnaldo Cesar da Silva,
1957-. II. Universidade Estadual de Campinas. Faculdade de
Engenharia Mecânica. III. Título.

Título em Inglês: Socioeconomic indicators analysis in sugarcane municipalities: case
studies in São Paulo and Alagoas states

Palavras-chave em Inglês: Economic indicators, Social Indicators, Ethanol, Sugarcane

Área de concentração: Planejamento de Sistemas Energéticos

Titulação: Mestre em Planejamento de Sistemas Energéticos

Banca examinadora: Suani Teixeira Coelho, Alexandre Gori Maia

Data da defesa: 30-07-2012

Programa de Pós Graduação: Engenharia Mecânica

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
COMISSÃO DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
PLANEJAMENTO DE SISTEMAS ENERGÉTICOS

DISSERTAÇÃO DE MESTRADO ACADÊMICO

**Análise de Indicadores Socioeconômicos em
Municípios Canavieiros: estudos de caso em
São Paulo e Alagoas**

Autor: Pedro Gerber Machado
Orientador: Arnaldo Cesar da Silva Walter

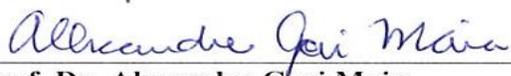
A banca examinadora composta pelos membros abaixo aprovou esta Dissertação:



Prof. Dr. Arnaldo Cesar da Silva Walter, Presidente
FEM - UNICAMP



Prof. Dra. Suani Teixeira Coelho
IEE - USP



Prof. Dr. Alexandre Gori Maia
IE - UNICAMP

Campinas, 30 de Julho de 2012

Agradecimentos

Ao Professor Arnaldo, que me orientou e acreditou em mim.

À minha mãe, Cristina, quem me deu a vida e compartilha comigo as mesmas ideias de um futuro.

Ao CNPQ.

À minha irmã, Cassia, por toda a força e incentivo, mesmo à distância.

Ao meu avô Lupércio e à minha avó Dulce (*In memoriam*). Por sempre me incentivarem e me apoiarem.

À minha tia Elisabete e ao meu tio John. Minha segunda família.

À UNICAMP, à Faculdade de Engenharia Mecânica e à sua seção de pós-graduação.

À Márcia e à Lilian do NIPE, pela competência e paciência.

Ao CTBE, pela infraestrutura.

Ao amigo Marcelo Cunha, pelas contribuições na etapa de qualificação.

Aos amigos e colegas do Programa de Sustentabilidade do CTBE, pelo apoio e contribuição para o meu crescimento.

À minha amiga Camila, com quem eu dividi grandes momentos.

Aos meus amigos Kauê, Yasmin, Leonardo, Elis, Mariana, Tiago, Fellipe, Marcos, Caio, Marília, André e Adriano, por me proporcionarem momentos mágicos, que eu nunca vou esquecer.

Resumo

MACHADO, Pedro Gerber, Análise de Indicadores Socioeconômicos em Municípios Canavieiros: estudos de caso em São Paulo e Alagoas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2012. 131 p. Dissertação (Mestrado).

Esta dissertação de mestrado tem como objetivo analisar os aspectos socioeconômicos associados à produção de cana-de-açúcar, em dois dos principais estados produtores: São Paulo, maior produtor da região Centro-Sul, e do país, e Alagoas, maior produtor da região Nordeste. Para tanto, foram utilizados indicadores municipais regularmente compilados e divulgados pelos órgãos oficiais do governo brasileiro. Foram utilizados alguns métodos estatísticos, como o teste de hipótese t na comparação dos grupos de municípios canavieiros e não canavieiros, a Análise de Cluster para separar, de forma não tendenciosa, os municípios canavieiros em grupos de municípios com melhores e piores indicadores sociais, a Análise Discriminante, com o intuito de identificar quais fatores explicam o maior ou menor estágio de desenvolvimento dos municípios canavieiros, e Gráficos Radar, para facilitar a análise da evolução dos indicadores sociais ao longo do tempo. Se não é possível afirmar categoricamente que os municípios canavieiros de São Paulo e Alagoas têm, na média, melhores indicadores socioeconômicos do que municípios similares não canavieiros, é possível concluir o oposto, ou seja, que os municípios canavieiros não têm piores indicadores do que os não canavieiros. O estudo da evolução dos indicadores de 1970 a 2000 permite concluir que há tendência de equalização dos municípios considerados canavieiros e não canavieiros no que diz respeito aos indicadores considerados, e nos dois estados estudados. Em São Paulo (e Alagoas de forma menos expressiva), de uma forma geral, os municípios que têm melhores indicadores têm a atividade econômica mais diversificada, ou seja, quanto maior a participação relativa de outros setores na economia (indústria e serviços), além da agropecuária, melhores os resultados. Especificamente quanto às variáveis que indicam a importância relativa da atividade canavieira, conclui-se que a presença da cana-de-açúcar não impacta negativamente os indicadores de qualidade de vida do município. Entretanto, a maior dependência da economia local da atividade canavieira está associada, em geral, à menor qualidade de vida, ou seja, são piores os indicadores. Também como regra geral, os municípios canavieiros com maior receita têm melhores indicadores. Da mesma forma, municípios que proporcionalmente recebem mais transferências estaduais e federais têm melhores indicadores.

Palavras-Chave: Indicadores socioeconômicos. Etanol. Cana-de-açúcar.

Abstract

MACHADO, Pedro Gerber, Análise de Indicadores Socioeconômicos em Municípios Canavieiros: estudos de caso em São Paulo e Alagoas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2012. 131 p. Dissertação (Mestrado).

This dissertation aims to analyze the socioeconomic aspects associated with the production of sugarcane in two major producing states: São Paulo, the largest producer in the Center-South region, and in country, and Alagoas, the largest producer in the Northeast region. For this purpose, municipal indicators regularly compiled and published by official agencies of the Brazilian government were used. Some statistical methods were applied, such as the hypothesis *t* test to compare the groups of municipalities that produce sugarcane in large-scale and those that do not produce sugarcane, the Cluster Analysis to separate in groups, in a non-biased way, the producer sugarcane municipalities according to their social indicators, Discriminant Analysis in order to identify which factors explain the higher or lower stage of development of municipalities with large sugarcane production, and diagrams for an easy analysis of trends in social indicators over time. To conclude about the first aim of this dissertation, if it is not possible to state categorically that the municipalities with large sugarcane production in the state of São Paulo and Alagoas have, on average, better socioeconomic indicators than similar municipalities without sugarcane, it is possible to conclude the opposite, namely that the municipalities with sugarcane do not have worse indicators than those without sugarcane. The study of the evolution of the indicators from 1970 to 2000 shows that there is a tendency for equalization of municipalities considered as sugarcane producers and non-sugarcane producers in relation to the indicators considered, and in the two states studied. In São Paulo (and Alagoas to some extent), in general, municipalities that have the best indicators are more diverse economically, i.e. the higher the relative share of other sectors in the economy (industry and services), in addition to farming, the better the results. Specifically concerning the variables that indicate the relative importance of sugarcane production, it was concluded that the presence of sugarcane does not negatively impact the indicators of quality of life in the municipality. However, the higher dependence of the local economy on sugarcane production is associated generally with lower quality of life, i.e. the indicators are worse. Also as a general rule, among the municipalities with large sugarcane production, those with higher taxes revenues have better indicators. Likewise, municipalities that receive proportionately more state and federal transfers have better indicators.

Key words: Socioeconomic indicators. Ethanol. Sugarcane.

Lista de Ilustrações

Figura 2.1 – Distribuição Normal e alguns resultados importantes.	24
Figura 2.2 – Distribuição Normal, distribuição com <i>skewness</i> e distribuição com <i>kurtosis</i>	29
Figura 2.3 - Transformações da distribuição normal.	30
Figura 2.4 - Distribuições originais e suas transformações mais comuns.	31
Figura 2.5 – Dispersão das relações homoscedásticas e heteroscedásticas.	32
Figura 3.1 – Evolução dos indicadores no estado de São Paulo de 1970-2000.	45
Figura 3.2 - Evolução dos indicadores no estado de Alagoas de 1970-2000.	46
Figura 3.3 - Evolução dos indicadores no estado de São Paulo em novos canaviais de 1980.	48
Figura 3.4 - Evolução dos indicadores no estado de São Paulo em novos canaviais de 1991.	49
Figura 3.5 – Evolução dos indicadores sociais em comparação geográfica localizada no estado de São Paulo – Grupo 1 – 1970 a 2000	51
Figura 3.6 - Evolução dos indicadores sociais em comparação geográfica no estado de São Paulo – Grupo 2 – 1970 a 2000.	52
Figura 3.7 - Evolução dos indicadores sociais em comparação geográfica no estado de São Paulo – Grupo 3 – 1970 a 2000.	53
Figura 4.1 - Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1970.	59
Figura 4.2 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1970.	60
Figura 4.3 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1980.	62
Figura 4.4 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1980.	63
Figura 4.5 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1991.	65
Figura 4.6 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1991.	67
Figura 4.7 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo, no ano de 2000.	70
Figura 4.8 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de São Paulo, em 2000	71
Figura 4.9 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1970.	76
Figura 4.10 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1980	78
Figura 4.11 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1991	81

Figura 4.12 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 2000	83
Figura A1 – (a) Histograma da variável IDH antes da transformação e; (b) depois da transformação (raiz do IDH).	96
Figura B1- Separação dos municípios em três clusters considerados indicadores de qualidade de vida no estado de São Paulo em 1991. Grupo 1, melhores resultados; Grupo 3, piores resultados.....	109
Figura C1 – Dispersões entre as variáveis <i>predictors</i> empregadas para 1970, em São Paulo.....	112
Figura C2– Dispersões entre as variáveis <i>predictors</i> empregadas para 1980, em São Paulo.....	113
Figura C3 – Dispersões entre as variáveis <i>predictors</i> empregadas para 1991, em São Paulo.....	115
Figura C4 – Dispersões entre as variáveis <i>predictors</i> empregadas para 2000, em São Paulo.....	117
Figura C5 – Dispersões entre as variáveis <i>predictors</i> empregadas, em Alagoas, em 1970.	118
Figura C6– Dispersões entre as variáveis <i>predictors</i> , para 1980, em Alagoas.....	120
Figura C7 – Dispersões entre as variáveis <i>predictors</i> empregadas, para 1991, em Alagoas.	122
Figura C8 – Dispersões entre as variáveis <i>predictors</i> empregadas, em 1991, em Alagoas.	124
Figura D1 – Expansão da atividade canavieira nos municípios do estado de São Paulo, de 1970 a 2000.	125
Figura F1 – IDH e participação da cana na economia – São Paulo – 1970.	127
Figura F2 - IDH e participação da cana na economia – São Paulo – 1980.....	128
Figura F3 - IDH e participação da cana na economia – São Paulo – 1991.....	128
Figura F4 - IDH e participação da cana na economia – São Paulo – 2000.....	129
Figura F5 – L de Theil e participação da cana na economia – São Paulo – 1970.....	129
Figura F6 - L de Theil e participação da cana na economia – São Paulo – 1980.	130
Figura F7 - L de Theil e participação da cana na economia – São Paulo – 1991.	130
Figura F8 - L de Theil e participação da cana na economia – São Paulo – 2000.	131
Figura F9 - IDH e participação da cana na economia – Alagoas – 1970.....	131
Figura F10 - IDH e participação da cana na economia – Alagoas – 1980.....	132

Figura F11 - IDH e participação da cana na economia – Alagoas – 1991.	132
Figura F12 - IDH e participação da cana na economia – Alagoas – 2000.	133
Figura F13 – L de Theil e participação da cana na economia – Alagoas – 1970.....	133
Figura F14 – L de Theil e participação da cana na economia – Alagoas – 1980.....	134
Figura F15 – L de Theil e participação da cana na economia – Alagoas – 1991.....	134
Figura F16 – L de Theil e participação da cana na economia – Alagoas – 2000.....	135

Lista de tabelas

Tabela 1.1 – Empregos, rendimento e escolaridade entre 2002 e 2009.....	14
Tabela 1.2 – Benefícios oferecidos em uma amostra de 47 estabelecimentos (2003).....	18
Tabela 1.3 – Indicadores Componentes do IDHM.....	19
Tabela 1.4-Coeficiente de Correlação de Pearson para IDHM e produção de cana	20
Tabela 3.1 – Resultados do teste t para o estado de São Paulo – 1970 a 2000.....	39
Tabela 3.2- Resultados do teste <i>t</i> para indicadores de 1970 no estado de São Paulo	40
Tabela 3.3 – Resultados do teste t para o estado de Alagoas – 1970 a 2000.....	42
Tabela 3.4 - Resultados do teste t para 1970 no estado de Alagoas	43
Tabela 3.5 – Resultados do teste Mann-Whitney para dois indicadores -1970 a 2000	47
Tabela 4.1 – Resultados do Cluster para a divisão em dois grupos, para São Paulo, em 1970	56
Tabela 4.2 – Correlações entre IDH e outros indicadores sociais de 1970-2000	57
Tabela 4.3 - Resultados da análise discriminante com dois grupos para São Paulo em 1970.....	58
Tabela 4.4 – Correlação entre indicadores econômicos e sociais – São Paulo 1970	59
Tabela 4.5 – Resultados do Cluster em 2 grupos para São Paulo em 1980	61
Tabela 4.6- Resultados da análise discriminante com dois grupos para São Paulo em 1980.....	61
Tabela 4.7 – Correlação entre indicadores econômicos e sociais – São Paulo 1980	62
Tabela 4.8 – Resultados da Análise Cluster em dois grupos, para São Paulo, em 1991	64
Tabela 4.9- Resultados da Análise Discriminante para a classificação em dois grupos, para São Paulo, em 1991	64
Tabela 4.10 – Correlação entre indicadores econômicos e sociais – São Paulo 1991.....	66
Tabela 4.11 – Resultados do Cluster em 2 grupos para São Paulo, em 2000.....	67
Tabela 4.12 - Resultados da Análise Discriminante para a classificação em dois grupos, para São Paulo, em 2000.....	68

Tabela 4.13 – Correlação entre indicadores econômicos e sociais – São Paulo 2000.....	69
Tabela 4.14 – Médias das participações econômicas dos setores e da população rural em São Paulo, em municípios canavieiros e vários anos.	72
Tabela 4.15 – Resultados do Cluster em 2 grupos para Alagoas em 1970	74
Tabela 4.16- Resultados da Análise Discriminante com dois grupos para Alagoas em 1970.....	74
Tabela 4.17 – Correlação entre indicadores econômicos e sociais – Alagoas 1970	75
Tabela 4.18 – Resultados do Cluster em dois grupos para Alagoas, em 1980	76
Tabela 4.19 - Resultados da análise discriminante para dois grupos de municípios, em Alagoas, em 1980.	77
Tabela 4.20– Correlação entre indicadores econômicos e sociais – Alagoas 1980	77
Tabela 4.21– Resultados do Cluster em dois grupos para Alagoas, em 1990	79
Tabela 4.22- Resultados da análise discriminante com dois grupos para Alagoas, em 1991.....	79
Tabela 4.23– Correlação entre indicadores econômicos e sociais – Alagoas 1991	80
Tabela 4.24 – Resultados do Cluster em dois grupos para Alagoas, em 2000	82
Tabela 4.25- Resultados da análise discriminante com dois grupos para Alagoas, em 2000.....	82
Tabela 4.26 – Correlação entre indicadores econômicos e sociais – Alagoas 2000	83
Tabela 4.27 – Médias das participações econômicas e da população rural em Alagoas	84
Tabela A1 – Resultado das transformações dos indicadores de bem-estar 1970 no estado de São Paulo...	97
Tabela A2 – Resultado do teste de Levene para 1970, no estado de São Paulo	98
Tabela A3 – Resultado das transformações dos indicadores de 1980, no estado de São Paulo	99
Tabela A4 – Resultados do teste <i>t</i> para 1980 no estado de São Paulo.....	99
Tabela A5 – Transformações dos indicadores de 1991 no estado de São Paulo	100
Tabela A6 – Resultados do teste de Levene para 1991 no estado de São Paulo	100
Tabela A7 – Transformações dos indicadores de 2000 no estado de São Paulo	101
Tabela A8 – Resultados do teste de Levene para 2000 no estado de São Paulo	101

Tabela A9 – Transformações dos indicadores de 1970 no estado de Alagoas	102
Tabela A10 – Resultado dos testes de Levene para 1970 no estado de Alagoas	102
Tabela A11 – Transformações dos indicadores de 1980 no estado de Alagoas	103
Tabela A12 – Resultado do teste de Levene para 1980 no estado de Alagoas.....	103
Tabela A13 – Transformações dos indicadores de 1991 no estado de Alagoas	104
Tabela A14– Resultado do teste de Levene para 1991 no estado de Alagoas.....	104
Tabela A15 – Transformações dos indicadores de 2000 no estado de Alagoas	105
Tabela A16 – Resultado do teste de Levene para 2000 no estado de Alagoas.....	106
Tabela B1 – Resultados da Análise Cluster em três grupos, para São Paulo, em 1991	107
Tabela B2 – Resultados da Análise Cluster com quatro grupos, para São Paulo, em 1991	107
Tabela B3 - Resultados da Análise Discriminante para a classificação em três grupos, para São Paulo, em 1991	108
Tabela B4 – Distâncias Euclidianas para os clusters – divisão em 2, 3 e 4 grupos	108
Tabela C1 – Resultado das transformações feitas para realização da Análise Discriminante – São Paulo 1970	110
Tabela C2 – Resultado do teste de Levene para aplicação da Análise Discriminante – 1970, no estado de São Paulo.....	111
Tabela C3 – Resultado das transformações para realização da Análise Discriminante – São Paulo, 1980	113
Tabela C4 – Resultado das transformações para realização da Análise Discriminante – São Paulo, 1991	114
Tabela C5 – Resultado do teste de Levene para as variáveis transformadas utilizadas na Análise Discriminante – 1991, no estado de São Paulo.....	114
Tabela C6– Resultado das transformações para aplicação da Análise Discriminante – São Paulo, 2000.	116
Tabela C7 – Resultado do teste de Levene para as variáveis aplicadas na Análise Discriminante – 2000, no estado de São Paulo	116

Tabela C8 – Resultado das transformações realizadas para aplicação da Análise Discriminante – Alagoas, 1970	117
Tabela C9 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante – 1970, no estado de Alagoas.....	118
Tabela C10 – Resultado da análise das variáveis empregadas na Análise Discriminante – Alagoas, 1980	119
Tabela C11 – Resultado do teste de Levene para as variáveis aplicadas na Análise Discriminante – 1980, no estado de Alagoas.....	119
Tabela C12 – Resultado das transformações aplicadas às variáveis para realização da Análise Discriminante – Alagoas, 1991.....	121
Tabela C13 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante – 1991, no estado de Alagoas.....	121
Tabela C14 – Resultado das transformações realizadas para as variáveis empregadas na Análise Discriminante – Alagoas, 2000.....	123
Tabela C15 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante -1991, no estado de Alagoas.....	123
Tabela E1 – Teste de hipóteses não paramétrico para análise da evolução dos indicadores – São Paulo de 1970 a 2000	126
Tabela An1– C Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de São Paulo entre 1970-1980 – valores sem transformação de Oliveira (2011)	136
Tabela An 2 - Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de São Paulo entre 1991-2000 – valores sem transformação de Oliveira (2011)	136
Tabela An3– Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de Alagoas entre 1970-1980 – valores sem transformação de Oliveira (2011).....	137
Tabela An 4 - Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de Alagoas entre 1970-1980 – valores sem transformação de Oliveira (2011).....	137

Lista de Abreviaturas e Siglas

CGEE – Centro de Gestão e Estudos Estratégicos

CONAB – Companhia Nacional de Abastecimento

EPI – Equipamento de Proteção Individual

GEE – Gases do Efeito Estufa

IBGE – Instituto Brasileiro de Geografia e Estatística

ICMS – Imposto sobre Circulação de Mercadorias e Serviços

IDH – Índice de Desenvolvimento Humano

IEA – International Energy Agency

INPC – Índice Nacional de Preços ao Consumidor

IPCC – Intergovernmental Panel on Climate Change

IPEA – Instituto de Pesquisa Econômica Aplicada

IPRS – Índice Paulista de Responsabilidade Social

IPVA – Imposto sobre a Propriedade de Veículos Automotores

MAPA – Ministério da Agricultura, Pecuária e Abastecimento

OIT – Organização Internacional do Trabalho

PIB – Produto Interno Bruto

RAIS – Relação Anual de Informações Sociais

UNICA – União da Indústria de Cana-de-açúcar

SUMÁRIO

Agradecimentos	v
Resumo.....	vi
Abstract.....	vii
Lista de Ilustrações.....	viii
Introdução.....	1
1. Revisão da literatura	4
1.1 Indicadores socioeconômicos e de avaliação da sustentabilidade	4
1.2 Revisão bibliográfica.....	13
1.2.1 Emprego e Renda.....	13
1.2.2 Qualidade do Emprego.....	16
1.2.3 Análise de indicadores socioeconômicos em municípios canavieiros	18
2. Revisão metodológica	23
2.1 Estatística Básica	23
2.2 Teste de Hipóteses	24
2.3 Análise Multivariada.....	25
2.4 Triagem dos dados.....	26
2.5 Normalidade	27
2.6 Linearidade	31
2.7 Homoscedasticidade.....	32
2.8 Multicolinearidade e singularidade.....	32
2.9 Análise de Cluster.....	33
2.10 Análise Discriminante	34
3. Resultados.....	37
3.1 Comparação das médias	37
3.1.1 São Paulo.....	38
3.1.2 Alagoas.....	41
3.2 Evolução do conjunto de indicadores.....	44
4. Análise de Cluster e Discriminante	55

4.1	São Paulo.....	55
4.1.1	1970.....	55
4.1.2	1980.....	60
4.1.3	1991.....	63
4.1.4	2000.....	67
4.2	Alagoas.....	73
4.2.1	1970.....	73
4.2.2	1980.....	76
4.2.3	1991.....	78
4.2.4	2000.....	81
5.	Conclusões e Recomendações.....	86
5.1.	Conclusões.....	86
5.2.	Sugestões.....	91
6.	Referências bibliográficas.....	92
	Apêndice A – Transformação dos indicadores socioeconômicos.....	96
	São Paulo.....	97
	1970.....	97
	1980.....	98
	1991.....	99
	2000.....	100
	Alagoas.....	101
	1970.....	101
	1980.....	103
	1991.....	104
	Apêndice B – Procedimento para escolha do número de clusters.....	107
	Apêndice C – Transformação das variáveis <i>predictors</i>	110
	São Paulo.....	110
	1970.....	110
	1980.....	112
	1991.....	114
	2000.....	115

Alagoas.....	117
1970.....	117
1980.....	119
1991.....	120
2000.....	122
Apêndice D – Novos municípios canavieiros no estado de São Paulo	125
Apêndice E – Teste de hipóteses não paramétrico de Mann-Whitney para novos canavieiros.....	126
Apêndice F – Evolução dos indicadores IDH e L de Theil.....	127
São Paulo	127
Alagoas.....	131
Anexo I.....	136

Introdução

Todas as sociedades necessitam de serviços energéticos para suprir as necessidades humanas básicas e para facilitar processos produtivos (IPCC, 2011). Atualmente, 85% de toda a oferta de energia primária vêm de combustíveis fósseis (IEA 2010; IPCC 2011). No entanto, sistemas energéticos predominantemente de origem fóssil não podem ser sustentados em longo prazo por sua distribuição desigual de recursos e sua grande contribuição para as emissões antropogênicas de Gases do Efeito Estufa (GEE). O uso de fontes de energia renováveis deverá ser crescente para substituir os combustíveis fósseis, tornando os sistemas de energia mais sustentáveis (IPCC, 2007). Os biocombustíveis, por sua vez, podem ser uma das soluções para os problemas climáticos resultantes da queima de energia fóssil. Dentro do debate sobre o clima, a segurança de suprimento e a volatilidade de preços dos combustíveis fósseis, os biocombustíveis tornaram-se uma opção promissora para reduzir as emissões de gases de efeito estufa e diversificar a oferta de energia. Como destaca Oliveira (2011), o interesse global na produção e no consumo de biocombustíveis vem aumentando, justificado, principalmente, pela necessidade de se mitigar as emissões de GEE.

No entanto, a crescente produção e utilização de bioenergia não são singularmente motivadas pelo seu potencial de mitigação de GEE (se produzidos sustentavelmente), mas também pela fácil implementação de bioenergia em infraestruturas energéticas existentes, a versatilidade da biomassa como fonte, a diversificação de fontes de energia e a subsequente segurança energética, a potencial contribuição para o desenvolvimento rural e o potencial de recuperação de terras degradadas (VAN DER HILST, 2012).

Políticas governamentais de apoio à produção de etanol da cana-de-açúcar, grande disponibilidade de terras e um clima favorável, ou seja, tropical, com bom regime de chuvas e temperaturas adequadas, tornaram o Brasil um país de liderança mundial nesse campo. Em 2009, o Brasil contribuiu com 33% da produção de etanol mundial e era também o maior exportador (HERRERAS, 2011). A importância do setor para a economia do país tomou grandes proporções, principalmente nas últimas quatro décadas. Em 2008 a participação no PIB brasileiro chegou a 2%, gerando mais de 1.280.000 empregos formais (UNICA, 2010). O número de usinas em 2010 era de 413, sendo 103 só produtoras de etanol, 11 apenas de açúcar e 297 mistas (MAPA, 2010), enquanto em 2011 o número total de usinas era 439. A previsão para 2012 é de produção de mais de 14,2 bilhões de litros de etanol hidratado e 9,7 bilhões de litros de etanol anidro, além de 38,8 milhões toneladas de açúcar (CONAB, 2012).

A dimensão da produção de cana-de-açúcar e de etanol, bem como o crescimento de seu consumo, que é consequência da introdução e do sucesso dos veículos *flex* na frota brasileira, justificam a análise não apenas dos aspectos ambientais, mas também sociais e econômicos da atividade, incluindo sua contribuição para o desenvolvimento do país e das regiões nas quais está presente.

Para Oliveira (2011), dois aspectos se destacam no estudo socioeconômico relacionado à produção de bioenergia: as condições sociais da população local e o acesso a recursos que proporcionem melhor qualidade de vida. Muitos autores (Moraes (2008), Hoffmann (2008), Fernandes (2011), Chagas (2011), etc) vêm analisando as condições de trabalho vigentes no setor, sobretudo aquelas consideradas mais degradantes, associadas ao corte manual da cana-de-açúcar. Outros analisam questões socioeconômicas considerando o rendimento dos trabalhadores, bem como suas características sociais, como raça, vínculo empregatício, horas de trabalho, etc. Finalmente, há estudos que focam o desenvolvimento social e econômico das regiões produtoras, sendo que a maioria dos estudos explora a relação entre produção de cana-de-açúcar e IDH.

Esta dissertação tem como objetivo analisar os aspectos socioeconômicos associados à produção de cana-de-açúcar em dois dos principais estados produtores: São Paulo, maior produtor da região Centro-Sul e do país, e Alagoas, maior produtor da região Nordeste. Para tanto, foram utilizados indicadores municipais regularmente compilados e divulgados pelos órgãos oficiais do governo brasileiro. Também é objetivo desta dissertação aprofundar estatisticamente as análises feitas por Oliveira (2011), reformulando as comparações das médias e os estudos de evolução dos indicadores.

Os objetivos específicos desta dissertação são:

- Comparar as médias dos indicadores socioeconômicos dos municípios canavieiros e não canavieiros, com o objetivo de identificar se há diferenças significativas;
- Analisar a evolução dos indicadores socioeconômicos dos municípios canavieiros e não canavieiros, e também daqueles que se tornaram canavieiros no período de tempo estudado;
- Separar os municípios canavieiros em grupos de melhores e piores desempenhos do ponto de vista dos indicadores socioeconômicos, e analisar quais fatores melhor explicam a existência de municípios melhores e piores, tentando entender por que os municípios estão no grupo de melhores ou piores indicadores, dentro de um conjunto de variáveis disponíveis para análise.

No Capítulo 1 desta dissertação é feita a revisão da literatura que trata das questões socioeconômicas relacionadas à produção de cana-de-açúcar e etanol, como emprego e renda, qualidade de emprego e análise de indicadores regionais. O capítulo também trata dos indicadores de sustentabilidade e socioeconômicos, os quais são utilizados nesta dissertação.

O Capítulo 2 descreve a metodologia utilizada para a realização das análises, e consequente cumprimento dos objetivos acima descritos.

No Capítulo 3 são apresentados os resultados da comparação dos municípios canavieiros e não canavieiros, bem como a análise da evolução dos indicadores. Também é feita uma comparação no âmbito de pequenas regiões geográficas no estado de São Paulo.

O Capítulo 4 apresenta os resultados das Análises de Cluster e Discriminante, utilizadas para analisar os fatores que explicam a existência de municípios canavieiros com melhores e piores indicadores.

No Quadro 1, a seguir, são sintetizados os principais objetivos desta dissertação, os métodos empregados e os casos estudados.

Para a análise estatística necessária, nesta dissertação foi utilizado o software *Statistica* 10 da StatSoft, Inc.

Quadro 1			
Objetivos da dissertação			
Objetivos	Métodos	Casos estudados	Seção
Comparação entre as médias de grupos de municípios.	Teste t	São Paulo e Alagoas, de 1970 a 2000	3.1
Análise da evolução dos indicadores; Comparação entre vizinhos geográficos.	Gráfico Radar	São Paulo e Alagoas, de 1970 a 2000. Somente São Paulo, no caso dos novos canavieiros	3.2
Separar municípios canavieiros em grupos de melhor e pior indicadores; Identificar variáveis responsáveis pela separação.	Análise de Cluster e Discriminante	São Paulo e Alagoas, de 1970 a 2000	Capítulo 4

1. Revisão da literatura

1.1 Indicadores socioeconômicos e de avaliação da sustentabilidade

Indicadores são empregados para informar o desempenho e a evolução de um sistema em direção a uma determinada meta; também são empregados para expressar tendências que não são facilmente perceptíveis (BELLEN, 2005). Em outras palavras, um indicador tem o propósito de identificar o funcionamento de um sistema e existem tantos indicadores quantos sistemas a serem monitorados. Suas principais funções são:

- Avaliar tendências e condições;
- Viabilizar a comparação entre lugares e situações;
- Avaliar resultados em relação às metas e aos objetivos;
- Prover informações de advertência; e
- Antecipar futuras condições e tendências.

Para cumprir com suas funções um indicador deve ser relevante, refletindo as características do sistema que o pesquisador deseja investigar. O indicador deve também ser de fácil compreensão, ser confiável e deve fornecer informações enquanto existe a possibilidade de agir sobre o sistema, modificando alguma tendência (HART, 1999). Outras qualidades de um indicador são apresentadas por Meadows (1998). Para a autora, um indicador deve ser democrático, ou seja, ter a participação da população na sua escolha, ter grandezas físicas e unidades facilmente entendidas. Se um indicador de estado do sistema é escolhido de modo errado, é quantificado imprecisamente, indica resultados defasados ou com viés, as decisões baseadas nesse indicador não poderão ser eficientes. Indicadores errôneos não permitem que o sistema seja levado ao estado esperado (MEADOWS, 1998).

Bellen (2005) considera um indicador uma representação de um atributo de um sistema. Para o autor, um indicador é uma representação, imagem ou abstração do atributo. Qualquer variável, e qualquer indicador, têm uma significância própria e sua característica mais importante é sua relevância para a definição de políticas e para o processo de tomada de decisão. Para ser representativo, o indicador tem de ser considerado importante para os tomadores de decisão e também pelo público.

Ainda segundo o autor, o indicador agrega e quantifica informações de modo que sua significância fique mais aparente. Indicadores podem ser quantitativos ou qualitativos e, em função das limitações que existem em relação a indicadores simplesmente numéricos, os indicadores qualitativos são mais adequados para a avaliação do desenvolvimento sustentável.

Indicadores de desenvolvimento sustentável, no entanto, devem ir além. Indicadores de sustentabilidade devem ter adicionados ao seu conceito a variável tempo, através de períodos, limites ou objetivos. A pergunta central torna-se, então, “quanto que certa atividade pode durar?” Se não relacionados ao tempo, os indicadores de sustentabilidade devem estar relacionados à

capacidade de sustentação, ou a valores *thresholds* (limiares, em português) que indicam limites críticos (MEADOWS, 1998).

Para bem analisar a sustentabilidade, Bellen (2005) diz que os indicadores dos sistemas devem ser inter-relacionados, sempre consideradas as restrições de disponibilidade e do custo de obtenção desses indicadores.

Hart (1999) vai além para conceituar indicadores de sustentabilidade comunitária, que devem mostrar as interconectividades entre mudanças na economia, o meio ambiente e a sociedade. O conceito de sustentabilidade comunitária, ou de cidades sustentáveis, está relacionado principalmente com questões espaciais e ambientais e, com menor importância, com relações sociais. O debate sobre a sustentabilidade das cidades, segundo Colantonio (2007), é se, com suas características específicas, as cidades podem se desenvolver sustentavelmente, já que, por um lado, a industrialização que criou as cidades gerou grandes impactos ambientais negativos; porém, por outro lado, as cidades fornecem serviços de saúde, saneamento e outras infraestruturas importantes para satisfazer as necessidades básicas da população.

A partir de Hart (1999) podem ser formuladas algumas perguntas para identificar a qualidade de um indicador para avaliação de comunidades sustentáveis. Nas esferas social e econômica, sete perguntas podem ser usadas para selecionar um indicador. São elas:

- O indicador trata da capacidade de carga¹ do capital humano da comunidade, como habilidades naturais das pessoas, educação e saúde?
- O indicador trata da capacidade de carga do capital social, como as conexões entre as pessoas na comunidade, as relações familiares, entre amigos, vizinho, grupos sociais, empreendimentos, governos e suas possibilidades em cooperarem e trabalharem juntos?
- O indicador trata da capacidade de carga do capital físico da comunidade e a capacidade de manter e aumentar a infraestrutura da cidade com fundos existentes?
- A comunidade consegue entender o indicador?
- O indicador fornece uma visão de longo prazo da comunidade?
- O indicador mede alguma conexão entre sociedade e economia?
- O indicador foca a sustentabilidade local em detrimento da sustentabilidade global?

A autora considera a última pergunta como automaticamente excludente, ou seja, caso a resposta seja “sim”, o indicador deve ser desqualificado para compor a representação da sustentabilidade.

Indo além do indicador individual, Maggino (2009) analisou os requisitos de um bom sistema de indicadores. Para a autora, um conjunto de indicadores deve ser avaliado de acordo com sua função, que pode ser de monitoramento, quando o objetivo é identificar e definir problemas existentes, controlar e identificar aspectos críticos do sistema, medir mudanças no tempo e

¹ Capacidade de carga é o tamanho da população que pode ser suportada indefinidamente pelos recursos disponíveis e pelos serviços de apoio ao capital natural, social, humano e construído.

melhorar o desempenho do sistema. O objetivo também pode ser relatar condições e a dinâmica de certa realidade, ou analisar as relações existentes entre diferentes componentes dessa realidade; Finalmente, pode ter a finalidade de avaliação, como em procedimentos de certificação.

Neste contexto, Bellen (2005) identificou três metodologias de avaliação de sustentabilidade, que foram consideradas em sua pesquisa como as mais relevantes e reconhecidas internacionalmente. O autor descreveu cada uma delas e as comparou a partir de categorias analíticas, tais como escopo, esfera, dados de participação e interface, levando em conta as características relevantes de um indicador de sustentabilidade.

Das três metodologias apresentadas em Bellen (2005), duas delas, o *Dashboard of Sustainability* e o *Barometer of Sustainability*, são sistemas de indicadores criados para a tomada de decisões, tanto em ambientes públicos como privados, e são capazes de representar estratégias de desenvolvimento.

O *Dashboard of Sustainability* (Painel de sustentabilidade, em português) representa cada dimensão - social, institucional, econômica e ambiental - através de índices que englobam os vários indicadores das dimensões individuais. Outro índice global de sustentabilidade resulta de uma função dos índices de cada dimensão. Nas dimensões social e econômica, o sistema é baseado nos indicadores apresentados no Quadro 2. Esses indicadores formam os índices, que são comparados a valores estabelecidos pela metodologia. A representação gráfica, como em um painel de carro, ajuda o entendimento de desempenho em cada dimensão.

Quadro 2 Indicadores do <i>Dashboard of Sustainability</i>	
Dimensão social	<ul style="list-style-type: none"> • Índice de pobreza • Igualdade de gênero • Padrão nutricional • Saúde • Mortalidade • Condições Sanitárias • Água Potável • Nível Educacional • Alfabetização • Moradia • Violência • População
Dimensão econômica	<ul style="list-style-type: none"> • Performance econômica • Comércio • Estado financeiro • Consumo de materiais • Consumo de energia • Geração e gestão de lixo • Transporte

Fonte: Adaptado de Van Bellen (2005)

O *Barometer of Sustainability* (Barômetro de sustentabilidade) mensura o bem-estar dos seres humanos e do ecossistema, sendo apresentado através de quatro índices. O Índice de Bem-estar Humano (HWI, em inglês) e o Índice de Bem-estar do Ecossistema (EWI, em inglês) são medidas abrangentes de qualidade de vida e do meio ambiente. O Índice de Bem-estar (WI) os justapõe para serem comparados. Outro índice foi criado para identificar quanto uma nação pode ganhar em Bem-estar por estar pressionando o meio ambiente: o Índice de Bem-estar/Estresse (WSI) (PRESCOTT-ALLEN, 2001). O Barômetro também possui representação gráfica, facilitando a compreensão de seus índices. Diferente do *Dashboard*, o *Barometer* é composto de apenas duas dimensões, a social e a ecológica. A dimensão social divide-se em outras cinco subdimensões: saúde e população, riqueza, conhecimento e cultura, comunidade e equidade. Essas subdimensões são representadas por vários indicadores, apresentados no Quadro 3.

Quadro 3 Indicadores do <i>Barometer of Sustainability</i>	
Saúde e população	Expectativa de vida no nascimento; expectativa de vida saudável ao nascer; taxa de mortalidade infantil; taxa total de fertilidade; população com alimentação insuficiente; prevalência de nanismo em crianças abaixo dos cinco anos de idade; prevalência de sub peso em crianças abaixo dos cinco anos de idade; prevalência de bebês com baixo peso ao nascer; população com distribuição de água e atendida pelo saneamento básico.
Riqueza	Produto Interno Bruto (PIB); PIB per capita; taxa de inflação anual; taxa de desemprego anual; valor presente do serviço da dívida como uma porcentagem das exportações de bens e serviços; valor presente do serviço da dívida como uma porcentagem do PIB; razão da dívida de curto prazo pelas reservas internacionais; dívidas públicas brutas como porcentagem do PIB; déficit governamental anual como porcentagem do PIB.
Conhecimento e cultura	Matrículas no ensino básico; matrículas no ensino médio; matrículas na universidade; número de linhas telefônicas e celulares por 100 pessoas; falhas por 100 linhas de telefone; usuários de internet.
Comunidade	Avaliação dos direitos políticos; avaliação da liberdade civil; avaliação da liberdade de imprensa; índice de percepção de corrupção; mortes por conflitos armados por ano; dispêndios militares como porcentagem do PIB; homicídios, estupros e assaltos.
Equidade	Razão entre a renda dos 20% mais ricos e dos 20% mais pobres; razão entre a renda média masculina e feminina; diferença média de formação entre homens e mulheres; participação feminina no parlamento.

Fonte: Adaptado de Prescott-Allen (2001)

Apesar do esforço na criação de metodologias para a avaliação da sustentabilidade, seja ela em ambientes públicos, privados ou em comunidades, a escolha dos indicadores depende principalmente da disponibilidade de dados, especialmente quando se trata de um estudo temporal e espacialmente amplo. Seguindo Oliveira (2011), a escolha dos indicadores para este estudo teve intenção de abranger várias subdivisões da dimensão social, e também econômica, levando em conta a existência dos dados nos anos e municípios estudados.

Quanto aos indicadores sociais, Jannuzzi (2001) os classifica como sendo indicadores de saúde, indicadores educacionais, de mercado de trabalho, demográficos, habitacionais, segurança pública, justiça, infraestrutura e renda e desigualdade. Os indicadores escolhidos nesta dissertação seguem esta classificação, limitada à disponibilidade dos dados.

A coleta de dados para este estudo foi realizada priorizando as bases de dados oficiais do governo brasileiro e de instituições estaduais, de acesso público. Prioridade foi dada às bases de dados eletrônicas do IPEADATA, do Instituto de Pesquisa Econômica Aplicada (IPEA). É uma base de dados que utiliza como fonte de dados os microdados provenientes de várias pesquisas do IBGE (por exemplo, Pesquisa Nacional por Amostra de Domicílios, Censos Demográficos) e dos vários ministérios (por exemplo, Censo Escolar, Mortalidade, Rais/Caged, entre outras), para diferentes anos (IPEADATA, 2012). São disponibilizadas três categorias de dados: macroeconômico, regional e social. Dada a natureza da pesquisa, a categoria social foi priorizada, e foram selecionados seis indicadores e dois índices para os anos de 1970, 1980, 1991 e 2000. Infelizmente, os dados de 2010 não estavam ainda disponibilizados até meados de 2012. De Oliveira (2011), as categorias e indicadores selecionados são apresentados no Quadro 4.

Quadro 4 Indicadores socioeconômicos selecionados para esta dissertação	
Categoria/Indicador	Definição
Educação	
Analfabetos - pessoas com 15 anos e mais - (%)	Percentual de pessoas de 15 e mais anos de idade que não sabem ler nem escrever um bilhete simples. Chamado de “Analfabetos”.
Saúde	
Esperança de vida ao nascer – (Anos)	Expectativa de anos de vida de uma pessoa nascida no ano de referência, supondo que as taxas de mortalidade por idade, estimadas para anos anteriores, se mantivessem constantes nos anos posteriores. Chamado de “Esperança de vida”.
Mortalidade infantil (por mil nascidos vivos)	Número de pessoas de cada mil nascidas vivas no ano de referência que não deverão completar um ano de vida. Chamado de “Mortalidade infantil”.
Equidade	
Índice L de Theil	Mede o grau de desigualdade na distribuição da renda domiciliar per capita entre os indivíduos. É o logaritmo da razão entre as médias aritmética e geométrica das rendas individuais, sendo nulo quando não existir desigualdade de renda entre os indivíduos, e tendente ao infinito quando a desigualdade tender ao máximo. Chamado de “L de Theil”.
Riqueza	
Pessoas pobres (P0) - (%)	Percentual de pessoas com renda domiciliar per capita inferior a R\$ 75,50, equivalente a 1/2 do salário mínimo vigente em agosto de 2000. Chamado de “Pessoas pobres”.
Infraestrutura	
Percentual de domicílios com iluminação elétrica (%)	Percentual de domicílios com iluminação elétrica. Chamado de “Iluminação elétrica”.
Percentual de domicílios com instalações sanitárias ligadas à rede geral (%)	Percentual de domicílios com instalações sanitárias ligadas à rede geral de esgoto. Chamado de “Instalações sanitárias”.
Desenvolvimento	
Índice de Desenvolvimento Humano Municipal (IDH-M)	É obtido pela média aritmética simples de três sub-índices, referentes às dimensões Longevidade (IDH-Longevidade), Educação (IDH-Educação) e Renda (IDH-Renda). Varia de zero a um. Chamado de “IDH”.

Fonte: IPEADATA (2012)

Os parâmetros apresentados acima, escolhidos como indicadores de qualidade de vida, ou de bem estar, são coerentes com as dimensões e subdimensões dos métodos de análise de sustentabilidade estudados por Van Bellen (2005). Há que se ponderar sobre as limitações da disponibilidade de dados e a ausência de outros indicadores considerados mais apropriados pelo autor.

Outros indicadores, aqui chamados de “*predictors*”, devido sua utilização na análise discriminante, são apresentados no Quadro 5. Os indicadores referentes a transferências tributárias e de receita municipal, bem como os dados de renda em cada setor da economia, foram considerados baseados em Silva (2008), que estudou a influência da produção de cana-de-açúcar no IDH municipal no estado de São Paulo. Os indicadores referentes à produção de cana-de-açúcar foram propostos para entender a importância e a dependência da cana-de-açúcar nos municípios estudados, e são baseados nos valores monetários da produção.

Os dados de renda foram coletados na RAIS, Relação Anual de Informações Sociais, um registro administrativo que os estabelecimentos são obrigados a responder com referência a 31 de dezembro do ano anterior, e que retrata os empregos formais no Brasil naquela data (RAIS, 2012).

O valor adicionado da produção de cana-de-açúcar foi calculado a partir do índice de participação do valor adicionado em relação ao valor de produção de cana-de-açúcar a preço básico, presente na Matriz de Insumo Produto do Brasil dos anos de 1975 e 2000 (IBGE, 2012). Os valores adicionados para a cana de 1971 e 1980 foram calculados usando o mesmo índice de 1975, e os valores adicionados de 2000 e 1990 foram calculados usando o índice de 2000 (CGEE, 2009).

Quadro 5 Indicadores econômicos “ <i>predictors</i> ” selecionados para esta dissertação	
Produção de cana-de-açúcar – R\$ (2000)	Produção de cana-de-açúcar anual em mil reais de 2000 (IPEADATA, 2012).
Participação da cana-de-açúcar na agricultura (%)	Razão entre o valor adicionado da cana-de-açúcar e o valor adicionado referente à agricultura no município.
Participação da cana-de-açúcar na economia (%)	Razão entre o valor adicionado da cana-de-açúcar e o PIB municipal.
Produção de cana-de-açúcar por habitante (mil R\$/habitante)	Razão entre o valor da produção de cana-de-açúcar no município e sua população.
Produção de cana-de-açúcar por quilômetro quadrado (mil R\$/km ²)	Razão entre o valor da produção de cana-de-açúcar no município e sua área total.
Cota-parte do fundo de participação municipal (FPM) (R\$)	Destina-se ao registro das transferências referentes ao FPM. O FPM, com objetivos redistributivos, é composto por 22,5% do produto da arrecadação do Imposto de Renda e do Imposto sobre Produtos Industrializados (IPEADATA, 2012).
Transferências correntes de tributos estaduais para os municípios (R\$)	Transferências correntes recebidas do nível de governo municipal (IPEADATA, 2012).
Participação da indústria na economia (%)	Razão entre o valor adicionado da indústria no município e o PIB municipal.
Participação dos serviços na economia (%)	Razão entre o valor adicionado dos serviços no município e o PIB municipal.
Participação da agropecuária na economia (%)	Razão entre o valor adicionado da agropecuária no município e o PIB municipal.
Receita total (R\$)	Receitas provenientes da realização de recursos financeiros oriundos de constituição de dívidas; da conversão em espécie, de bens e direitos; recursos recebidos de outras pessoas de direito público ou privado, tributária, de contribuição, patrimonial, industrial, agropecuária, de serviços e tributos de competência do município: impostos, taxas e contribuição de melhoria (IPEADATA, 2012).
Rendimento médio do trabalhador da indústria (R\$)	Rendimento médio do trabalhador do setor industrial no município, em reais de 2000 (RAIS, 2012).
Rendimento médio do trabalhador dos serviços (R\$)	Rendimento médio do trabalhador no setor de serviços no município, em reais de 2000 (RAIS, 2012).
Rendimento médio do trabalhador do comércio (R\$)	Rendimento médio do trabalhador no setor comercial no município, em reais de 2000 (RAIS, 2012).
Rendimento médio do trabalhador da agropecuária (R\$)	Rendimento médio do trabalhador do setor da agropecuária no município, em reais de 2000 (RAIS, 2012).

Fonte: IPEADATA (2012); RAIS (2012)

1.2 Revisão bibliográfica

Os estudos já realizados sobre aspectos socioeconômicos da produção de etanol em larga escala, no Brasil, abrangem principalmente questões de qualidade de trabalho, renda e desenvolvimento rural. Alguns desses trabalhos visam estimar a importância da atividade canavieira, assim como seus impactos sobre a sociedade. Há diferenças metodológicas e divergência de pontos de vista sobre os impactos socioeconômicos da indústria sucroenergética.

1.2.1 Emprego e Renda

Moraes et al. (2010) tratam da grande difusão da indústria sucroenergética e sua grande capilaridade, gerando empregos e ajudando a desenvolver o interior do país. Para o estado de São Paulo, a cana-de-açúcar representa ainda maior importância, estando sua produção presente em 63,3% dos municípios. O estado emprega 50,8% dos trabalhadores do setor, e os demais estão distribuídos em 24 unidades federativas.

O número de empregos vem aumentando consideravelmente nos últimos anos. Entre 2002 e 2006 o número de empregos no cultivo de cana-de-açúcar cresceu 15,3%, 73,6% na produção de açúcar e 8,8% na produção de etanol (HOFFMANN, 2008). Esse aumento foi maior ainda para a produção de etanol quando comparado com o ano de 2009, que teve um aumento de 106,2%, gerando 135.058 empregos em 2009 (PNAD, 2009).

Fernandes et al. (2011) mostram que esse crescimento se deu principalmente na Região Centro-Oeste, já que gerou mais postos de trabalho no período estudado pelos autores, de 1995 a 2009, especialmente nos estados de Mato Grosso do Sul e Goiás. Na Região Nordeste, o estado de Pernambuco apresentou grande número de perdas no mercado de trabalho do setor. Os autores verificaram também que o sub-setor sucroalcooleiro de destaque foi o da indústria de açúcar e álcool, em detrimento da lavoura de cana-de-açúcar. Apesar do estado de São Paulo ter tido perdas de postos de trabalho na lavoura, foi no estado que mais cresceu a geração de empregos na indústria de açúcar e álcool. Segundo os autores, tais resultados mostram que a dinâmica no mercado de trabalho formal no setor sucroalcooleiro, no Brasil, está associada com o avanço dessa atividade, mormente, para novas fronteiras agrícolas localizadas no Centro-Oeste.

O rendimento dos trabalhadores também teve um aumento importante entre os anos de 2002 e 2006, e o resultado está bastante associado à escolaridade. A Tabela 1.1 mostra essa evolução.

Tabela 1.1 – Empregos, rendimento e escolaridade entre 2002 e 2009

		2002	2004	2006	2009	Variação 2002 - 2009
Cana	Empregos (1.000)	491	527,7	566	576,3	17,4%
	Rendimento	417	496,3	568,1	733,7	75,9%
	Escolaridade	2,8	3,2	3,7	4,5	60,7%
Etanol	Empregos (1.000)	65,5	86,7	71,3	135,1	106,2%
	Rendimento	1.192	883	1.254	1.298	8,9%
	Escolaridade	7,3	7,9	8,6	8,6	17,8%
Açúcar	Empregos (1.000)	92,8	110,1	161,1	125,3	35,0%
	Rendimento	1.181	765,2	1.201,6	1.183	0,17%
	Escolaridade	6,5	6,5	7,9	8,2	26,2%

¹ Valores deflacionados para o ano de 2002, utilizando o INPC (Índice Nacional de Preços ao Consumidor) médio dos meses de setembro e outubro dos anos estudados.

Fonte: HOFFMAN (2008), PNAD (2009)

Moraes (2007) também estuda a distribuição de idade e a escolaridade dentro dos três grupos de produção do setor. A autora estudou as características do trabalhador entre os anos de 1992 e 2005. Como é de se esperar, o rendimento do trabalhador aumenta com sua escolaridade. Na comparação entre as atividades do setor, houve uma leve mudança em relação ao nível de escolaridade e o respectivo rendimento. Para todos os níveis de educação, a atividade de fabricação de açúcar tinha, em 2005, um rendimento maior. Em 2010, comparando com os dados da RAIS, a produção de álcool teve maior rendimento em quase todos os níveis de escolaridade, com exceção dos níveis superior completo e incompleto e mestrado completo no mercado de trabalho formal (RAIS, 2010). Essa mudança reflete o crescimento da importância do etanol tanto no setor, como na economia do país.

Outro aspecto importante é a pequena participação das mulheres no setor. No ano de 2010, elas representaram 9,2% de todos os empregados do setor, sendo 5,6% dos trabalhadores do refino do açúcar e 11% no cultivo da cana.

Em estudo sobre gênero e corte da cana, Caumo et al. (2011) identificaram as características e diferenças entre os trabalhadores homens e mulheres. Em seus resultados, os autores mostram que quanto à raça/cor dos trabalhadores, observou-se a predominância de brancos e de pardos, tanto entre os homens quanto entre as mulheres, e que a inserção das mulheres no corte da cana-

de-açúcar ocorre em idade mais avançada do que a dos homens; porém elas permanecem atuando na atividade até uma faixa etária mais avançada. Além disso, o grupo de mulheres estudado é, em sua maioria, da própria cidade tomada como estudo de caso. Com relação à produtividade dos trabalhadores entrevistados, os homens cortam, em média, 374,3 metros de cana-de-açúcar (linhas de plantio) por dia de trabalho, quantidade bem superior ao rendimento das mulheres, que é de 216,5 metros por dia.

A idade do trabalhador pouco variou entre os anos de 2005, tratado no trabalho de Moraes (2007), e de 2010 (último ano disponível da RAIS). A faixa com maior concentração de trabalhadores no setor é de 30 a 39 anos. Grande parte das faixas etárias teve crescimento, de acordo com o crescimento do número de trabalhadores.

As diferentes regiões produtoras também tiveram grandes mudanças quanto ao rendimento salarial do trabalhador. Tendo 35% dos trabalhadores em 2010, a região Nordeste é ainda a com menores níveis de salários médios. Numa comparação direta entre as regiões, no ano de 2005 os trabalhadores do Nordeste recebiam, em média, 40% a menos que os da região Sudeste (em valores monetários de dezembro de 2005) (MORAES, 2007). Em 2010 a diferença subiu para 46%, em média (em valores monetários de dezembro de 2010) (RAIS, 2012).

Hoffmann et al. (2008) também compararam o rendimento dos trabalhadores envolvidos com a produção da cana-de-açúcar e o daqueles envolvidos com outras culturas (arroz, banana, café, citros, mandioca, milho e soja). Os autores mostraram que, em média, os trabalhadores envolvidos com o cultivo de soja e citros têm rendimentos mais elevados, enquanto a menor renda é dos trabalhadores envolvidos com o cultivo de milho e mandioca. É importante observar que, segundo os autores, na cultura da soja os trabalhadores têm, em média, maior nível de escolaridade e exercem funções mais especializadas. Enquanto 41,1% dos trabalhadores da soja são operadores de máquinas (tratores), na produção de citros e cana eles são 14,2% e 4,3%, respectivamente. Os autores mostraram ainda que, para todas as culturas, houve um aumento do rendimento real durante o período de 1992 a 2006.

Devido as atuais mudanças na legislação (Lei nº 11.241, de 19 de setembro de 2002) e às emissões de gases do efeito estufa, a mecanização da colheita será obrigatória, gradualmente, na grande maioria dos canaviais no estado de São Paulo. Por isso, há uma grande preocupação quanto o deslocamento dos trabalhadores para outros setores, bem como sua requalificação para que possam ser absorvidos pelo próprio setor, em áreas novas como colheita mecanizada, plantio, etc.

Baccarin et al. (2011) dizem que, apesar do crescimento da produção setorial, o nível de ocupação das empresas sucroalcooleiras praticamente se manteve entre 2007 e 2010. Enquanto decrescia expressivamente o número de “Trabalhadores Canavieiros não Qualificados”, em que se encontram os cortadores de cana, aumentava, em intensidade um pouco menor, o número na categoria “Outras Ocupações Sucroalcooleiras”, em que a qualificação profissional é maior. Os autores dizem ainda que houve aumento considerável do número na categoria “Trabalhadores na

Mecanização Agrícola”, embora em número absoluto bem menor do que a queda dos “Trabalhadores Canavieiros não Qualificados”.

Trabalhando com uma usina de 1.000.000 toneladas moídas por ano, como referência, Vieira (2003) mostra que, com a mecanização da colheita, a usina deverá demitir 397 colaboradores até 2011, 1.032 até 2016 e 688 até 2021, totalizando 2.117 trabalhadores para apenas uma usina estudada. Essas demissões somam mais de R\$ 9.000.000 em massa salarial (em valores de 2001). O autor utilizou como referência o corte manual de 7,5 t/homem, e que cinco máquinas substituiriam 383 homens no caso estudado. No caso estudado, essa substituição equivaleria no final de 2021 à redução da massa salarial em R\$ 9.759.918 (em reais de 2001). Com as novas contratações, a massa salarial acumulada de 2006 a 2021 chegaria a apenas R\$ 2.364.750, com contratação total de 177 empregados.

1.2.2 Qualidade do Emprego

O trabalho de corte de cana depende basicamente da força, destreza e agilidade do trabalhador. As atividades são de alto risco para a saúde e Rocha (2007) indica que as principais doenças estão ligadas a execução de movimentos que exigem adoção de postura inadequada e forçada, e estar exposto a condições ambientais adversas, como radiação solar, calor intenso e grande quantidade de poeira e fuligem. O trabalho executado pelo cortador excede os limites de tolerância do sistema osteomuscular, podendo causar doenças como dorsalgias, cervicalgias, tenossinovites, tendinites bursites e artroses. Muitos estudos indicam ainda mortes nos canaviais, decorrentes da intensa atividade do trabalhador. Além das mortes ocorridas nos canaviais, há aquelas não registradas, e que ocorrem após um tempo determinado. Doenças como câncer, provocado pelo uso de agrotóxicos, contato com fuligem da cana, além de doenças respiratórias, alérgicas, e da coluna, aliadas a quase total impossibilidade dessas doenças serem adequadamente tratadas em razão da inexistência de recursos financeiros (MENDONÇA, 2006).

O esforço realizado pelo cortador pode ser considerado por muitos como forçado, por estar diretamente ligado ao seu salário. O trabalhador recebe de acordo com a quantidade cortada e, assim, o cortador está cada vez mais eficiente. É apontado ainda que a maioria dos trabalhadores não tem controle da pesagem ou da metragem de sua produção diária, que é feita pela usina. Muitas denúncias apontam para a manipulação e fraude desses dados pelas usinas, que pagam menos do que os trabalhadores teriam direito (MENDONÇA, 2006; ORTIZ, 2007; NOVAES, 2007). A produtividade quase dobrou em 20 anos, sem que houvesse mudança nos instrumentos de trabalho (ALVES, 1992). Segundo Alves (1992), a produtividade de um cortador no início dos anos 1990 era 10 t/dia.

No sistema de corte manual, as exigências na seleção de trabalhadores são físicas e o tipo de contrato de trabalho é por tempo determinado, ou seja, por safra. Nesse tipo de contrato os trabalhadores não recebem, por lei, o seguro desemprego no final do contrato (NOVAES, 2007). O cortador recebe três refeições diárias, e algumas usinas ainda têm projetos de incentivo de

produção e permanência do trabalhador, com cestas básicas ou auxílios em dinheiro, além do salário. Mas as refeições são consideradas pobres em nutrientes por Rocha (2007), por não conterem legumes e vegetais, nem mesmo carnes para suprir as necessidades protéicas de um trabalho que requer alto esforço físico; porém, isto varia de usina para usina.

Quanto aos acidentes e doenças de trabalho, o cultivo de cana-de-açúcar tem maus resultados. A atividade foi responsável entre 1997 e 1999 por 40% (14.661) dos *acidentes-tipo* (acidentes decorrentes da atividade do trabalhador) ocorridos na agricultura no estado de São Paulo, e 52% (2.069) das *doenças de trabalho* (acidentes ocasionados por qualquer tipo de doença peculiar a determinado ramo de atividade). Em relação ao total de acidentes de trabalho na agricultura no estado, naquele período, 28% dos acidentes e 38% das doenças estão no cultivo da cana. Apenas 0,15% do total dos acidentes de trabalho no estado, naquele período, resultaram em óbito e apenas 0,11% em parada permanente (TEIXEIRA, 2003). Essa participação dos acidentes na atividade canavieira deve ser analisada levando em conta a participação dos trabalhadores da cana no total da agropecuária. No que se trata de acidentes de trabalho, a UNICA (União das Indústrias de Cana-de-Açúcar) mostra, com dados recentes, que em 2010 houve 6.075 acidentes na área agrícola, e 2.552 na área industrial/administrativa das suas associadas (88 usinas para esses dados). Os investimentos em saúde e segurança pelas associadas à UNICA chegaram a mais de R\$ 87 milhões no ano de 2010. Isso representa 0,3% do faturamento declarado das usinas (UNICA, 2011).

O trabalho forçado também é uma preocupação dentro do setor. De acordo com a convenção 29 de 1930 da OIT (Organização Internacional do Trabalho), trabalho forçado é definido como “todo trabalho ou serviço exigido de uma pessoa sob a ameaça de sanção e para o qual não se tenha oferecido espontaneamente.” Nesse aspecto, o setor canavieiro também apresenta maus resultados. Apesar de representar apenas 7% dos casos denunciados em todo o país, esses casos são responsáveis por 31% (1.911) de todos os trabalhadores envolvidos em casos semelhantes à escravidão no ano de 2009. No estado de São Paulo não houve nenhum registro de trabalho forçado até 2009, mas vários casos de irregularidades trabalhistas foram encontrados, de acordo com auditoria feita pelo Grupo Estadual Rural no mesmo ano. Nessas auditorias, foram registrados casos de falta de repouso após seis horas consecutivas de trabalho; excesso de jornada; prorrogação de jornada sem autorização, falta de registro de entrada e saída dos funcionários, trabalho aos domingos sem autorização, irregularidades em EPIs, banheiro sem higiene, etc. (Repórter Brasil, 2010).

Por outro lado, há indícios de ações positivas para a melhoria das condições de trabalho. Na Tabela 1.2 são mostrados os benefícios oferecidos pelas usinas em uma amostra de 47 estabelecimentos.

Tabela 1.2 – Benefícios oferecidos em uma amostra de 47 estabelecimentos (2003)

Benefício	% da amostra
Plano de saúde	95,7%
Plano odontológico	93,5%
Transporte	93,3%
Seguro de vida coletivo	91,5%
Refeição	87,0%
Auxílio farmácia	85,1%
Tratamento auditivo	63,8%
Cesta de Natal	59,1%
Cooperativa de crédito	37,8%
Cesta básica	43,5%
Auxílio educação	35,6%
Auxílio doença	20,0%

Fonte: adaptado de Barbosa (2008)

A UNICA e suas associadas têm ações para a requalificação dos trabalhadores para os novos postos criados com o crescimento da mecanização da colheita. Em dois anos foram investidos quase R\$ 11 milhões na área agrícola e mais de R\$ 13 milhões na área industrial, representando esses investimentos 0,09% do faturamento total declarado das associadas para o ano de 2009.

1.2.3 Análise de indicadores socioeconômicos em municípios canavieiros

Alguns estudos foram feitos com o objetivo de verificar os impactos da produção de etanol, cana-de-açúcar e açúcar nos indicadores socioeconômicos de uma região. As abordagens adotadas diferem, mas em grande parte deles o IDHM (Índice de Desenvolvimento Humano de Municípios) é o indicador empregado para avaliação da qualidade de vida. Como anteriormente comentado, o IDH tem três componentes: educação, longevidade e renda. Na Tabela 1.3 são apresentadas informações sobre os indicadores que compõem o índice.

Tabela 1.3 – Indicadores Componentes do IDHM

Indicador	Definição
Esperança de vida ao nascer (em anos)	Número médio de anos que as pessoas viveriam a partir do nascimento.
Taxa de alfabetização de adultos (%)	Percentual de pessoas acima de 15 que sabem ler e escrever.
Taxa bruta de frequência escolar (%)	Proporção entre o número total de pessoas em todas as faixas etárias que freqüentam os cursos fundamental, segundo grau ou superior, em relação ao total de pessoas na faixa etária de 7 a 22 anos.
Renda per capita	Razão entre o somatório da renda de todos os indivíduos e a população total.
Índice de longevidade (IDHM-L)	Índice do IDHM relativo à dimensão Longevidade. É obtido a partir do indicador esperança de vida ao nascer, através da fórmula: (valor observado do indicador - limite inferior) / (limite superior - limite inferior); os limites inferior e superior são equivalentes a 25 e 85 anos, respectivamente.
Índice de Educação (IDHM-E)	Índice do IDHM relativo à Educação. Obtido a partir da taxa de alfabetização e da taxa bruta de frequência à escola, convertidas em índices por: (valor observado - limite inferior) / (limite superior - limite inferior), com limites inferior e superior de 0% e 100%. O IDHM-Educação é a média desses dois índices, com peso 2 para o da taxa de alfabetização e peso 1 para o da taxa bruta de frequência.
Índice de Renda (IDHM-R)	Deriva da renda municipal per capita. Primeiro convertem-se os valores anuais máximo e mínimo expressos em dólar PPC (Paridade do Poder de Compra), em valores mensais expressos em reais. Em seguida, são calculados os logaritmos da renda média municipal per capita e dos limites máximo e mínimo de referência. O logaritmo é usado porque expressa melhor o fato de que um acréscimo de renda para os mais pobres é proporcionalmente mais relevante do que para os mais ricos. Finalmente, para se chegar ao índice de renda municipal (IDHM-R) aplica-se a fórmula: $IDH-R = (\log \text{ de renda média municipal per capita} - \log \text{ do valor de referência mínimo}) / (\log \text{ do valor de referência máximo} - \log \text{ do valor de referência mínimo})$.
Índice de Desenvolvimento Humano Municipal (IDHM)	É obtido pela média aritmética simples dos três índices, referentes às dimensões Longevidade (IDHM Longevidade), Educação (IDHM-Educação) e Renda (IDHM-Renda).

Fonte: Relatório de Desenvolvimento Humano (2000)

Silva (2008) fez uma regressão múltipla para identificar o quanto o IDH-M é influenciado pela presença do setor canavieiro no município. Nessa regressão, a variável dependente é o IDHM e, para variáveis explanatórias, buscou-se selecionar um conjunto de indicadores que pudessem, teoricamente, afetar o desenvolvimento humano de cada município, bem como as variações desse indicador entre os municípios. Porém, foi tomado o cuidado para não serem tomadas variáveis

que compõem o próprio IDHM. Algumas das variáveis consideradas são os impostos arrecadados, despesas com saúde, importância relativa de cada setor, etc. Além de, é claro, a participação do setor no município.

Para identificar a participação, a autora utilizou variáveis *dummies* (binárias). A existência ou não de usinas no município é registrada por uma variável binária. O mesmo ocorre quando o cultivo da cana-de-açúcar é a principal atividade agrícola do município.

Silva (2008) conclui que, se não são considerados os efeitos cruzados, a presença do setor eleva a média do IDHM para os municípios que têm como atividade agrícola principal a cana-de-açúcar. Porém, ao considerar os efeitos cruzados das variáveis binárias (que indicam a presença do setor) sobre outras variáveis (como arrecadação de impostos, por exemplo), o setor apresenta impactos negativos, ou seja, o impacto da variável em questão sobre o IDHM diminui, ou é mesmo negativo.

Em Campeão et al. (2009) é estudada a relação do IDHM com a quantidade produzida de cana. Os autores trabalham com o coeficiente de correlação de Pearson, que indica se há correlação entre duas variáveis métricas em dois casos: Os municípios produtores foram divididos em quatro grupos, com número aproximadamente igual de integrantes; outro caso é da análise conjunta de todos os municípios. Para todos os municípios os autores obtiveram um coeficiente de correlação 0,17, estatisticamente significativo. Esse valor indica uma fraca relação entre as variáveis IDHM e a quantidade produzida de cana.

Para o caso no qual há separação dos municípios em grupos, em apenas dois o coeficiente de correlação entre as variáveis é estatisticamente significativo, como mostrado na Tabela 1.4.

Tabela 1.4-Coeficiente de Correlação de Pearson para IDHM e produção de cana

Grupo (em toneladas)	Correlação	p-valor
Até 600	0,11	0,004
Entre 600 e 2.800	0,005	0,90
Entre 2.800 e 32.000	-0,02	0,62
Mais que 32.000	0,24	0,000

Fonte: Adaptado de Campeão et al. (2009)

No trabalho de Chagas (2010) é feito um estudo de regiões produtoras de cana em todo o Brasil, e de seus IDH. O autor compara regiões com produção de cana e sem produção, além de utilizar o método *propensity score*, que foi introduzido por Rosenbaum e Rubin (1983). *Propensity score* é a probabilidade de um indivíduo (nesse caso, um município) ser alocado a uma condição particular em um estudo, dado um conjunto de covariâncias. A utilização do *score* justifica-se pelo caráter geográfico do problema, e permite comparar regiões produtoras com regiões não produtoras, consideradas similaridades em termos de clima, solo, proximidade, distância de usina, legislação; tais aspectos definem as covariâncias.

Na primeira parte da análise, sem a consideração do *propensity score*, conclui-se que o IDH das regiões produtoras é sempre maior do que o das não produtoras, independente da porcentagem de área agrícola utilizada na plantação de cana-de-açúcar. Com a consideração do *propensity score*, conclui-se que os valores do IDH das regiões produtoras de cana são menores, mas que estatisticamente não é possível concluir que haja diferença entre regiões produtoras e não produtoras. O resultado indica que a presença da cana de açúcar em certa localidade não é determinante para suas condições sociais (CHAGAS, 2010).

Outros dois estudos incluem o IDH em suas avaliações, mas não se restringem a esse indicador. Camargo e Toneto (2008) estudaram, além do IDH, o IPRS (Índice Paulista de Responsabilidade Social), os rendimentos na agricultura, comércio e indústria, a arrecadação de impostos (IPVA, ICMS e total de impostos) e o PIB per capita dos 645 municípios do estado de São Paulo. Na pesquisa os municípios paulistas foram divididos em quatro grupos:

BASE – grupo dos municípios nos quais não existem usinas e a produção de cana é menor que a mediana do estado;

CANA – grupo no qual não existem usinas, mas há forte presença de cana (maior que a mediana do estado);

CANA+USINA – grupo no qual existem usinas e forte presença de cana;

RMSP – região metropolitana de São Paulo.

O trabalho consistiu em comparar os indicadores escolhidos desses quatro grupos. No geral, os autores encontraram forte relação entre a presença da cana e usina (grupo CANA+USINA) e os resultados dos indicadores. As comparações foram feitas através de médias e, ao que tudo indica, não foram feitos testes estatísticos para confirmar os melhores resultados obtidos pelo grupo CANA +USINA.

Oliveira (2011) realiza uma ampla análise para todos os municípios dos estados maiores produtores de cana-de-açúcar: São Paulo, Paraná, Minas Gerais, Alagoas e Pernambuco. Os indicadores analisados foram: IDHM, Analfabetos, Esperança de vida ao nascer, Mortalidade infantil, Pobreza (% de pessoas pobres), Distribuição de renda (Índice L de Theil), Domicílios com luz elétrica (%) e Domicílios com instalações sanitárias (%). Para o estado de São Paulo também foram analisados o IPRS (Índice Paulista de Responsabilidade Social) e seus subíndices (renda, longevidade e escolaridade).

A análise foi feita comparando os grupos “com cana” e os “sem cana”. Os municípios foram classificados de acordo com a produção de cana em cada ano. O grupo dos municípios canavieiros é aquele que contribui com a produção de 90% do total no estado, em cada ano. O grupo dos municípios não canavieiros foi definido entre aqueles que têm a mesma faixa populacional do grupo anterior, e não têm significativa produção de cana em cada ano.

Foram calculadas as médias dos indicadores e seus desvios padrão, e o teste de hipótese t de Student foi empregado para garantir confiança estatística à comparação das médias dos indicadores entre diferentes grupos.

A autora conclui que para todos os casos analisados, com raras exceções, não há evidências de que a atividade canavieira proporcione desvantagens aos municípios sob um enfoque socioeconômico. Dos resultados da comparação de indicadores entre municípios canavieiros e não canavieiros pode-se concluir que, nos municípios nos quais a produção de cana-de-açúcar é mais significativa, os resultados socioeconômicos são melhores. Os melhores resultados e a vantagem relativa dos municípios canavieiros são claros no estado de São Paulo, onde todos os indicadores são melhores com confiança estatística mínima de 95% em todos os anos.

Esta dissertação tem como base principal o trabalho realizado por Oliveira (2011), com o emprego dos mesmos indicadores socioeconômicos e buscando aprofundar as análises feitas pela autora. Além disso, existe forte conexão entre as análises em Camargo e Toneto (2008), principalmente na comparação das médias, Campeão et al. (2009) nas correlações, e Silva (2008) nos indicadores utilizados para análise discriminante.

Uma importante diferença entre os trabalhos citados e esta dissertação é o tratamento dos dados relacionados à produção de cana. Além de tratar com grandeza monetária, a produção é estudada apenas em termos relativos, seja por área, habitante ou produto interno bruto do município. Autores como Camargo e Toneto (2008) conseguiram introduzir a presença das usinas em seu estudo; assim, com a separação dos grupos em com usina e sem, puderam analisar a presença do setor industrial da indústria sucroenergética. Aqui, nesta dissertação, houve a tentativa de incluir as atividades industriais usando a participação da indústria na economia municipal, porém não foi possível tratar isoladamente o impacto da atividade industrial do setor.

Oliveira (2011) teve importante influência na escolha dos indicadores. Estes foram escolhidos com base na literatura e também na sua disponibilidade. Além de dar início a pesquisa com a comparação das médias e análise da evolução usando gráficos radar.

2. Revisão metodológica

Neste capítulo são apresentadas as metodologias usadas nas análises realizadas nesta dissertação. É apresentada uma introdução aos testes de hipótese e informações sobre Análises de Cluster e Discriminante, incluindo seus pré requisitos de aplicação.

2.1 Estatística Básica

O objetivo principal da estatística é permitir inferências sobre uma população a partir de uma amostra da mesma. Em uma amostra, existe uma distribuição de frequência que indica quantas vezes um resultado ocorreu dentro da amostra. Dessa distribuição de frequência podem-se tirar algumas características, principalmente seus centros de distribuição (WONNACOTT & WONNACOTT, 1981).

Os centros de distribuição são três: a moda, a mediana e a média. A moda de uma distribuição é o valor mais frequente da mesma. A moda não é necessariamente uma tendência central, pois quase sempre depende do agrupamento arbitrário de dados. A mediana é o valor abaixo do qual recaem 50% dos valores da amostra. A média é a mais comum das medidas, somando-se todas as observações iniciais e dividindo o resultado pelo número total de observações, como indicado na equação 1.

$$\frac{1}{n} \cdot \sum_{i=1}^n X_i \quad \text{Equação 1}$$

Igualmente importante como conhecer a média de uma distribuição é conhecer sua dispersão. As duas medidas mais conhecidas de dispersão são variância e desvio padrão. A variância, conforme indicado na equação 2 é a soma do desvio de cada observação a contar da média, elevada ao quadrado, já que isso evita o cancelamento dos desvios negativos pelos desvios positivos. A soma é então dividida pelo número de observações, menos um.

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{Equação 2}$$

O desvio padrão é simplesmente a raiz quadrada da variância. Isso reduz os resultados às mesmas unidades das observações X_i (WONNACOTT & WONNACOTT, 1981).

Das distribuições de frequência, a normal é a mais conhecida. Uma vez que características naturais, como altura do ser humano, seu peso e outros parâmetros têm o gráfico de suas distribuições em forma de sino, emprega-se o termo normal (SIRKIN, 2006). A representação gráfica de distribuição de parâmetros é feita na Figura 2.1 e, como pode ser visto, toda curva normal é simétrica e assintótica em relação ao eixo das abscissas. A figura mostra também as probabilidades de casos ocorrerem dentro das áreas apresentadas (em função do desvio padrão).

A distribuição de probabilidade de uma normal padronizada, que é chamada assim por ter média zero e desvio padrão um, segue a função da equação 3 $p(z) = \frac{1}{\sqrt{2 \cdot \pi}} e^{-(1/2) \cdot z^2}$ Equação

3 (WONNACOTT & WONNACOTT, 1981). Na medida em que o valor de z^2 cresce, ou seja, que se afasta de zero, o expoente negativo de e (constante matemática) cresce, fazendo decrescer a probabilidade $p(z)$, tendendo para zero nas duas extremidades.

$$p(z) = \frac{1}{\sqrt{2 \cdot \pi}} e^{-(1/2) \cdot z^2} \quad \text{Equação 3}$$

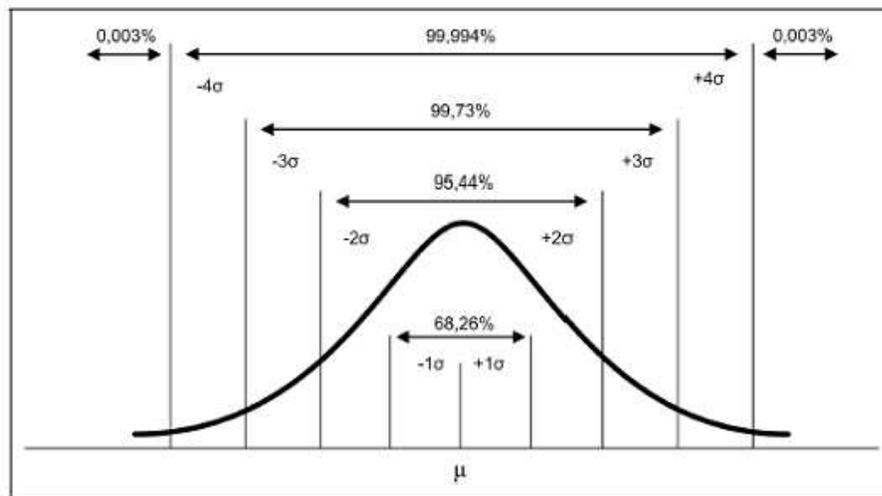


Figura 2.1 – Distribuição Normal e alguns resultados importantes.
Fonte: Action (2011)

2.2 Teste de Hipóteses

O teste de hipótese utilizado neste trabalho é o mesmo aplicado por Oliveira (2011) na comparação das médias dos indicadores dos municípios não canavieiros e canavieiros. O teste t é o método mais comumente usado para avaliar diferenças entre as médias de dois grupos. Teoricamente, o teste t pode ser utilizado tanto para grandes amostras quanto para pequenas (aproximadamente 10 observações), desde que as variáveis sejam normalmente distribuídas e que a variância nos dois grupos não seja estatisticamente diferente. Como será tratado posteriormente, a normalidade da distribuição pode ser avaliada com o teste de *Shapiro-Wilk* e com os índices de *kurtosis* e *skewness*, enquanto a homogeneidade da variância pode ser testada com o teste de *Levene*. Se essas condições (normalidade e homogeneidade) forem atendidas, o teste de hipótese

t pode ser aplicado com coerência (STATSOFT, 2011). Apesar de existirem testes t que não necessitam de homogeneidade de variância, o teste utilizado nesta dissertação, aplicado pelo software *Statistica*, exige tal condição.

Matematicamente, a hipótese de que as médias dos dois grupos estudados são diferentes é testada. Nesse caso, é necessário escolher um intervalo de confiança, que nesta dissertação será de 95%. Um intervalo de confiança representa o conjunto de hipóteses aceitáveis e, em geral, qualquer hipótese que fica fora do intervalo de confiança pode ser considerada implausível ou rejeitável. A hipótese de que a diferença entre as médias dos municípios canavieiros e não canavieiros é igual a 0 é chamada de hipótese nula, ou H_0 . A equação 4 indica o cálculo de t .

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{x_1x_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{Equação 4}$$

Sendo $S_{x_1x_2} = \sqrt{\frac{(n_1 - 1) \cdot S_{x_1}^2 + (n_2 - 1) \cdot S_{x_2}^2}{n_1 + n_2 - 2}}$, S_x^2 o desvio padrão de cada amostra, e n_i o tamanho de cada amostra. Se o resultado de t for inferior, em módulo, a um nível de significância de 5%, a hipótese H_0 é plausível (WONNACOTT & WONNACOTT, 1981).

2.3 Análise Multivariada

Conjuntos de dados complexos podem surgir de diferentes maneiras. Uma delas é quando o tamanho da amostra é relativamente pequeno, mas muitas variáveis inter-relacionadas são medidas em cada unidade. Em outras situações, menos variáveis podem estar presentes, mas o tamanho da amostra pode ser muito grande, o que dificulta a assimilação. Em outros casos, é claro, pode-se ter muitas variáveis bem como grandes amostras. A coleção de grandes conjuntos de dados é mais comum, em parte por causa do desenvolvimento de dispositivos computacionais cada vez mais sofisticados e, em parte, porque é possível armazenar grandes conjuntos de dados eletronicamente de tal forma que uma recuperação instantânea e a classificação dos dados não são problemas (KRZANOWSKI, 2007).

Para tratar de conjuntos de dados complexos usa-se a análise multivariada, que se refere a toda técnica estatística que analisa, simultaneamente, múltiplos indivíduos ou medidas. Muitos fenômenos envolvem diversas dimensões, que podem estar entrelaçadas ou altamente relacionadas (HAIR et al., 2010; RAYKOV & MARKOULIDES, 2008).

A análise multivariada pode ser categorizada por: incluir vários indicadores (ou variáveis independentes), múltiplos critérios (ou resultados, ou variáveis dependentes), ou ambos. Técnicas

estatísticas multivariadas também podem ser distinguidas pelas relações esperadas. Os diferentes procedimentos multivariados são usados para testar hipóteses sobre o grau de associação entre variáveis, as diferenças entre grupos, e a previsão de diferenças entre os grupos (TINSLEY & BROWN, 2000).

Rencher (2002) indica quatro tipos básicos de dados e as possíveis análises que podem ser feitas utilizando métodos multivariados. De acordo com o autor, os dados podem ser combinações ou extensões de:

- Uma amostra com diversos parâmetros medidos em cada unidade da amostra;
- Uma única amostra com dois grupos de parâmetros medidos em cada unidade da amostra;
- Duas amostras com diversos parâmetros medidos em cada unidade da amostra;
- Três ou mais amostras com diversos parâmetros medidos em cada unidade da amostra.

O autor indica ainda análises possíveis com os dados. Neste trabalho, no estudo dos indicadores sociais municipais dos estados de São Paulo e Alagoas, tem-se duas amostras com oito parâmetros. Como será mostrado mais a frente, os municípios são divididos em grupos de acordo com os indicadores aqui chamados de bem estar. As amostras têm suas médias comparadas, uma combinação linear das variáveis que melhor separe as unidades das amostras em grupos é feita (Análise Discriminante), e ainda uma função das variáveis que aloque as unidades em uma das amostras ou grupos é criada (Análise Discriminante de Classificação).

Hair et al. (2010) denominam os métodos de análise como de dependência e de interdependência, sendo os de interdependência aqueles usados para variáveis que não podem ser denominadas dependentes ou independentes. O interesse é analisar todas as variáveis para identificar uma estrutura comum a todas as unidades. Se os casos, ou unidades das amostras, devem ser agrupados, o método de Análise de Cluster deve ser usado. Os autores ainda indicam Análise Discriminante Múltipla, ou simplesmente Análise Discriminante para as situações em que a amostra total deve ser dividida em grupos tendo por base uma variável não métrica. O objetivo desta análise é entender as diferenças nos grupos formados e estimar em qual deles uma nova unidade da amostra se encontrará, baseado em suas variáveis (métricas) medidas.

2.4 Triagem dos dados

Em um procedimento de análise como o que foi feita nesta dissertação, Raycov e Markoulides (2008), e Tabachnick e Fidell (2007), indicam que a primeira ação a ser feita é a revisão dos dados. Primeiramente deve-se revisar cuidadosamente se os dados foram inseridos corretamente, e que nenhum valor foi trocado. Quando o número de casos é grande e sua análise individual quase impossível, os autores sugerem o cálculo das médias para cada indicador, e a verificação de que seus valores mínimos e máximos estão dentro de uma faixa plausível.

Hair et al (2010) consideram o uso de gráficos adequado para identificar correlações antes de qualquer análise empírica usando métodos estatísticos. Uma simples plotagem de cada caso da

amostra, escolhendo-se dois indicadores, pode identificar a relação, negativa, ou positiva, ainda que apenas em uma análise superficial.

Casos discrepantes devem receber maior atenção por exercerem grande influência sobre os resultados. Casos discrepantes (em inglês, *outliers*) são casos ou observações com uma combinação única de características identificadas como distintas de outras observações. Tipicamente, é julgado como uma observação não usualmente alta ou baixa, ou uma combinação única de diversas variáveis que faz a observação se sobressair em relação às outras (HAIR et al., 2010).

Tabachnick e Fidell (2007) encontraram quatro razões para a existência de discrepâncias: 1) erro na coleta de dados, 2) problemas ao identificar dados inexistentes, 3) a observação discrepante não deveria fazer parte da amostra que se deseja estudar e 4) a distribuição da variável tem valores mais extremos que o resto da amostra.

A procura de discrepâncias em análises multivariadas pode ser feita utilizando a distância de *Mahalanobis*, que identifica a distância de um indivíduo da amostra em relação a centroide de todos os casos (RAYCOV & MARKOULIDES, 2008). A distância *mahalanobis* é a mesma que a distância euclidiana quando as variáveis independentes não são correlacionadas. A distância euclidiana é tratada na seção sobre Análise de Cluster.

Identificadas as discrepâncias, há agora a necessidade de saber em quais variáveis a observação discrepante está mais distante dos outros casos. Essa identificação ajudará entender se a observação faz realmente parte da amostra e, no caso de modificar os dados em vez de apagá-los, deve-se saber quais dados modificar (TABACHNICK & FIDELL, 2007).

Para utilizar os métodos de análise multivariada são feitas algumas considerações em relação aos dados utilizados. São elas normalidade, linearidade e homoscedasticidade. Cada consideração será tratada a seguir.

2.5 Normalidade

A consideração mais fundamental em análise multivariada é a normalidade, referindo-se à forma da distribuição dos dados para uma variável métrica e sua correspondência com a distribuição normal. Se o desvio em relação a distribuição normal for suficientemente grande, todos os resultados de testes estatísticos são inválidos, pois normalidade é uma exigência para usar os testes estatísticos z , F e t . Normalidade multivariada significa que todas as variáveis são normais individualmente, e suas combinações também são normais (HAIR et al, 2010).

Testar uma amostra e suas variáveis quanto à normalidade não é praticamente possível, mas existem implicações a essa condição que podem ser testadas empiricamente. Para tanto, dois índices de distribuição são utilizados: *skewness* e *kurtosis* (obliquidade e curtose, em português). *Skewness* caracteriza a simetria da distribuição, e uma variável com distribuição univariada normalmente distribuída tem este índice igual à zero. *Kurtosis*, por sua vez, caracteriza a forma

da distribuição em termos de ser a distribuição normal achatada ou apresentar algum pico, relativa à distribuição normal (RAYCOV & MARKOULIDES, 2008). No caso de *skewness* positiva, há um acúmulo de casos à esquerda, e à direita a cauda é muito longa; com *skewness* negativa há um acúmulo de casos à direita, e à esquerda a cauda é muito longa. Valores de *kurtosis* acima de zero indicam uma distribuição que é muito pontiaguda com caudas curtas e espessas, enquanto os valores de *kurtosis* abaixo de zero indicam uma distribuição que é muito plana (também com muitos casos nas caudas) (TABACHNICK & FIDELL, 2007). A Figura 2.2 mostra os possíveis desvios da normalidade. Os autores Hair et al. (2010) e Tabachnick e Fidell (2007) sugerem a aplicação de um teste de significância para *skewness* e para *kurtosis* para testar o valor obtido numa hipótese nula igual a zero. Para tanto, calcula-se o desvio padrão para cada índice. O desvio padrão para *skewness* é aproximadamente

$$S_s = \sqrt{\frac{6}{N}} \quad \text{Equação 5}$$

Sendo S_s o desvio padrão e N o número de casos na amostra. O valor do índice obtido é comparado com zero utilizando a distribuição z , em que

$$z = \frac{S - 0}{S_s} \quad \text{Equação 6}$$

Sendo S o valor encontrado de *skewness*. Para *kurtosis* o desvio padrão é aproximadamente

$$S_k = \sqrt{\frac{24}{N}} \quad \text{Equação 7}$$

E, então, o valor encontrado de *kurtosis* é comparado com zero através da distribuição z , em que

$$z = \frac{K - 0}{S_k} \quad \text{Equação 8}$$

Se os valores de z ultrapassarem os valores críticos especificados, então a distribuição não é normal. O valor crítico da distribuição z depende do nível de significância escolhido que, preferencialmente, é de 1 ou 5% (HAIR et al., 2010; TABACHNICK & FIDELL, 2007).

Um teste estatístico usado para verificar a normalidade é o teste de Shapiro-Wilk W, que testa a hipótese nula de que uma amostra deriva de uma população normalmente distribuída. Se o teste for significativo, a hipótese de que a respectiva distribuição seja normal deve ser rejeitada. Em 1965 o teste de Shapiro-Wilk foi originalmente estruturado estritamente para amostras com menos de 50 indivíduos; porém, em 1995 foi aprimorado por Patrick Royston, tornando possível utilizá-lo para amostras de 3 a 5000 indivíduos (RAZALI, 2011). O software *STATISTICA*,

utilizado nesta dissertação, no entanto, indica a adequação do uso desse teste para amostras com até 2000 indivíduos (STATSOFT, 2011).

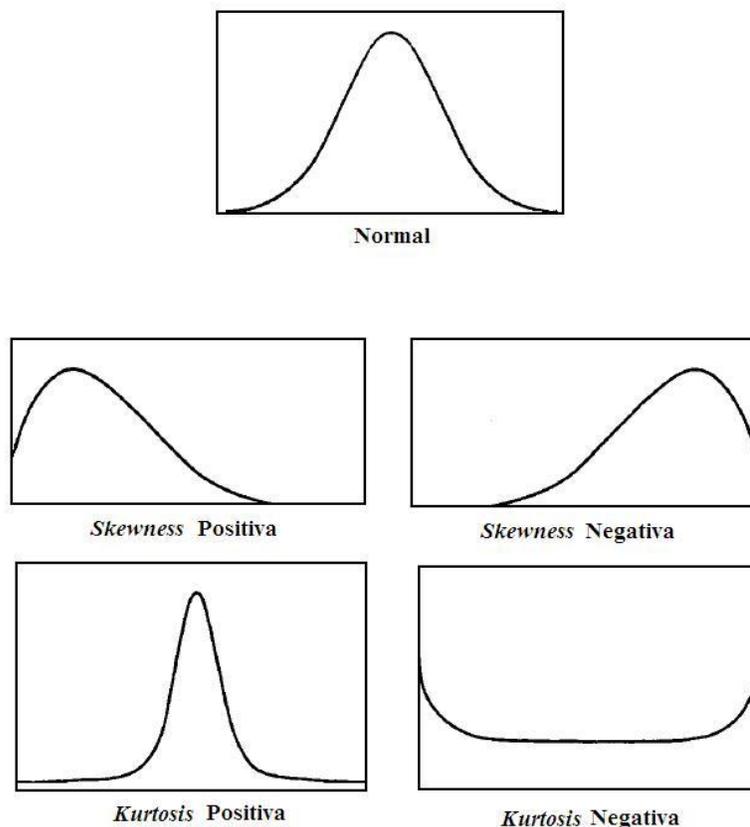


Figura 2.2 – Distribuição Normal, distribuição com *skewness* e distribuição com *kurtosis*.
Fonte: Tabachnick e Fidell (2007)

Para ajustar a não normalidade de uma variável, é considerável aplicar transformações tais como raiz quadrada, elevar os valores ao quadrado, aplicar logaritmo, etc. Porém, a aplicação dessas transformações dificulta a análise dos dados e dos resultados.

Alguns autores dão exemplo de como transformar variáveis, de acordo com seus desvios da normalidade. Tabachnick e Fidel (2007) e Stevens (2009) utilizam funções matemáticas para cada tipo de desvio, seja ele *kurtosis* e *skewness* positiva ou negativa.

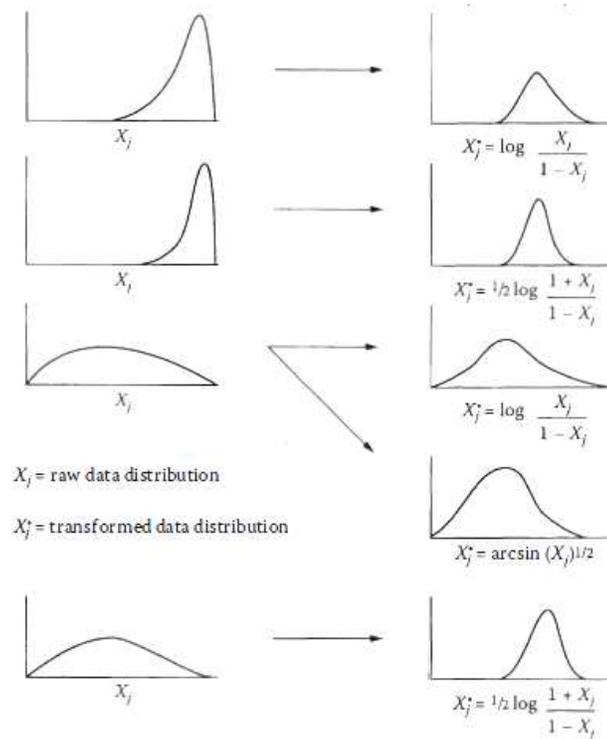


Figura 2.3 - Transformações da distribuição normal.

Fonte: Stevens (2009)

A Figura 2.3, de Stevens (2009), mostra algumas transformações mais complexas, especialmente quando há grande concentração de observações na extrema direita (*skewness* negativa e *kurtosis* positiva), e quando há distribuições abauladas.

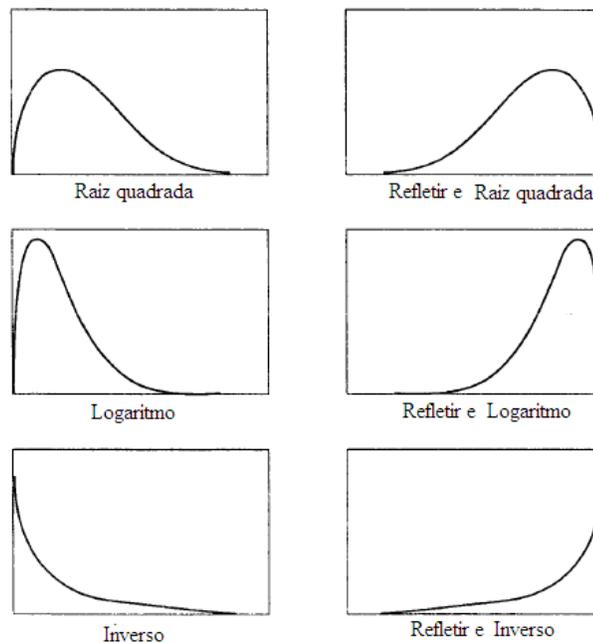


Figura 2.4 - Distribuições originais e suas transformações mais comuns.
 Fonte: Tabachnick e Fidell (2007)

Já Tabachnick e Fidell (2007) concentram-se em casos extremos de *skewness* e *kurtosis*, tanto positivos quanto negativos, como mostra a Figura 2.4

O procedimento chamado “refletir uma distribuição” corresponde a: encontra-se o maior valor observado de uma variável e somar um a esse valor; depois, subtraem-se os valores de cada observação dessa nova constante (TABACHNICK & FIDELL, 2007).

2.6 Linearidade

Supor linearidade é o mesmo que dizer que duas variáveis estão relacionadas através de uma linha reta. A linearidade é importante por uma questão prática, já que a aplicação do r de Pearson (índice de correlação) só capta a relação linear entre variáveis e, se existe alguma componente não linear nessa relação, o coeficiente ignorará (TABACHNICK & FIDELL, 2007).

A maneira mais comum de avaliar a linearidade é examinar a dispersão das variáveis e identificar qualquer padrão não linear nos dados. De acordo com Tabachnick e Fidel (2007), quando duas variáveis são lineares e normais, o gráfico resultante possui uma forma oval. Porém, em alguns casos a relação entre variáveis simplesmente não é linear. Para contornar o problema de não linearidade, Hair et al. (2010) consideram novamente a transformação das variáveis, ou a criação de novas variáveis para representar a porção não linear dessa relação.

2.7 Homoscedasticidade

A hipótese de homoscedasticidade implica que a variância dos valores para uma variável contínua deve ser basicamente a mesma para os valores das outras variáveis. A homoscedasticidade está relacionada com a hipótese de normalidade, pois quando a normalidade é atingida, a relação entre as variáveis se torna homoscedástica (TABACHNICK & FIDELL, 2007). A heteroscedasticidade, o oposto de homoscedasticidade, pode vir, principalmente, de duas fontes: tipo de variáveis, já que muitas variáveis têm a tendência natural de diferenças na sua dispersão, ou distribuição enviesada de uma ou duas variáveis. A Figura 2.5 mostra esta dispersão desigual entre duas variáveis.

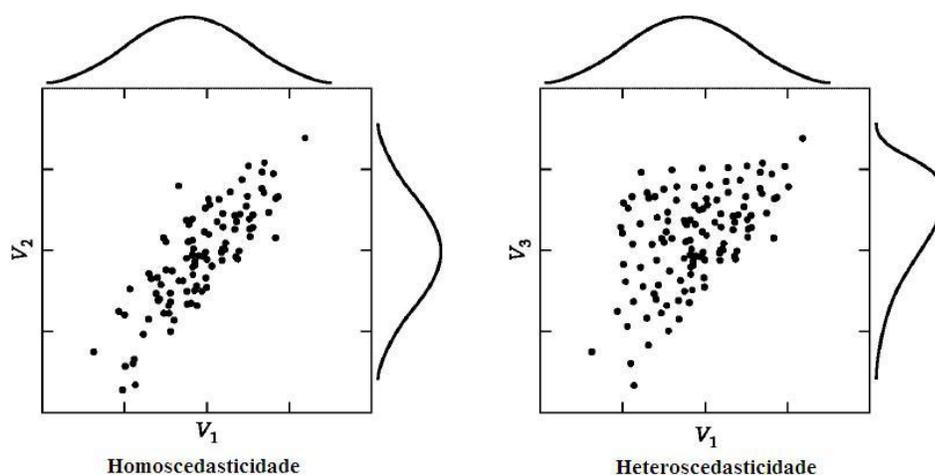


Figura 2.5 – Dispersão das relações homoscedásticas e heteroscedásticas.

Fonte: Hair et al. (2010)

Existem testes estatísticos para avaliar se uma variável é homoscedástica. O mais comum é o teste Levene, usado para analisar se as variâncias de uma variável são iguais dentro de um número qualquer de grupos (HAIR et al., 2010).

A heteroscedasticidade também pode ser remediada usando-se transformação dos dados, assim como normalidade.

2.8 Multicolinearidade e singularidade

Multicolinearidade e singularidade são problemas que ocorrem quando variáveis são linearmente relacionadas. Com a multicolinearidade, as variáveis são correlacionadas e, com singularidade, as variáveis são redundantes, ou seja, uma variável é a combinação de duas ou mais variáveis. Quando as variáveis são multicolineares ou singulares, elas contêm informações tautológicas e não são necessárias na mesma análise. Em outras palavras, existem menos variáveis que

deveriam aparecer na análise (TABACHNICK & FIDELL, 2007). A multicolinearidade pode ser verificada analisando-se a matriz de correlação. O coeficiente de correlação fornece um retrato instantâneo de quão próximos dois parâmetros variam juntos. O coeficiente de correlação pode ser expresso em termos de suas observações originais, de acordo com a equação 9 (WONNACOTT & WONNACOTT, 1978).

$$r = \frac{\sum (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}} \quad \text{Equação 9}$$

No entanto, é importante perceber que nem sempre a correlação indica causa e efeito. O fato de evoluírem no mesmo passo pode ser consequência da influência de uma terceira variável (WONNACOTT & WONNACOTT, 1981). A correlação mais utilizada é a correlação de Pearson, que varia de -1 a +1 (STATSOFT, 2012).

2.9 Análise de Cluster

Na análise de cluster o objetivo é encontrar padrões nos dados para agrupar as observações da amostra em *clusters* ótimos, nos quais essas observações sejam similares, mas os grupos sejam tão diferentes quanto possível. Os grupos são chamados de naturais e deve fazer sentido no seu resultado final (RENCHER, 2002). Seja para ser utilizada como uma etapa preliminar de outra análise, ela pode ser aplicada em diversos campos do conhecimento, como psicologia e ciências sociais, biologia, estatística, etc. (TAN et al., 2006). Tan et al. (2006) dividem os métodos de *clustering* em hierárquico ou divisivo (*partitional*, em inglês), exclusivo, ou sobreposto ou *fuzzy*. Quando um método de cluster é hierárquico, os cluster possuem *sub clusters*, e quando se trata de um método divisivo, não existe a possibilidade de existir *sub clusters*, ou seja, o cluster é a menor forma de divisão dos dados. O cluster hierárquico pode ser visto como um conjunto de clusters divisivos, e vice-versa. Já os clusters exclusivos ou sobrepostos estão relacionados com a classificação dos indivíduos da amostra. Se um indivíduo é classificado apenas em um cluster, este é chamado de exclusivo, e se é possível classificá-lo em mais de um cluster, normalmente hierárquicos, o método é chamado de sobreposto. O método é chamado de *fuzzy* quando todos os indivíduos podem ser classificados em todos os clusters, porém com pesos que variam de 0 a 1, sendo 0 não pertencente ao grupo e 1 completamente pertencente ao grupo.

Dos métodos de cálculo e de formações dos grupos, *joining* ou *tree clustering*, *two-way joining* e *K-means clustering*, somente o último foi aplicado neste trabalho e, por isso será tratado aqui. Na técnica de cluster *K-means* primeiramente escolhem-se *K* centroides, normalmente a média dos indivíduos da amostra, sendo *K* o número de centroides especificado pelo pesquisador ou usuário. Os pontos da amostra são então designados aos grupos cujos centroides estão mais próximos, e um novo centroide é calculado. Os centroides são calculados até que nenhum indivíduo seja realocado de grupo.

A distância euclidiana é utilizada para designar os indivíduos a algum dos K grupos e, para medir a qualidade do cluster, é utilizada uma medida de Soma Quadrática dos Erros (SQE), representada na equação 10.

$$SQE = \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2 \quad \text{Equação 10}$$

Sendo c_i o centroide do cluster C_i e $dist$ a distância euclidiana entre a centroide e o indivíduo da amostra. Segundo Tan et al.(2006), a centroide que minimiza a SQE é a média dos indivíduos no espaço euclidiano. Uma maneira de se pensar no plano euclidiano é como um conjunto de pontos que satisfazem a determinadas relações, expressas em termos de distância e de ângulo (neste caso, trata-se apenas de distância). Cada ponto é representado no espaço euclidiano pela equação 11, ou seja, a raiz quadrada do quadrado dos pontos de cada indivíduo, no caso deste trabalho, o valor dos indicadores estudados para cada município.

$$\sqrt{x_1^2 + x_2^2 + \dots + x_i^2} \quad \text{Equação 11}$$

2.10 Análise Discriminante

A análise discriminante, segundo Rencher (2002), tem duas finalidades. Uma delas é a descrição da separação de grupos, em que funções lineares das variáveis são usadas para descrever as diferenças entre dois ou mais grupos. O autor chama este tipo de análise de discriminante descritiva, que inclui identificar a contribuição de uma variável para a separação dos grupos. Tabachnick e Fidell (2007), no entanto, utilizam a análise discriminante para a classificação em cada grupo. A atribuição de observações em grupos é feita por funções lineares ou quadráticas das variáveis (funções de classificação) e são usadas para designar um indivíduo da amostra a um dos grupos. Os valores medidos no vetor de observação para um indivíduo são avaliados pelas funções de classificação para encontrar o grupo mais provável ao qual pertence o indivíduo.

Para determinar a função determinante para dois grupos, assume-se que as duas populações a serem comparadas possuem a mesma matriz de covariância, mas vetores de médias distintos. A determinação da função é feita com as amostras $y_{11}, y_{12}, \dots, y_{1n_1}$ e $y_{21}, y_{22}, \dots, y_{2n_2}$ das duas populações. Cada um destes vetores é composto por medidas de p variáveis. A função discriminante é uma combinação linear dessas p variáveis, que maximiza as distâncias dos vetores das médias dos grupos. Uma combinação linear $z = \mathbf{a}'\mathbf{y}$ transforma cada vetor de observações em um escalar, como mostram as equações 12 e 13.

$$z_{1i} = \mathbf{a}'\mathbf{y}_{1i} = a_1y_{1i1} + a_2y_{1i2} + \dots + a_p y_{1ip} \quad i=1,2,\dots,n_1 \quad \text{Equação 12}$$

$$z_{2i} = \mathbf{a}'\mathbf{y}_{2i} = a_1y_{2i1} + a_2y_{2i2} + \dots + a_p y_{2ip} \quad i=1,2,\dots,n_2 \quad \text{Equação 13}$$

Com isso, os vetores de medições das p variáveis de cada indivíduo da amostra, neste caso os municípios, são agregados formando n_1+n_2 escalares. As médias dos escalares para cada observação j de cada grupo i é encontrada, e depois comparada para encontrar a máxima diferença padronizada (*standardized*). A média de cada escalar é representada pela equação 14.

$$\bar{Z}_i = \sum_{j=1}^{n_i} \frac{z_{ij}}{n_i} \quad i=1,2 \quad \text{Equação 14}$$

Por Rencher (2002) a média pode ser calculada como $\mathbf{a}'\bar{\mathbf{y}}_i$, com $\bar{\mathbf{y}}_i = \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n_i$, com $i=1,2$. Destas duas igualdades, pode-se encontrar o vetor \mathbf{a} que maximiza a diferença e $(\bar{Z}_1 - \bar{Z}_2)^2 / s_z^2$. A potência ao quadrado é para evitar valores negativos. Portanto, conclui-se da equação 15,

$$\frac{(\bar{Z}_1 - \bar{Z}_2)^2}{s_z^2} = \frac{[\mathbf{a}'(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)]^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} \quad \text{Equação 15}$$

em que \mathbf{S} é a matriz covariância formada pelas p variáveis, que o vetor \mathbf{a} que maximiza a diferença é $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$. O vetor \mathbf{a} será um vetor com p elementos da função discriminante. Estes elementos são escolhidos pelo software *STATISTICA*, empregado neste trabalho, utilizando o método *forward stepwise analysis*. O método refere-se ao modo como as variáveis são incluídas na função, de acordo com seu lambda e lambda parcial de *Wilk*. Resumidamente, o lambda parcial de *Wilk* é usado para identificar o poder de discriminação da variável entre os grupos, sendo 0 poder discriminatório perfeito e 1 nenhum poder discriminatório. O lambda de *Wilk* também possui a mesma função, ou seja, identificar o poder discriminatório da variável, porém, o lambda parcial de *Wilk* está relacionado com o poder da variável sozinha, sem a presença de outras variáveis no modelo, enquanto o lambda de *Wilk* está relacionado com a presença de outras variáveis (STATSOFT, 2011). A descrição matemática de lambda parcial de *Wilk* e lambda de *Wilk* pode ser encontrada em Rencher (p. 161, 2002). No primeiro passo do método, os Lambdas parciais de *Wilk* de cada variável são calculados e o menor deles é escolhido para entrar no modelo. O segundo passo é calcular o valor de lambda para as $p-1$ variáveis restantes. Após escolher o menor lambda, ele é novamente calculado para cada $p-2$ variáveis restantes, e assim por diante. Entre cada seleção de variável, a variável previamente incluída no modelo é reavaliada para comprovar sua contribuição para discriminação entre os grupos utilizando o lambda de *Wilk*. A variável com maior lambda será retirada do modelo (RENCHE, 2002).

Outro tipo de função que compõe a análise discriminante é a função de classificação, que, a partir das variáveis escolhidas pelo método *forward stepwise*, classificam o indivíduo da amostra em um dos dois grupos. Quando o indivíduo deve ser classificado, o procedimento utilizado nesta dissertação pode ser visto em Rencher (p. 300, 2002) e Tabachnick e Fidell (p. 387, 2007). As matrizes de covariância das duas populações devem ser iguais, porém normalidade não é necessária. A classificação é feita pelo resultado da multiplicação dos coeficientes da equação de classificação pelos valores de cada variável de cada indivíduo. A equação que gerar o maior

score classificará o indivíduo ao seu grupo. O vetor de coeficientes da equação de classificação é encontrado multiplicando-se a matriz covariância interna aos grupos por um vetor de médias de cada variável (TABACHNICK & FIDELL, 2007).

3. Resultados

Para analisar a relação entre qualidade de vida, que se entende pode ser representada pelos oito indicadores apresentados, e a produção de cana-de-açúcar nos estados de São Paulo e Alagoas, foram feitos os seguintes procedimentos:

- Comparação entre os valores médios dos indicadores dos grupos de municípios canavieiros e não canavieiros, nos estados de São Paulo e Alagoas, e análise dos resultados;
- Classificação dos municípios canavieiros em *clusters* (ou grupos), tendo por base os indicadores de qualidade de vida (bem-estar);
- Emprego da análise discriminante utilizando indicadores econômicos, de finanças públicas e indicadores relativos à produção de cana-de-açúcar nos municípios objeto de análise dos dois estados.

Essas análises complementam aquelas feitas por Oliveira (2011), inclusive no que se diz respeito à busca de maior rigor estatístico para a utilização de testes e métodos de análise. As transformações feitas aos dados originais para corrigir problemas de normalidade, homogeneidade de variância e linearidade são apresentadas resumidamente em cada seção.

3.1 Comparação das médias

A análise dos impactos da atividade canavieira sobre os indicadores de qualidade de vida dos municípios nas quais a mesma é significativa foi feita, em um primeiro estágio, pela comparação das médias dos valores dos indicadores entre os municípios não canavieiros e canavieiros; o teste estatístico *t*, apresentado na seção de revisão metodológica, foi empregado para se verificar discernimento na diferenciação dos grupos. É importante lembrar que os valores são transformados, e sua comparação deve ser feita levando em conta as consequências da aplicação das transformações feitas no apêndice A na hora de concluir qual é média é maior.

Na comparação entre as médias dos indicadores dos municípios canavieiros e não canavieiros, os municípios foram divididos em dois grupos; no grupo de municípios canavieiros, em um dado ano, foram alocados os produtores de cana-de-açúcar que somam 90% da produção total do estado. Já o grupo dos não canavieiros é formado pelos municípios restantes, excluindo aqueles que não tinham, naquele ano, população dentro da faixa de população máxima e mínima dos municípios canavieiros.

O procedimento de exclusão justifica-se pelo fato de que municípios muito grandes ou muito pequenos em população podem apresentar indicadores significativamente distintos e, assim, evita-se distorção na análise. Da observação dos valores numéricos dos indicadores, nota-se que os mesmos têm algum grau de correlação com a população.

Além disso, os grupos foram definidos para a mesma faixa populacional, para evitar viés devido ao número de habitantes de cada município. Por isso, foram testados, dentro dos municípios não-canavieiros, apenas os municípios com população comparável com o grupo de canavieiros.

3.1.1 São Paulo

Em 1970 havia 539 municípios no estado de São Paulo, sendo destes 91 considerados canavieiros, pelo método de alocação mencionado.

No apêndice A são apresentadas as transformações empregadas com vistas a atingir os pré requisitos de aplicação do teste de t utilizado nesta dissertação (foi aplicada a variante do teste t que requer homogeneidade de variância). Na Tabela 3.1 são apresentados os resultados da comparação dos grupos de municípios canavieiros e não canavieiros, para os oito indicadores considerados, tendo por base o teste t para verificação do discernimento estatístico na afirmação de eventual diferença entre os grupos. O teste t permite explorar a hipótese de que os grupos de municípios canavieiros e não canavieiros são diferentes entre si, para um dado indicador. A hipótese de diferenciação entre os grupos é rejeitada se o teste for não significativo.

O teste de hipóteses t foi aplicado para um nível de significância de 5%. Os valores marcados em verde na Tabela 3.1 são os melhores quando da comparação e, para os critérios adotados nesta dissertação, conclui-se que são estatisticamente diferentes. Nota-se que, em função da transformação aplicada nesta dissertação, não é possível a simples comparação dos valores numéricos apresentados na tabela com aqueles apresentados por Oliveira (2011). Como ilustração, na Tabela 3.2 são comparados os resultados obtidos da análise dos indicadores, de São Paulo, em 1970, nesta dissertação e em Oliveira (2011). No Anexo I são apresentados os resultados dos valores médios e desvios padrão dos indicadores, em todos os anos, sem as transformações, resultado do trabalho de Oliveira (2011).

A conclusão é que, da mesma forma como concluído no estudo de Oliveira (2011), mas agora com maior rigor estatístico, os municípios canavieiros têm, na média, melhores indicadores do que os municípios não canavieiros, comparados municípios com a mesma faixa de população. No estado de São Paulo, a conclusão é válida para todos os oito indicadores considerados e para todos os anos analisados. É importante destacar que, mesmo com a transformação, a maioria das variáveis não possui distribuição normal, o que pode levar a falsas interpretações dos resultados, principalmente das variáveis que não satisfazem as suposições para aplicação do teste t .

Tabela 3.1 - Resultados do teste *t* para o estado de São Paulo – 1970 a 2000

Variável	1970				1980			
	Com cana	Sem cana	<i>t</i>	<i>p</i>	Com cana	Sem cana	<i>t</i>	<i>p</i>
L de Theil	0,57±0,07	0,61±0,07	4,980	0	-0,41±0,11	-0,37±0,120	3,710	0
Pessoas pobres	5,98±0,95	4,70±1,46	-8,004	0	4,02±0,97	5,47±1,24	11,580	0
Mort. infantil	1,89±0,07	1,95±0,09	6,148	0	7,28±0,79	7,64±0,84	4,094	0
Analfabetos	4,92±0,58	5,45±0,72	6,485	0	4,41±0,56	4,72±0,63	4,725	0
IDH	0,71±0,04	0,66±0,05	-7,554	0	0,02±0,01	0,04±0,02	7,772	0
Esperança de vida	2,6±0,44	2,95±0,51	6,157	0	7,76±0,15	7,69±0,16	-4,112	0
Ilumin. Elétrica	0,94±0,16	0,67±0,22	-10,629	0	0,02±0,02	0,07±0,05	9,088	0
Inst. sanitárias	0,54±0,25	0,28±0,26	-8,512	0	0,53±0,20	0,27±0,23	-10,969	0

Variável	1991				2000			
	Com cana	Sem cana	<i>t</i>	<i>p</i>	Com cana	Sem cana	<i>t</i>	<i>p</i>
L de Theil	0,63±0,06	0,69±0,07	-7,306	0	-0,35±0,08	-0,331±0,080	-2,994	0,002
Pessoas pobres	3,89±0,91	4,83±1,24	-8,017	0	3,93±0,65	4,670±1,014	-9,111	0
Mort. infantil	4,89±0,57	5,36±0,71	-6,866	0	1,12±0,11	1,186±0,138	-5,569	0
Analfabetos	3,71±0,46	3,82±0,57	-2,009	0	3,13±0,41	3,373±0,552	-5,384	0
IDH	0,03±0,01	0,04±0,02	-4,562	0	0,89±0,01	0,879±0,020	6,725	0
Esperança de vida	2,19±0,48	2,55±0,51	-7,045	0	2,54±0,41	2,767±0,484	-5,515	0
Ilumin. Elétrica	1,00±0,01	1,01±0,03	-3,973	0	0,99±0,00	0,983±0,032	4,515	0
Inst. sanitárias	1,09±0,06	1,18±0,09	-9,878	0	1,06±0,04	1,157±0,091	-13,100	0

Para os quatro anos estudados, apenas a variável “Analfabetos” se adequou à normalidade após a transformação, em todos os anos, e o mesmo com as variáveis “IDH” em 1970 e “Esperança de vida” em 1991. A conclusão baseia-se na interpretação do parâmetro lambda de Wilki, que se encontra no apêndice A.

A Tabela 3.2 mostra um exemplo de comparação entre o trabalho feito por Oliveira (2011) e esta dissertação. Não há diferenças nas conclusões para o ano apresentado.

Tabela 3.2- Resultados do teste *t* para indicadores de 1970 no estado de São Paulo

Variável	Após transformação		Oliveira (2011)	
	Canavieiros	Não canavieiros	Canavieiros	Não canavieiros
L de Theil	0,570±0,079	0,614±0,075	0,334 ± 0,090	0,383 ± 0,092
Pessoas pobres	5,985±0,951	4,704±1,465	57,91 ± 11,35	70,2 ± 14,3
Mortalidade infantil	1,894±0,074	1,957±0,092	79,73 ± 13,89	92,93 ± 20,04
Analfabetos	4,925±0,582	5,450±0,725	24,7 ± 5,8	30,2 ± 8,0
IDH	0,711±0,040	0,668±0,052	0,508 ± 0,058	0,449 ± 0,071
Esperança de vida	2,600±0,444	2,958±0,517	55,45 ± 2,29	53,40 ± 3,06
Iluminação elétrica	0,944±0,167	0,676±0,227	74,5 ± 15,0	47,5 ± 22,9
Instalações sanitárias	0,543±0,257	0,281±0,269	35,3 ± 21,1	15,2 ± 17,9

Segundo Stevens (2009), a não normalidade dos valores tem pequeno impacto sobre o erro tipo I, que é a rejeição de uma hipótese nula quando ela é verdadeira. Ou seja, no caso tratado desta seção, erro tipo I implicaria aceitar a existência de uma diferença entre os grupos enquanto não existe tal diferença. Portanto, rigorosamente não é possível afirmar que, exceto no caso do indicador “Analfabetos”, no caso do estado de São Paulo, os municípios canavieiros têm melhores indicadores de qualidade de vida do que os municípios similares em população, e não canavieiros. Por outro lado, com maior rigor do que a conclusão apresentada por Oliveira (2011), pode-se afirmar que há fortes indícios de que os municípios canavieiros de São Paulo têm, na média, melhores indicadores do que os não canavieiros. E isso em todo o período analisado.

Por ser bastante conservadora, uma conclusão mais segura, como apresentado em Oliveira (2011), é que a atividade canavieira, considerados os oito indicadores analisados nesta dissertação, não traz desvantagens do ponto de vista da qualidade de vida, comparados com resultados dos municípios em que a atividade canavieira é irrelevante ou inexistente.

3.1.2 Alagoas

Em 1970 o estado de Alagoas tinha 94 municípios, sendo 70 não canavieiros e 24 canavieiros, de acordo com o método escolhido nesta dissertação para a definição dos grupos. Porém, Maceió, que estaria entre os municípios canavieiros, foi excluído da análise, para seguir a mesma metodologia usada por Oliveira (2011) e evitar distorções na análise comparativa; por ser a capital do estado, e ter várias outras atividades econômicas relevantes, a comparação entre os grupos seria distorcida.

Seguindo procedimento anteriormente descrito, os grupos de municípios canavieiros e não canavieiros foram comparados, considerados os oito indicadores analisados, e o teste t foi aplicado para se buscar maior rigor estatístico na análise e nas conclusões. Os resultados são apresentados na Tabela 3.3, sendo que em verde estão marcados os parâmetros que são estatisticamente melhores quando da comparação entre os grupos. No Anexo I são apresentados os resultados dos valores médios e desvios padrão dos indicadores, em todos os anos, sem as transformações (resultado do trabalho de Oliveira(2011)).

Tabela 3.3 - Resultados do teste *t* para o estado de Alagoas – 1970 a 2000

Variável	1970				1980			
	Com cana	Sem cana	<i>t</i>	<i>p</i>	Com cana	Sem cana	<i>t</i>	<i>p</i>
L de Theil	-0,50±0,08	-0,49±0,13	0,083	0,933	0,56±0,06	0,61±0,11	-1,495	0,138
Pessoas pobres	0,97±0,16	0,62±0,31	-5,416	0	4,96±0,62	3,49±1,09	6,469	0
Mort. infantil	194,04±10,97	208,46±32,94	2,284	0,024	155,43±19,88	163,23±30,93	-1,293	0,198
Analfabetos	4,05±1,06	4,21±0,95	0,115	0,908	4,31±0,72	4,38±1,10	0,233	0,816
IDH	-0,58±0,04	-0,63±0,06	-3,816	0	1,06±0,02	1,07±0,02	6,469	0
Esperança de vida	48,43±1,12	47,09±3,31	-2,143	0,034	52,62±2,25	51,82±3,41	1,201	0,232
Ilum. elétrica	-0,83±0,29	-1,087±0,33	-3,637	0	-0,42±0,16	-0,65±0,24	4,474	0
Inst. sanitárias	0±0	0,01±0,06	0,016	0,986	0,03±0,08	0,01±0,08	1,434	0,154

Variável	1991				2000			
	Com cana	Sem cana	<i>t</i>	<i>p</i>	Com cana	Sem cana	<i>t</i>	<i>p</i>
L de Theil	0,61±0,08	0,65±0,07	2,042	0,043	-0,32±0,07	-0,25±0,09	3,375	0,001
Pessoas pobres	1,19±0,18	1,05±0,26	-2,878	0,004	3,98±0,82	3,64±0,98	-2,116	0,036
Mort. infantil	75,77±12,09	85,28±16,34	2,998	0,003	1,70±0,10	1,70±0,10	0,374	0,709
Analfabetos	4,30±1,02	4,22±1,00	-0,834	0,406	3,56±0,91	3,27±0,95	-1,785	0,077
IDH	0,69±0,02	0,68±0,03	-2,153	0,033	0,59±0,05	0,58±0,04	-1,238	0,218
Esperança de vida	7,58±0,15	7,47±0,20	-2,950	0,003	2,53±0,60	2,57±0,59	0,365	0,715
Ilum. elétrica	1,09±0,06	1,16±0,07	3,606	0	1,05±0,03	1,08±0,05	2,090	0,039
Inst. sanitárias	0,01±0,06	0,01±0,06	-0,941	0,349	-1,40±0,88	-1,96±0,84	-3,131	0,002

Como ilustração, na Tabela 3.4 são comparados os resultados obtidos da análise dos indicadores, de Alagoas, em 1970, nesta dissertação e por Oliveira (2011).

Tabela 3.4 - Resultados do teste *t* para 1970 no estado de Alagoas

Variável	Com transformação		Oliveira (2011)	
	Canavieiros	Não canavieiros	Canavieiros	Não canavieiros
L de Theil	-0,504±0,082	-0,491±0,130	0,314 ± 0,054	0,342 ± 0,118
Pessoas pobres	0,973±0,160	0,625±0,315	89,76 ± 2,88	95,60 ± 3,36
Mortalidade infantil	194,040±10,971	208,466±32,941	191,93 ± 14,60	207,85 ± 29,49
Analfabetos	4,050±1,064	4,218±0,950	69,13 ± 7,93	69,90 ± 7,72
IDH	-0,585±0,045	-0,636±0,067	0,265 ± 0,025	0,227 ± 0,035
Esperança de vida	48,437±1,119	47,091±3,310	48,67 ± 1,54	0,342 ± 0,118
Iluminação elétrica	-0,831±0,298	-1,087±0,331	71,39 ± 13,28	10,01 ± 9,93
Instalações sanitárias	0±0	0,010±0,062	0,0 ± 0,0	0,62 ± 3,04

O teste *t* para a variável “L de theil”, em 1980, não permite a conclusão de que os grupos de municípios canavieiros e não canavieiros são estatisticamente distintos quanto à distribuição de renda. Apesar de a distribuição da variável continuar não normal mesmo após a transformação, houve melhora no índice de *kurtosis* (assim como no caso da variável “IDH”), o que é importante para conclusões mais seguras em função dos resultados do teste estatístico *t*, como explica Statsoft (2011).

Os resultados do teste de hipótese *t* para o ano de 1991 indicam diferenças em relação às conclusões de Oliveira (2011). Após as transformações, conclui-se que estatisticamente não há diferença entre os grupos quanto à variável “IDH”.

As variáveis “Mortalidade infantil”, “Esperança de vida” e “Analfabetos” atingiram a condição de normalidade, em todos os anos. Para os anos de 1980, 1991 e 2000 há diferenças entre os resultados de Oliveira (2011) e os resultados obtidos com a análise das variáveis transformadas, nesta dissertação. Para o ano de 2000, para quatro indicadores (“Mortalidade Infantil”, “Analfabetos”, “IDH” e “Esperança de vida”) não é possível afirmar que há diferença entre os grupos, consideradas as variáveis transformadas.

No estado de Alagoas, apesar de se atingir a condição de distribuição normal para grande parte dos indicadores e anos, o que diminui a probabilidade de se cometer o erro de tipo I (isto é, afirmar que os grupos são diferentes quando na realidade não o são), o baixo número de indivíduos na amostra diminui o poder (*power*, em inglês) do teste estatístico, que está associado

à confiança estatística de se fazer a interpretação correta, ou seja, concluir que os grupos são realmente diferentes.

Stevens (2009) indica que o poder do teste depende do nível de significância adotado, do tamanho da amostra e do tamanho do efeito (*effect size*, em inglês), sendo que tal aspecto representa quanto os grupos se diferem em população em relação às variáveis estudadas. O autor mostra que, para amostras com menos de 20 indivíduos deve-se escolher níveis de significância mais liberais, como 10 ou 15%. Como nesta dissertação adotou-se nível 5% e têm-se poucos indivíduos na amostra (mas mais do que 20), deve-se chamar a atenção do leitor para as conclusões, que podem ser conservadoras. Ou seja, a rigor, a vantagem dos municípios canavieiros, do ponto de vista de alguns parâmetros, e para alguns anos, talvez não exista.

Além disso, apesar do teste de Shapiro-Wilk ter sido criado para amostras menores que 50 indivíduos, Razali (2011) testou-o, entre outros testes de normalidade, e concluiu que nenhum teste de normalidade tem bom desempenho quando a amostra é menor que 30 (é o caso de Alagoas). E mais, o autor indica que os resultados dos testes devem ser analisados em conjunto com os valores de *skewness* e *kurtosis*.

No caso de Alagoas, pode-se afirmar que, em geral, o grupo de municípios canavieiros apresenta melhores indicadores que o grupo de não canavieiros. Diferentemente de São Paulo, em Alagoas há maior similaridade entre os grupos e o teste *t* não indica, exceto para os indicadores “Pessoas Pobres” e “Iluminação elétrica”, diferenças estatisticamente significativas, de forma contínua, no período analisado. Ou seja, para Alagoas, não é possível afirmar que há superioridade do grupo dos municípios canavieiros em relação aos indicadores sociais. Por outro lado, tampouco é possível a conclusão oposta, isto é, de que os municípios canavieiros têm piores indicadores do que os municípios não canavieiros de mesmo porte.

3.2 Evolução do conjunto de indicadores

Concluído que os municípios canavieiros apresentam, em geral, e na média, melhores indicadores do que os municípios não canavieiros de mesmo porte, é importante analisar como tem sido a evolução dos indicadores, e se a vantagem dos municípios canavieiros tem aumentado ou reduzido ao longo dos anos.

Para analisar a evolução do conjunto dos indicadores nos dois grupos de municípios objeto de comparação, foram feitos diagramas “radar”, por estado e por ano. A seguir são apresentados os diagramas para os estados de São Paulo e Alagoas. Em função das distintas escalas dos vários indicadores, e para a representação de todos os resultados em um único diagrama, foi necessária a criação de um procedimento de normalização. Para isso, as médias dos parâmetros para os grupos de municípios com cana e sem cana foram normalizadas em uma escala que varia entre 0 e 1. Ao melhor valor do indicador verificado no País, entre os anos estudados, se atribuiu o valor unitário, e ao pior resultado o valor nulo. Fez-se uma escala linear, e por simples proporção foi calculado o

valor normalizado do valor médio do indicador nos dois grupos de municípios, em cada ano. Nos gráficos apresentados a seguir, os resultados são tanto melhores quanto mais distantes os parâmetros estiverem do centro do diagrama.

Na Figura 3.1 é apresentada a evolução dos indicadores dos municípios canavieiros e dos não canavieiros no estado de São Paulo. Da análise das médias dos valores dos indicadores, pode-se identificar uma tendência de aproximação dos dois grupos, com exceção do indicador “L de Theil”, para o qual a aproximação se dá apenas no último ano estudado. Para todos os indicadores, a aproximação entre os grupos é mais clara em 2000. A evolução dos indicadores IDH e L de Theil podem ser observadas geograficamente no apêndice F.

No caso do estado de São Paulo, apesar da tendência de aproximação dos grupos, os municípios canavieiros têm sempre os melhores resultados em relação aos municípios não canavieiros. Com o tempo, a menor diferença entre os grupos indica perda da vantagem relativa dos municípios canavieiros. As informações disponíveis não permitem uma conclusão sobre as razões da tendência observada.

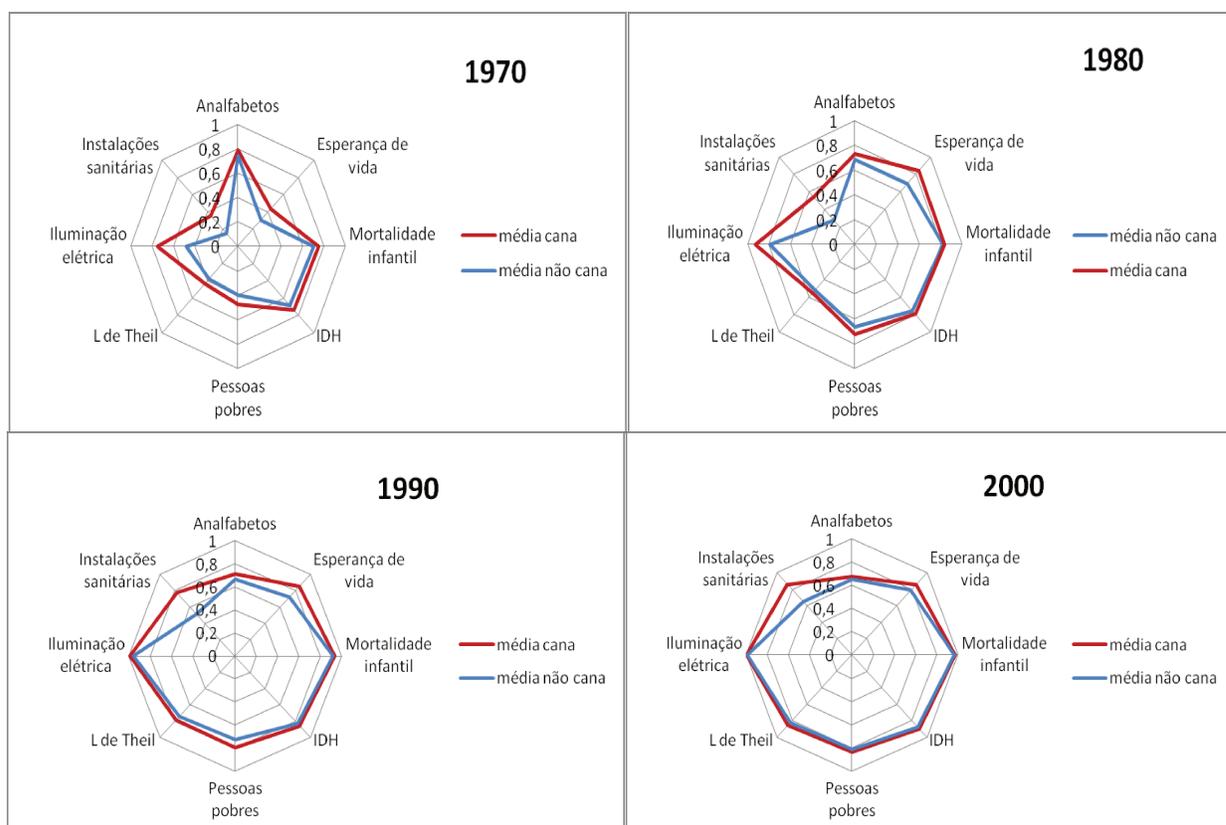


Figura 3.1 – Evolução dos indicadores no estado de São Paulo de 1970-2000.

No estado de Alagoas também houve evolução dos indicadores para os dois grupos, mas tomando como referência (para a normalização) os melhores resultados para o país, os gráficos para o

estado chamam a atenção pelos maus resultados. Em particular, chama atenção a pequena melhora no indicador “Domicílios com acesso a saneamento” Ao longo das quatro décadas, houve aproximação entre os grupos para os indicadores “Pessoas pobres”, “Mortalidade infantil”, “IDH”, “Esperança de vida” e “Iluminação elétrica”. Para os outros três indicadores a tendência de aproximação não é clara. A Figura 3.2 mostra os gráficos radar para o estado de Alagoas.

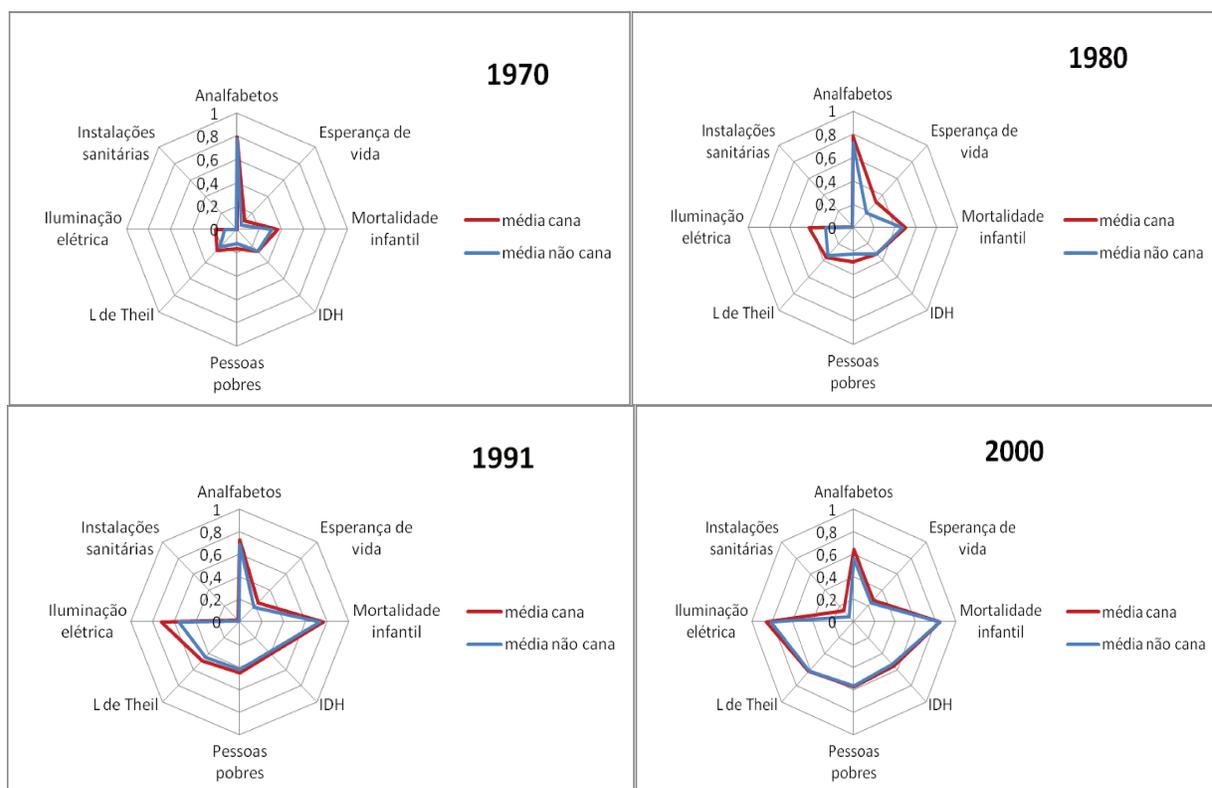


Figura 3.2 - Evolução dos indicadores no estado de Alagoas de 1970-2000.

Para a análise do impacto da entrada da cana-de-açúcar nos municípios do estado de São Paulo, ou da expansão da atividade canieira, foram definidos dois grupos de municípios, como descrito abaixo:

- Grupo dos municípios não canieiros: Aqueles municípios que nunca estiveram no grupo dos municípios canieiros, ou seja, aqueles que desde 1970 estiveram no grupo dos não canieiros e não mudaram de grupo nos anos seguintes;
- Grupo dos municípios canieiros: Grupo dos municípios que não eram canieiros na década de 1970 e passaram a ser classificados como canieiros em 1980 e permaneceram assim até 2000. Ou seja, houve introdução ou expansão da atividade canieira no município entre 1970 e 1980, quando da primeira expansão da produção de etanol no país, e tal condição permaneceu até 2000.

- O grupo de municípios não canavieiros foi ajustado para que a faixa de população do grupo objeto de comparação seja a mesma. Para tanto, a faixa de população em 1970 foi utilizada.

Assim, pode-se analisar se a introdução ou a expansão da atividade canvieira no município foi benéfica do ponto de vista dos indicadores empregados neste trabalho,. Esta foi feita apenas para o estado de São Paulo porque não houve expansão significativa em Alagoas, em novos municípios.

Analisando primeiramente a média dos indicadores em cada grupo no ano de 1970, percebe-se que o grupo de municípios canavieiros já possuía resultados melhores para todos os indicadores, ou seja, no ano de 1970, quando esse grupo de municípios não tinha produção de cana-de-açúcar expressiva, os indicadores já eram melhores. No período 1980-2000, após a expansão ou entrada da cana, a tendência é de aproximação dos dois grupos. A Figura X mostra a evolução das médias dos indicadores dos dois grupos no período 1980-2000.

Isto pode ser comprovado pelo teste de hipótese não paramétrico aplicado para esta análise, cujos resultados são apresentados no apêndice E. Tabela 3.5 mostra um exemplo do teste não paramétrico de Mann-Whitney, que identifica se os dois grupos são diferentes entre si. O valor de p mostra o valor do nível de significância. Pode-se observar que os dois grupos têm as médias dos indicadores significativamente diferentes no ano de 1970, com exceção de “L de Theil”. Quando se analisa a evolução dos indicadores individualmente, constata-se que os indicadores “Esperança de vida”, “Mortalidade infantil”, “IDH”, “Iluminação elétrica” e “Instalações sanitárias” tendem a se aproximar, enquanto não se percebe uma tendência clara para o indicador “Analfabetos”. Já para os indicadores “Pessoas pobres” e “L de Theil”, os grupos tendem a se diferenciar, ou seja, o grupo de municípios novos canavieiros tem um crescimento do parâmetro normalizado maior que o grupo dos não canavieiros (Lembrar que quanto maior o valor do indicador na análise feita, melhor o resultado). A Figura 3.3 mostra a evolução para os novos canavieiros de 1980.

Tabela 3.5 – Resultados do teste Mann-Whitney para dois indicadores - 1970 a 2000

Variável	z	p
Esperança de vida 1970	-3,747	0
Esperança de vida 1980	-1,708	0,087
Esperança de vida 1991	-1,699	0,089
Esperança de vida 2000	-1,678	0,093
IDH 1970	-3,778	0
IDH 1980	-3,654	0
IDH 1991	-3,362	0
IDH 2000	-2,966	0,003

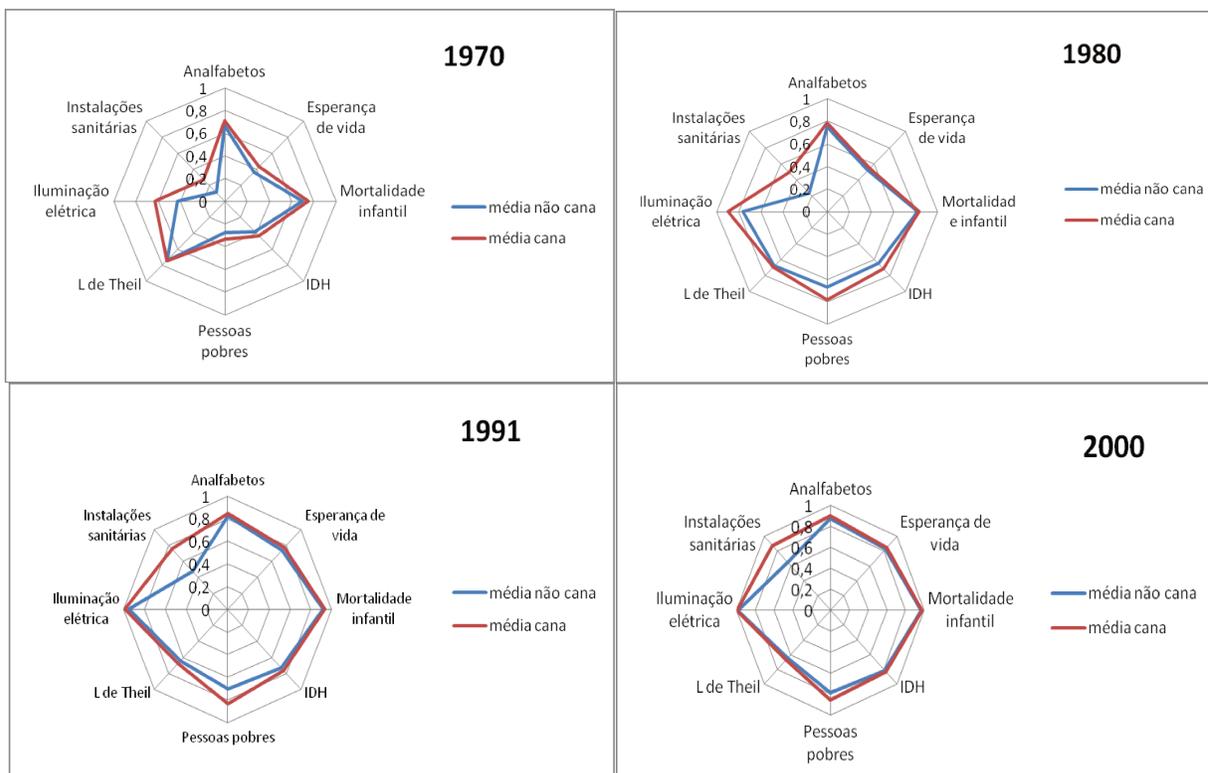


Figura 3.3 - Evolução dos indicadores no estado de São Paulo em novos canavieiros de 1980.

A mesma análise foi feita para os municípios que se tornaram canavieiros entre 1980 e 1991, período da segunda expansão da produção de etanol. Assim, os novos municípios canavieiros foram comparados com os municípios não canavieiros, e que nunca se tornaram canavieiros até 2000. Os resultados, apresentado em diagramas radar, correspondem à Figura 3.4. Como pode ser visto, o grupo de novos municípios canavieiros em 1991 já tinha indicadores com médias maiores para os anos de 1970 e 1980, quando ainda não eram significativamente canavieiros. Da mesma forma que na análise anteriormente apresentada, a conclusão é que os grupos de municípios tendem a se aproximar quanto à media dos indicadores. Em outras palavras, a expansão ou a entrada da cana nesses municípios não trouxe, em geral, vantagens significativas em relação aos municípios de mesmo porte que nunca se tornaram canavieiros.

A expansão da atividade canaveira no estado de São Paulo, entre 1970 e 2000, pode ser vista através de mapas com os principais municípios produtores, que são apresentados na Figura D1, no apêndice D.

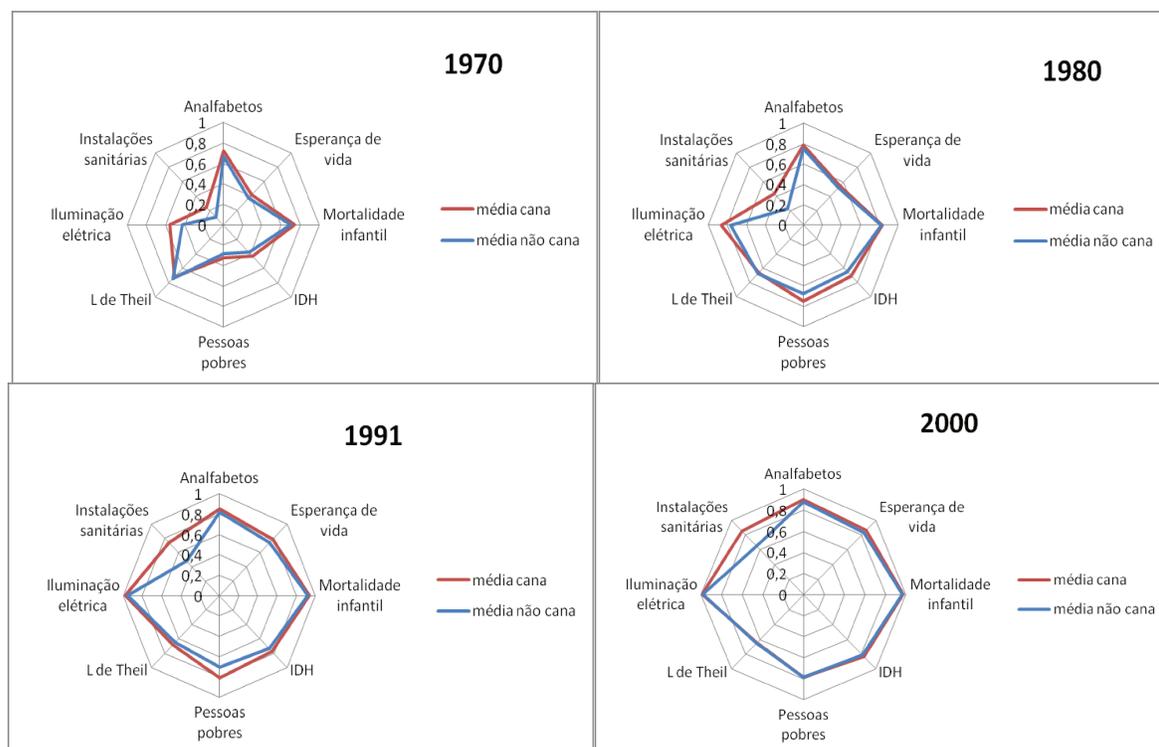


Figura 3.4 - Evolução dos indicadores no estado de São Paulo em novos canavieiros de 1991.

Cabe observar que o procedimento de análise adotado tem alguns problemas, que devem gerar distorção. Primeiro, nos dois grupos há municípios que estão localizados em regiões distintas. Como pode ser visto nos mapas apresentados no Apêndice D, os novos municípios canavieiros em 1980 estão localizados sobretudo em regiões agropecuárias, no norte e no oeste de São Paulo. Por outro lado, os municípios do grupo não canavieiro estão distribuídos por todo o estado, e também se localizam em áreas que não têm vocação agrícola. Segundo aspecto, alguns dos municípios não canavieiros atingiram alto grau de industrialização e/ou diversificação da economia no período; é provável que esses municípios afetem as médias dos indicadores.

Na tentativa de ao menos minimizar o problema acima mencionado, a análise de evolução foi feita para três grupos de novos municípios canavieiros em 1980, vizinhos entre si, e que foram comparados com grupos de municípios próximos que não se tornaram canavieiros no período considerado. O objetivo é comparar grupos de municípios que devem ter características mais similares, e reduzir o aspecto da dispersão geográfica. Os três grupos, em diferentes regiões do estado, foram definidos pelos municípios abaixo listados.

Grupo 1, na região próxima a Ribeirão Preto:

- Novos municípios canavieiros em 1980: Altinópolis, Brodowski e Jardinópolis;
- (Municípios) Não canavieiros no período: Santo Antônio da Alegria, São José da Bela Vista e Franca.

Grupo 2, na região próxima a São José do Rio Preto:

- Novos canavieiros em 1980: Tabapuã e Olímpia.
- Não canavieiros: Cajobi e Severina.

Grupo 3, na região oeste do estado:

- Novos canavieiros em 1980: Maracá e Paraguaçu Paulista.
- Não canavieiros: Borá, Lutécia, João Ramalho e Cruzália.

Dado o pequeno número de municípios em cada grupo, a comparação não pode ser considerada um estudo estatístico, mas sim um exercício de análise dos indicadores sociais estudados nesta dissertação.

No Grupo 1, os municípios novos canavieiros em 1970 (quando ainda não eram considerados canavieiros) já tinham alguns indicadores melhores do que os não canavieiros (que nunca se tornariam canavieiros). Em 1980, possivelmente logo após a transição, o grupo de novos municípios canavieiros passou a ter piores resultados do que o outro grupo, para alguns indicadores (“Esperança de vida”, “Mortalidade infantil” e “L de Theil”); em 1991 e 2000 a tendência é de aproximação entre os grupos. Já outros indicadores como “IDH” e “Pessoas Pobres” apresentam um avanço maior na década de 80 no grupo de novos canavieiros, para depois tender a aproximar com o outro grupo. Já para os indicadores de infraestrutura (acesso à rede elétrica e ao saneamento), a grande vantagem do grupo dos municípios canavieiros foi perdida para o outro grupo, possivelmente por conta da homogeneidade de ações de empresas e do poder público estadual e federal. Como pode ser visto na Figura 3.5, também para a comparação específica do caso aqui chamado de Grupo 1, a tendência geral é de aproximação dos grupos quanto aos valores médios dos indicadores.

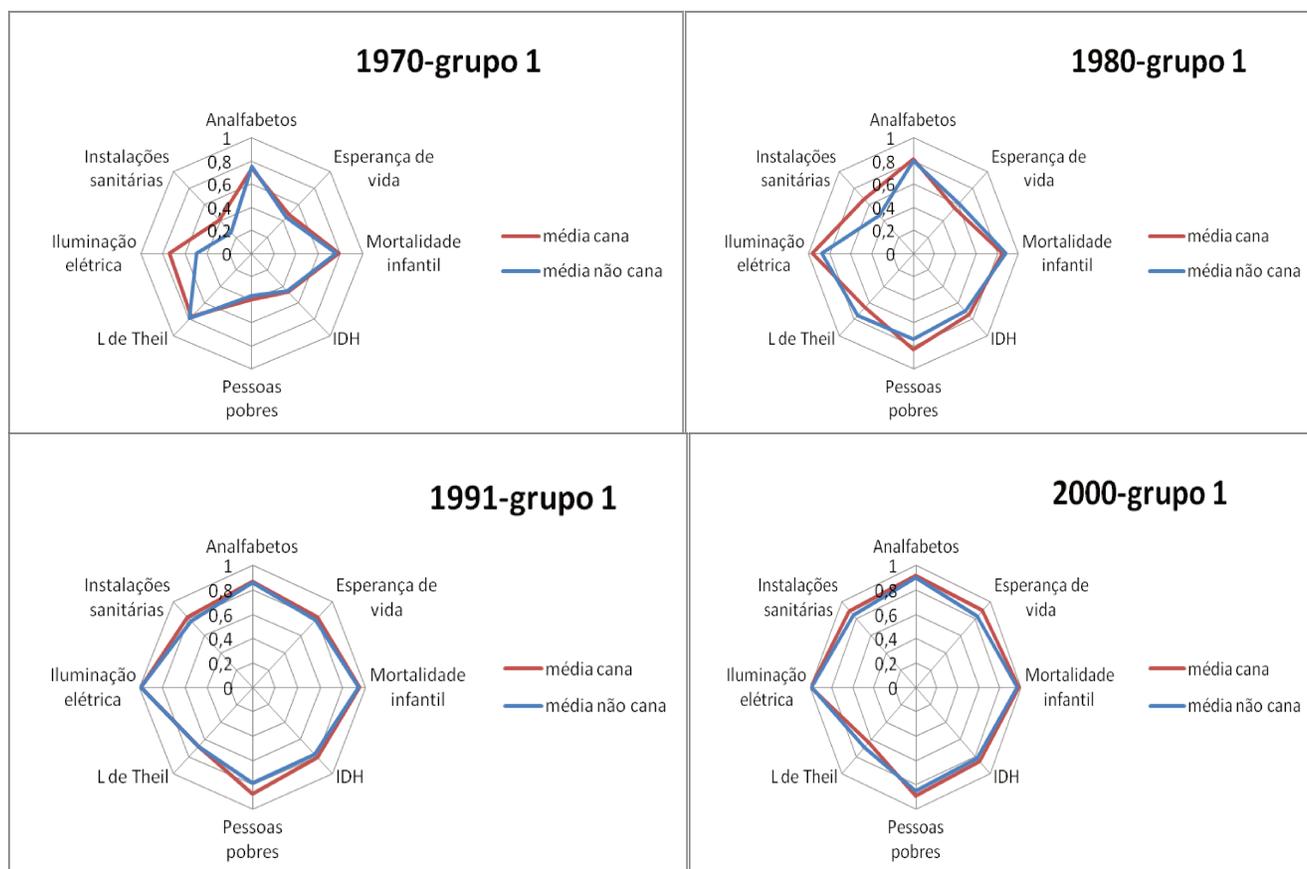


Figura 3.5 – Evolução dos indicadores sociais em comparação geográfica localizada no estado de São Paulo – Grupo 1 – 1970 a 2000

No caso do Grupo 2, os municípios novos canavieiros já tinham melhores indicadores do que o outro grupo em 1970, exceto para “Desigualdade de renda” e “Acesso à rede elétrica”. Na transição, em 1980, comparativamente os novos municípios canavieiros tiveram grande melhora no indicador “Desigualdade de renda” e perda relativa no indicador “Pessoas pobres”. Regra geral, a partir de 1991 a tendência é de aproximação dos grupos quanto aos indicadores, não sendo possível concluir por clara vantagem ou desvantagem dos municípios canavieiros em relação ao outro grupo. Os resultados são apresentados na Figura 3.6.

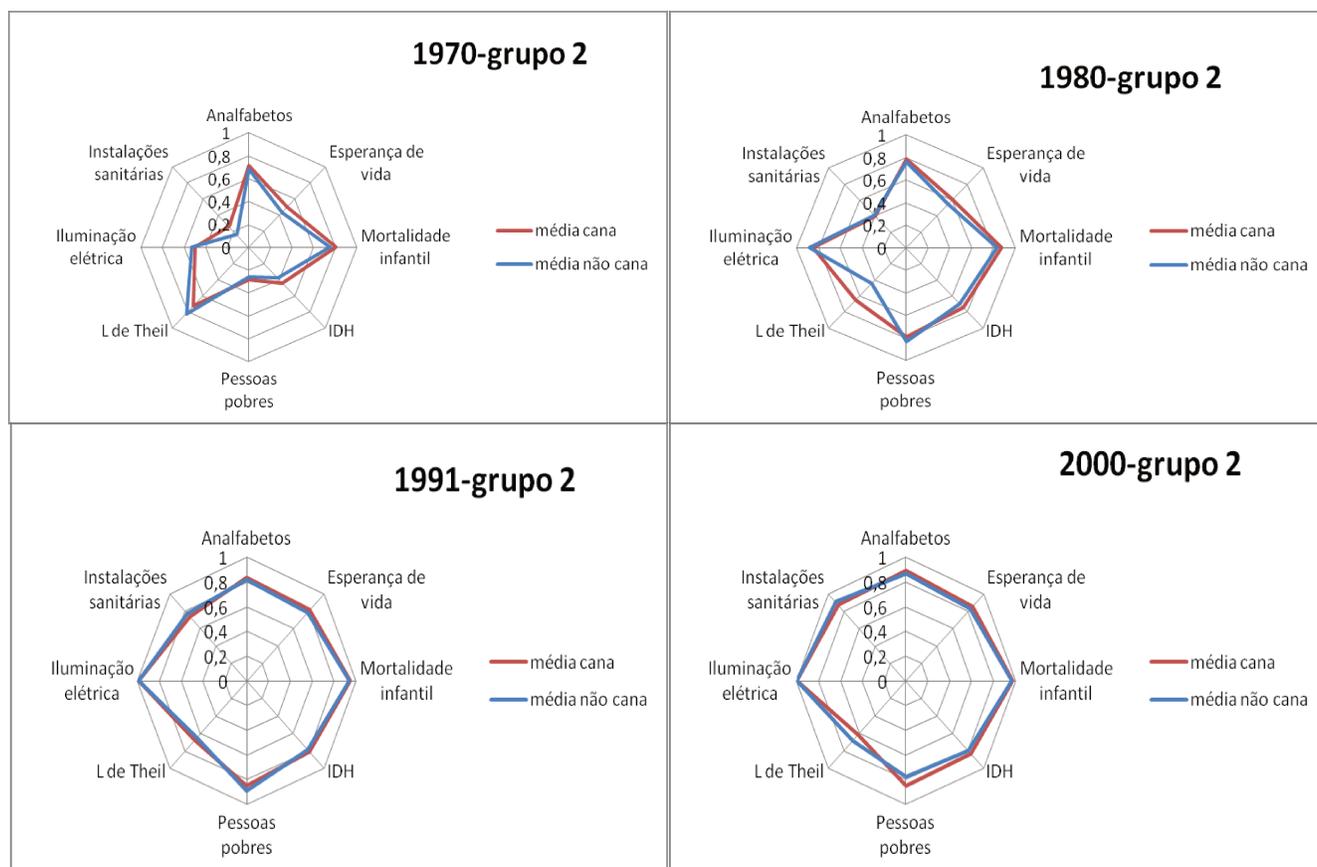


Figura 3.6 - Evolução dos indicadores sociais em comparação geográfica no estado de São Paulo – Grupo 2 – 1970 a 2000

No Grupo 3, apresentado na Figura 3.7 observa-se que os municípios canavieiros tinham alguns indicadores piores do que os do outro grupo, e que houve clara melhora em 1980. A vantagem relativa foi mantida em 1991, mas significativamente reduzida em 2000.

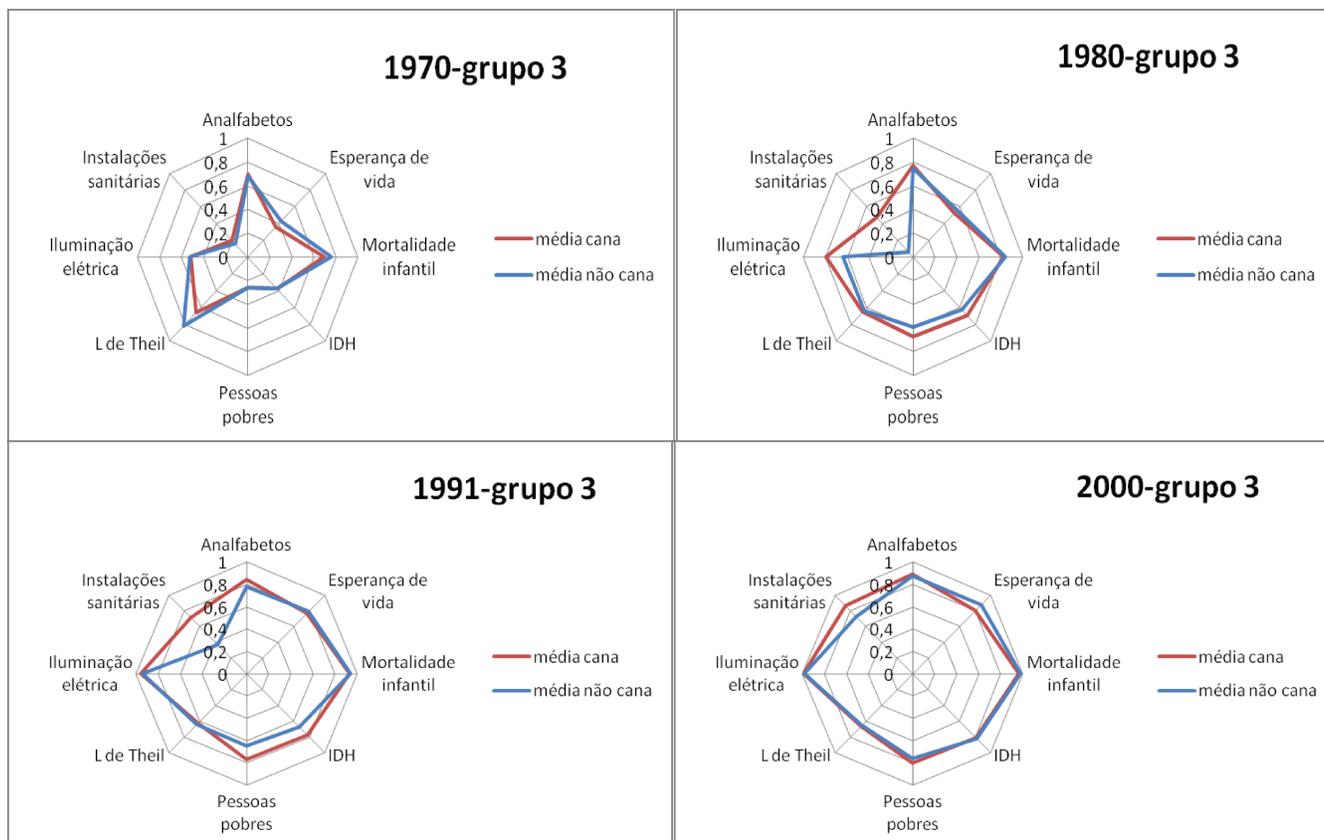


Figura 3.7 - Evolução dos indicadores sociais em comparação geográfica no estado de São Paulo – Grupo 3 – 1970 a 2000

Da análise dos três grupos é possível identificar que a entrada da cana-de-açúcar em nos municípios analisados não representou, em termos relativos, nenhum claro impacto negativo, ou positivo. Na comparação dos dois grupos de municípios, na análise da evolução dos indicadores, não existe uma tendência que indique um impacto negativo ou positivo da entrada/expansão da cana. O que é clara é a evolução dos indicadores nos três casos, mas de forma menos marcante no que diz respeito aos indicadores que sinalizam distribuição de renda.

Da mesma forma que no caso da análise mais geral, nos três grupos considerados os novos municípios canavieiros já apresentavam indicadores melhores em 1970, antes da introdução ou expansão da atividade canvieira, em relação ao outro grupo de municípios (6 indicadores no Grupo 1, 6 indicadores no Grupo 2 e 5 indicadores no Grupo 3)

A análise de evolução dos indicadores, apresentada nesta seção, indica que, se por um lado, não é possível afirmar que a atividade canvieira causou claros benefícios aos municípios no qual ela tem importância, tendo por base os oito indicadores aqui considerados, tampouco cabe considerar a hipótese de que houve malefícios. Em uma análise mais geral, o grupo de municípios no qual a atividade canvieira é importante apresenta, na média, indicadores melhores do que o grupo não canvieiro, mas com tendência de aproximação dos resultados.

Os municípios que se tornaram canavieiros quando da expansão da atividade canavieira, entre meados dos anos 1970 e meados dos anos 1980, já tinham, em geral, melhores indicadores do que os do grupo de comparação. Não foi possível chegar a uma conclusão definitiva, mas provavelmente isso se deve ao fato de que já eram municípios com tradição em agrícola, em uma época em que a diversificação da atividade econômica era pouco relevante no país. A tendência de equalização dos dois grupos de municípios deve-se, por um lado, ao desenvolvimento econômico em função da industrialização e diversificação da atividade econômica dos não canavieiros, e à eficácia de políticas públicas que não têm relação com a principal atividade econômica dos municípios.

É claro que análises gerais, sem a consideração de aspectos locais, tais como a vocação econômica do município, vantagens relativas (por exemplo, a qualidade do solo), e a existência de infraestrutura, além da não consideração de aspectos históricos, não são adequadas. Esta conclusão sugere que a metodologia para a análise da evolução dos indicadores e dos impactos da introdução ou expansão da atividade canavieira precisa ser melhorada.

4. Análise de Cluster e Discriminante

Nesta dissertação a Análise de Cluster foi utilizada para a divisão não tendenciosa dos municípios canavieiros em grupos daqueles que têm melhores e piores indicadores. Feita essa divisão, foi utilizada a Análise Discriminante para identificar parâmetros (outros que aqueles empregados na classificação original entre os grupos), chamados de *predictors*, que melhor permitem a alocação de municípios entre os grupos. No caso específico, buscou-se a identificação de fatores explicativos para a existência de municípios canavieiros com melhores e piores indicadores, ou seja, de fatores que explicam porque de municípios com melhores indicadores.

Tanto a Análise de Cluster quanto a Análise Discriminante foram feitas com os indicadores transformados. Os grupos definidos na Análise de Cluster, ou os Clusters, foram formados de acordo com o procedimento apresentado na revisão metodológica.

Inicialmente a metodologia Análise de Cluster foi empregada para a formação de dois, três e quatro grupos. Comparativamente, foi feita a escolha do agrupamento mais adequado para dar sequência ao procedimento de análise. Os critérios utilizados para seleção do número de clusters são:

- Melhor separação entre os clusters, com clara definição de grupos de municípios com melhores e piores indicadores;
- Formação de grupos com menor número de municípios nos grupos de melhores e/ou piores indicadores, o que facilitaria a identificação de casos de maior interesse;
- Melhor resultado estatístico na distinção entre os grupos, quando da aplicação da Análise Discriminante.

A aplicação do procedimento para os dados de 1991, para o estado de São Paulo é apresentada como exemplo no Apêndice B. Para todos os anos, para os estados de São Paulo e de Alagoas, o mesmo procedimento foi repetido.

4.1 São Paulo

4.1.1 1970

Em 1970, primeiro ano da série de análises, 91 municípios produziam 90% de toda cana-de-açúcar no estado de São Paulo. Na primeira análise dois clusters foram formados usando as variáveis transformadas dos oito indicadores escolhidos. Estas transformações são apresentadas no Apêndice C. Dos dois clusters formados, o cluster 1 é considerado o de melhor desempenho, sendo, na média, pior apenas quanto ao indicador “L de theil”. Na Tabela 4.1 são apresentadas os

valores médios dos indicadores em cada grupo, ou cluster, com uso das variáveis originais. Já o valor de p representa o resultado da análise de variância entre os grupos, e indica se a diferença entre os grupos é estatisticamente significativa.

Tabela 4.1 – Resultados do Cluster para a divisão em dois grupos, para São Paulo, em 1970

	Grupo 1	Grupo 2	p
Amostra	55	36	-
L de Theil	0,360	0,288	0
Pessoas pobres	51,765	66,881	00
Mortalidade infantil	76,765	83,991	0,016
Analfabetos	21,749	28,947	0
IDH	0,542	0,456	0
Esperança de vida	55,945	54,755	0,016
Iluminação elétrica	0,824	0,631	0
Instalações sanitárias	0,473	0,187	0

A formação de dois grupos permite a identificação de um grupo com indicadores estatisticamente maiores e outro com indicadores estatisticamente menores, ou seja, um grupo de municípios com melhores indicadores socioeconômicos do que o outro. Nessa divisão de grupos, porém, um indicador não se adequa ao que se supunha seria lógico: o indicador de desigualdade de renda, L de Theil, é melhor, na média, indicando menor desigualdade, no grupo de municípios que têm piores resultados para os outros indicadores. Isso é mais evidente para o estado de Alagoas, mas em São Paulo também pode ser observado.

Na Tabela 4.2 são apresentadas as correlações entre IDH, que em princípio é o indicador que sintetiza o desenvolvimento do município, e os outros indicadores. O fato da correlação ser positiva reflete a aparente contradição, uma vez que a tendência observada é de que quanto maior o IDH, maior a concentração de renda no município. Por outro lado, existe claramente uma redução nos valores das correlações entre IDH e L de Theil com o passar do tempo.

A hipótese de Kuznets, proposta pelo economista Simon Kuznets, é de que o desenvolvimento econômico (ou o crescimento da renda), primeiro aumenta para, em seguida, diminuir, ou seja, no início do processo de desenvolvimento econômico há tendência de acumulação de renda, para depois começar o processo de melhor distribuição da mesma (Kuznets, 1955). Assim, a associação entre a renda e sua distribuição, em uma dada sociedade, resultaria uma parábola, ou curva em U invertido. Godoy et al. (2004) explicam que, inicialmente, com níveis baixos de renda, e durante os primeiros estágios do crescimento econômico, a desigualdade de renda é baixa porque a maioria das pessoas é pobre e trabalham principalmente na agricultura, e em sua maioria na agricultura de subsistência. Com as transformações e a evolução da agricultura, a desigualdade de renda cresce com a ida de algumas pessoas para trabalhos urbanos, com salários mais elevados, nas indústrias ou nos serviços. Com mais pessoas indo para as cidades e

trabalhando em processos industriais e no setor de serviços, a desigualdade de renda decai na medida em que há contração da agricultura como uma fonte de renda e emprego. Da mesma forma, a desigualdade também cai em sociedades industrialmente desenvolvidas devido à implementação de políticas públicas pelos governos, ajudando aqueles que estão na base da distribuição de renda. Assim, pode-se perceber a evolução do desenvolvimento e da desigualdade de renda na Tabela 4.2. Enquanto em 1970 existia certa correlação positiva entre desenvolvimento e desigualdade, ou seja, quanto maior o desenvolvimento maior a desigualdade, em 2000 essa relação passa a ser estatisticamente não significativa; seguindo a hipótese de Kuznets, a correlação pode se tornar negativa com o passar dos anos.

Ainda a respeito do que é apresentado na Tabela 4.2, em 1970 as correlações eram fracas entre IDH e L de Theil e Mortalidade infantil. Já em 2000, além do L de Theil, as correlações passaram a ser mais fracas, com o tempo, entre IDH e “Iluminação elétrica” e “Instalações sanitárias”, o que indica a melhora da infraestrutura nos municípios independente de outros resultados quanto ao desenvolvimento socioeconômico.

Tabela 4.2 – Correlações entre IDH e outros indicadores sociais de 1970-2000

Ano	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	Esperança de vida	Iluminação Elétrica	Instalações sanitárias
IDH 1970	0,256	-0,905	-0,383	-0,867	0,378	0,823	0,721
IDH 1980	0,194	-0,827	-0,273	-0,779	0,275	0,726	0,653
IDH 1991	0,028	-0,845	-0,523	-0,792	0,510	0,552	0,604
IDH 2000	0,071	-0,837	-0,727	-0,804	0,721	0,518	0,483

Feita a opção por dois clusters, seguiu-se com a Análise Discriminante. A Análise Discriminante feita para o ano de 1970 diferencia-se dos demais anos pelo fato de que, para aquele ano, não há disponibilidade de dados referentes a finanças públicas e dados relacionados à contribuição da renda dos setores na economia dos municípios. Por isso, apenas sete variáveis são disponibilizadas para o modelo, e apenas seis foram incluídas (pelo software). Das variáveis escolhidas, apresentadas na Tabela 4.3, a “Participação da cana na economia” tem maior relevância na discriminação entre os grupos, de acordo com seu coeficiente da função discriminante, seguida por “Cana/habitante” e “Participação dos serviços”; os coeficientes da função discriminante devem ser analisados em módulo.

Das funções de classificação podem-se identificar algumas tendências, a partir das variáveis incluídas no modelo. Um município com forte participação da cana na economia, a tendência é sua classificação no grupo daqueles que têm piores indicadores, ou seja, o Grupo 2. Todas as outras variáveis indicam tendência direta de classificação no Grupo 1, ou seja, em função da importância; por exemplo, maior participação de serviços ou da indústria na economia. A análise deve ser feita mantendo as outras variáveis fixas, e alterando a variável estudada. Na prática, no entanto, os indicadores são relacionados e possuem limites superiores, ou seja, suas combinações são limitadas. Os coeficientes apresentados na Tabela 4.3 são resultado do cálculo feito com as variáveis transformadas, cujas operações são apresentadas no Apêndice C. As variáveis *predictors* no modelo respondem por 44,8% da discriminação entre os grupos, de acordo com o lambda de Wilks da última variável a entrar no modelo. Isso significa que se fossem utilizadas essas variáveis para a classificação dos municípios nos dois grupos, o acerto seria de aproximadamente 45%. Mas, deve ser lembrado, não esse é o objetivo da Análise Discriminante neste trabalho, e sim o de encontrar fatores que explicam a existência de municípios com resultados diferentes do ponto de vista socioeconômico.

Tabela 4.3 - Resultados da análise discriminante com dois grupos para São Paulo em 1970

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Participação da agropecuária	0,545	878,696	874,570	0,326
Participação cana na economia	0,660	-142,529	-126,487	-1,718
Cana/km²	0,546	0,552	0,377	0,239
Participação de serviços	0,587	1109,516	1086,647	1,236
Cana/habitante	0,576	75,124	68,375	1,308
Participação da indústria	0,551	819,033	811,055	0,714

Para dar suporte à Análise Discriminante, as correlações entre variáveis de qualidade de vida adotadas neste trabalho, e as de importância relativa de setores da economia (e também relacionadas à produção de cana) foram calculadas e são apresentadas na Tabela 4.4. A correlação entre “Cana na economia” e os indicadores sociais, quando estatisticamente significativos, tende a ser inversamente proporcional ao que se entende como melhores indicadores.

Tabela 4.4 – Correlação entre indicadores econômicos e sociais – São Paulo 1970

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,397	0,313	0,041	0,166	-0,370	-0,041	-0,410	-0,377
Cana/km ²	-0,508	-0,156	-0,010	-0,010	-0,073	0,012	0,117	-0,010
Participação da indústria	0,025	-0,516	-0,043	-0,283	0,389	0,038	0,545	0,469
Participação dos serviços	0,591	-0,153	0,041	-0,345	0,393	-0,028	0,151	0,325
Participação da agropecuária	-0,421	0,592	0,013	0,501	-0,634	-0,016	-0,619	-0,664

Para ilustrar a análise da correlação entre IDH e participação da cana na economia, foi plotada a dispersão das variáveis, que mostrada na Figura 4.1. Em 1970, existiam municípios com valores mais altos de IDH e com baixa participação da cana na sua economia, bem como baixos IDH e baixa participação da cana. Porém, não existiam municípios com maior IDH (por exemplo, > 0,5) e alta participação da cana.

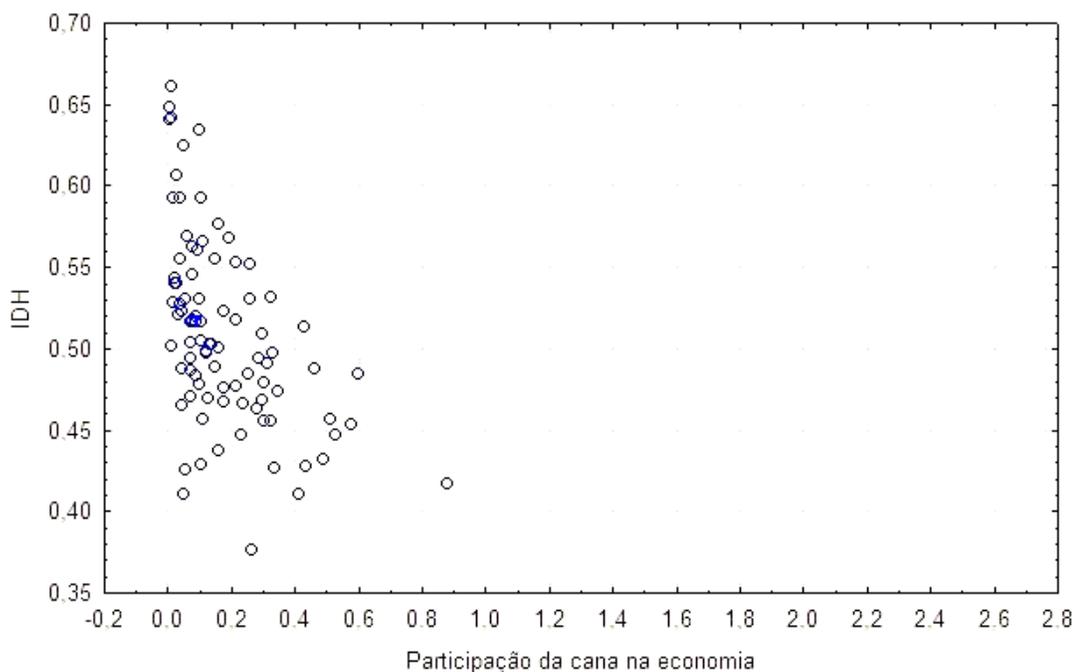


Figura 4.1 - Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1970.

Na Figura 4.2 É apresentada a distribuição espacial dos dois clusters de municípios canavieiros no estado de São Paulo, em 1970.

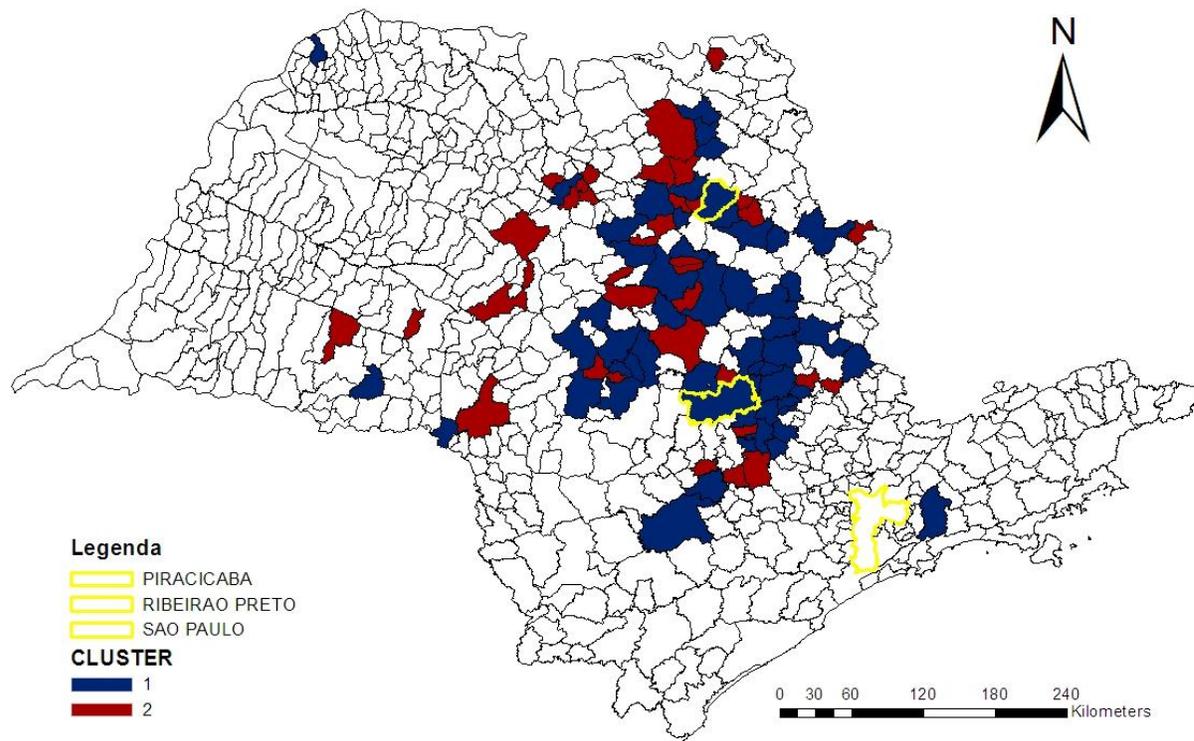


Figura 4.2 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1970.

4.1.2 1980

No ano de 1980 o número de municípios canavieiros cresce para 116 no estado de São Paulo, ou seja, a produção estadual agora é menos concentrada e 90% da produção de cana-de-açúcar é dividida entre mais municípios no estado.

Tabela 4.5 – Resultados do Cluster em 2 grupos para São Paulo em 1980

	Grupo 1	Grupo 2	p
Membros	72	44	-
L de Theil	0,375	0,428	0,005
Pessoas Pobres	13,087	23,892	0
Mortalidade infantil	49,106	61,227	0
Analfabetos	17,734	23,225	0
IDH	0,725	0,677	0
Esperança de vida	61,242	58,766	0
Iluminação elétrica	0,953	0,878	0
Instalações sanitárias	0,625	0,399	0,005

O cluster de pior desempenho dos indicadores para o ano de 1980 é o grupo 2, como visto na Tabela 4.5.

A análise discriminante resultante para o ano de 1980 apresentada na Tabela 4.6 revela que a expressiva participação da cana na agricultura classifica o município no grupo 2, porém o indicador “Cana/habitante” classifica no grupo 1, de melhor desempenho. A integração dos indicadores deve ser levada em conta. Por exemplo, uma grande participação da agropecuária atrelada a uma baixa participação dos serviços unida a uma alta participação da cana na agricultura e baixa produção de cana-de-açúcar no município deve classificá-lo no grupo 2, de pior desempenho. Estas variáveis respondem por apenas 16,7% da discriminação entre os grupos. A Tabela 4.7 mostra a relação entre as atividades econômicas dos municípios e os indicadores sociais, assim como a Figura 4.3 mostra a dispersão do IDH com a participação da cana na economia.

Tabela 4.6- Resultados da análise discriminante com dois grupos para São Paulo em 1980

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Participação da agropecuária	0,971	-14,111	-9,812	1,429
Participação da cana na agricultura	0,864	-0,482	2,307	1,030
Cana/habitante	0,857	6,704	4,491	-1,045
Participação dos serviços	0,833	-24,211	-26,195	-0,294

Algumas correlações estatisticamente significativas suportam a hipótese que a presença da cana no município tem proporção inversa aos indicadores de qualidade de vida, como número de analfabetos e IDH. A participação da agropecuária também apresenta esta tendência, com proporcionalidade negativa em relação à Iluminação elétrica, IDH, número de analfabetos e porcentagem de pessoas pobres. No ano de 1980 no estado de São Paulo a correlação entre participação da indústria e porcentagem de municípios com instalações sanitárias era negativa e entre participação de da agropecuária positiva, o que difere do ano de 1970 estudado.

Tabela 4.7 – Correlação entre indicadores econômicos e sociais – São Paulo 1980

Variável	L de theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,277	0,126	0,028	0,482	-0,407	-0,031	-0,117	0,124
Cana/km ²	-0,499	-0,420	-0,009	0,205	-0,058	0	0,345	-0,350
Participação da indústria	-0,126	-0,486	0,088	-0,405	0,366	-0,094	0,497	-0,504
Participação dos serviços	0,130	-0,094	0,057	-0,170	0,182	-0,062	0,180	-0,178
Participação da agropecuária	0,049	0,550	-0,124	0,514	-0,482	0,133	-0,613	0,620

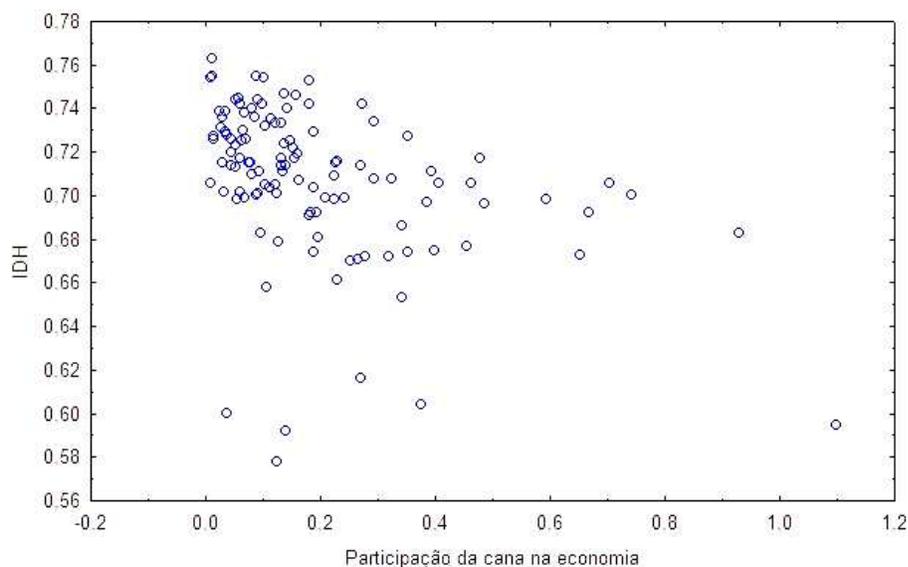


Figura 4.3 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1980.

Finalmente a dispersão geográfica dos municípios em cada cluster é apresentada na Figura 4.4.

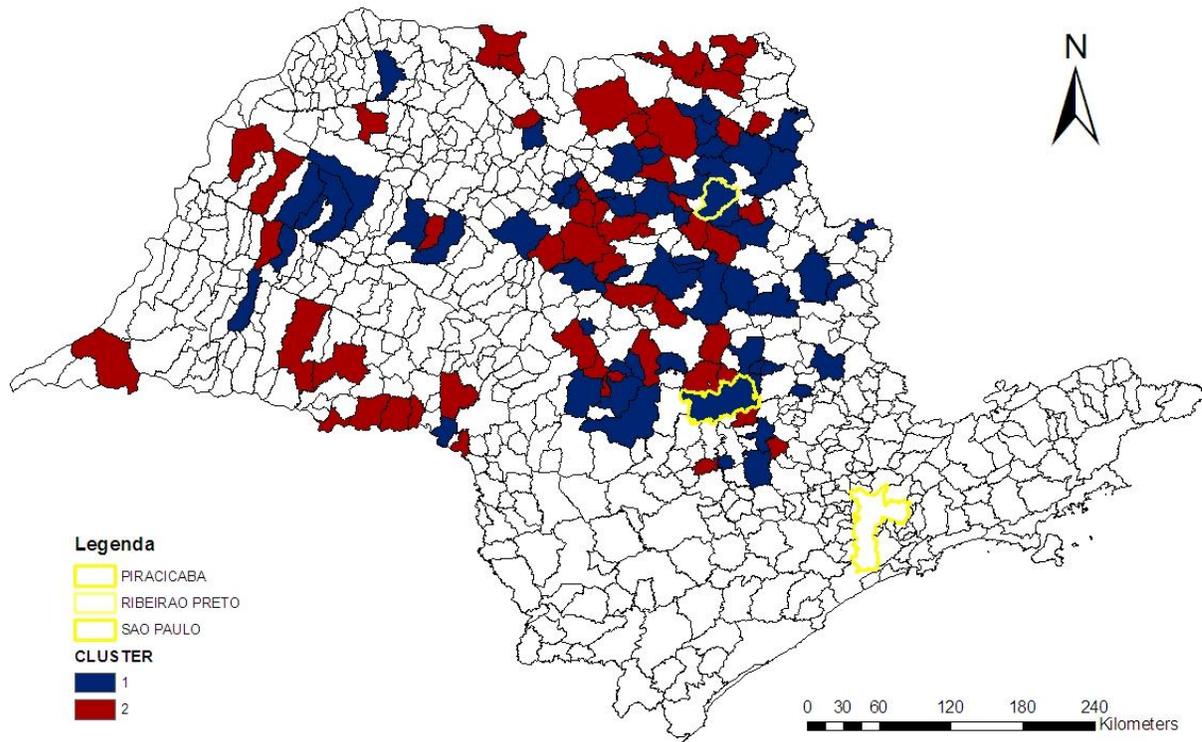


Figura 4.4 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1980.

4.1.3 1991

Em 1991 havia 148 municípios considerados canavieiros no estado de São Paulo e sua divisão em dois grupos resultou um grupo com 98 municípios e outro com 50 municípios. Os 98 municípios com melhores indicadores foram alocados no Grupo 1, como pode ser observado na Tabela 4.8. Os valores de p são referentes à comparação entre as variáveis transformadas, porém os resultados dos valores médios apresentados na tabela são dos valores originais de cada indicador, para facilitar o entendimento. O Grupo 1, que tem os melhores indicadores, tem médias significativamente distintas (estatisticamente) do segundo grupo para sete dos oito parâmetros (a única exceção é indicador “L de Theil”).

Tabela 4.8 – Resultados da Análise Cluster em dois grupos, para São Paulo, em 1991

	Grupo 1	Grupo 2	p
Membros	98	50	-
L de Theil	0,407	0,419	0,423
Pessoas pobres	12,618	22,619	0
Mortalidade infantil	21,924	28,872	0
Analfabetos	12,782	16,408	0
IDH	0,777	0,716	0
Esperança de vida	70,481	67,865	0
Iluminação elétrica	0,993	0,971	0
Instalações sanitárias	0,838	0,640	0

A Análise Discriminante foi então aplicada para as classificações com dois clusters formados. A Tabela 4.9 mostra os principais índices e coeficientes obtidos na análise, quando os municípios são classificados em dois grupos. Neste modelo, as variáveis representam apenas 7,5% da discriminação entre os grupos.

Tabela 4.9- Resultados da Análise Discriminante para a classificação em dois grupos, para São Paulo, em 1991

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Renda serviços	0,936	190,051	187,513	-0,502
Cana/habitante	0,935	-28,550	-31,326	-1,664
Cana na economia	0,929	81,809	87,114	1,356
Transferência estadual	0,944	-0,255	-0,335	-0,990
Cota-parte municipal	0,935	-0,605	-0,378	0,815
Cana na agricultura	0,927	8,897	8,144	-0,337

A função discriminante de dois grupos mostra que as variáveis “Cana por habitante” e “Cana na economia” são as que mais contribuem para a discriminação entre os grupos, o que indica uma forte relação entre a produção de cana nos municípios e seu desempenho na qualidade de vida, seguidas das variáveis “Transferência municipal” “Cota-parte municipal”, o que indica a influência das finanças públicas no desenvolvimento.

Na análise das funções de classificação, os coeficientes de cada variável não devem ser comparados, já que não são coeficientes padronizados (ou *standardized*, em inglês), assim, a

comparação deve ser feita entre grupos. Se um novo município deve ser classificado entre os grupos 1 e 2, uma renda média no setor dos serviços, por exemplo, tenderia para uma classificação no grupo 1, já que seu coeficiente é maior para este grupo. As variáveis relacionadas à cana, “Cana por habitante” e “Cana na economia” seguem padrões diferentes. Enquanto “Cana por habitante” tende a classificar para o grupo 1, “Cana na economia” tende a classificar no grupo 2, de pior desempenho. Ou seja, a larga produção de cana (lembrando que aqui a produção está em unidades monetárias) classificaria um novo município no grupo de melhores desempenhos dos indicadores de bem-estar, porém, a concentração da atividade econômica da cana-de-açúcar classifica este município no grupo de piores desempenhos dos indicadores. Esta mesma conclusão pode ser tirada ao analisar a participação da cana na agricultura do município. Ou seja, quanto maior concentração do setor na atividade agrícola canavieira, pior o desempenho de seus indicadores sociais. Para completar esta análise, o gráfico de dispersão entre “IDH” e “participação da cana na economia” é mostrado na Figura 4.5.

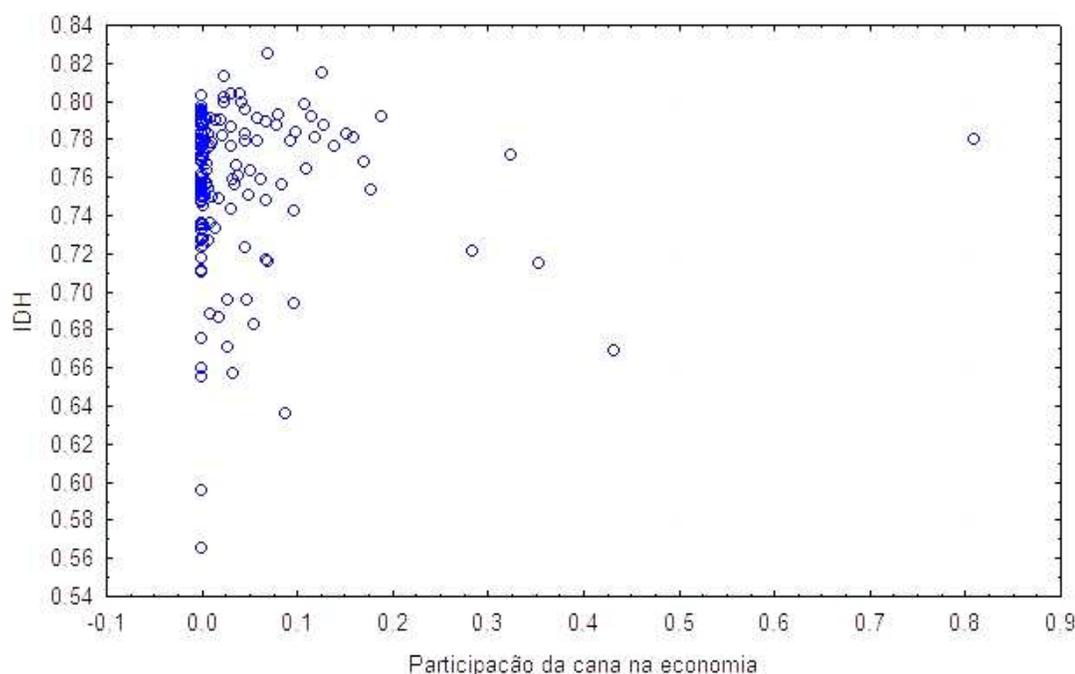


Figura 4.5 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo no ano de 1991.

Uma concentração de IDH altos em pequenas participações da produção de cana-de-açúcar na economia pode ser notada, porém a correlação entre as variáveis não é significativa, como mostra a Tabela 4.10.

Tabela 4.10 – Correlação entre indicadores econômicos e sociais – São Paulo 1991

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,026	-0,004	0,103	-0,008	-0,002	-0,115	-0,066	0,051
Cana/km ²	-0,021	-0,076	0,049	-0,108	0,076	-0,052	0,056	0,131
Participação da indústria	-0,031	-0,042	-0,034	-0,039	0	0,034	-0,023	0,033
Participação de serviços	-0,029	-0,009	-0,035	0,059	-0,030	0,029	0,045	0,057
Participação da agropecuária	0,045	0,042	0,052	-0,001	0,024	-0,048	-0,007	-0,064

Como nenhuma das correlações calculadas se mostrou significativas, nada se pode concluir sobre a relação entre as variáveis selecionadas.

As regiões de Piracicaba e Ribeirão Preto estão destacadas na Figura 4.6 para evidenciar o desenvolvimento da região. Outras regiões, como a região sul e extremo norte tem a maior concentração de municípios do grupo 2, de pior desempenho dos indicadores de qualidade de vida. Os municípios considerados canavieiros no ano de 1990 no estado de São Paulo presentes na região do Pontal do Paranapanema também se encontram no grupo 2 (Teodoro Sampaio e Caiuá).

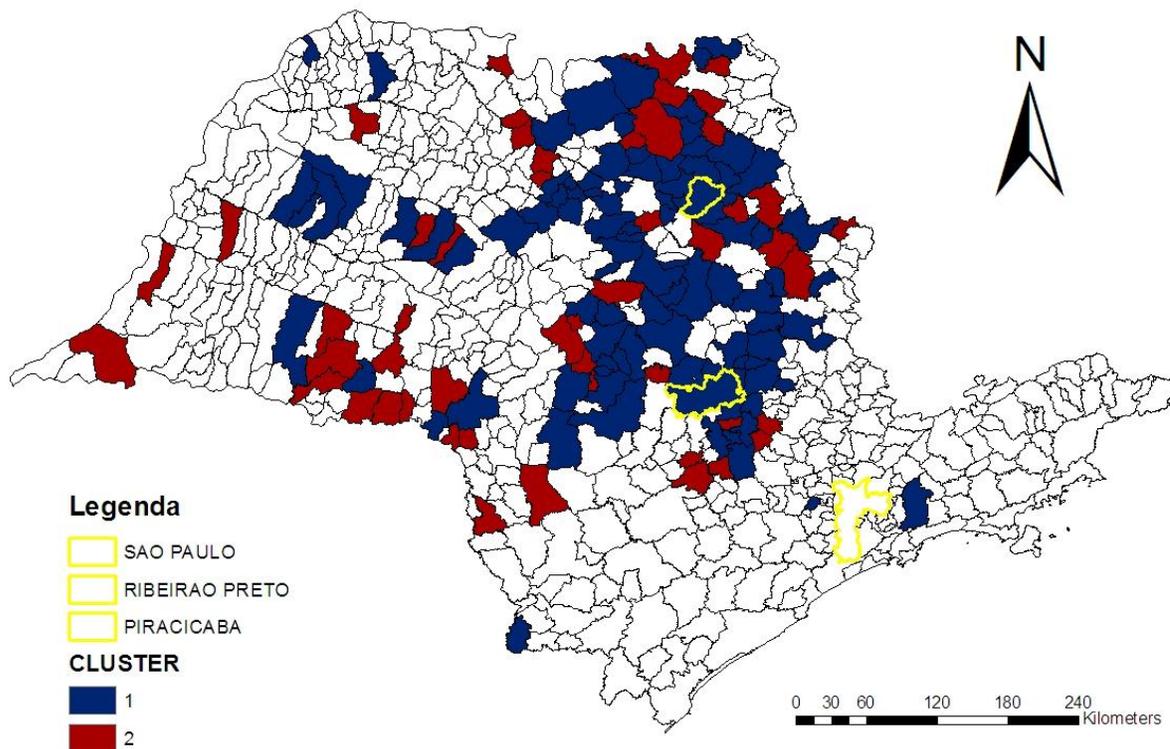


Figura 4.6 - 2 grupos de desempenho dos indicadores de qualidade de vida no estado de São Paulo em 1991.

4.1.4 2000

Pelos critérios adotados nesta dissertação, em 2000 o número de municípios considerados canavieiros subiu para 183; divididos em dois grupos estatisticamente distintos, o Grupo 1 é o de municípios com melhores resultados, para sete de oito indicadores. O agrupamento em 2000 também foi feito utilizando-se os dados transformados, para se buscar normalidade, porém as médias apresentadas na Tabela 4.11 são as dos valores originais.

Tabela 4.11 – Resultados do Cluster em 2 grupos para São Paulo, em 2000

	Grupo 1	Grupo 2	P
Amostra	93	90	-
L de Theil	0,456	0,449	0
Pessoas pobres	12,669	19,187	0
Mortalidade infantil	11,232	16,267	0
Analfabetos	8,373	11,563	0
IDH	0,811	0,772	0
Esperança de vida	73,974	71,074	0
Iluminação elétrica	0,997	0,992	0
Instalações sanitárias	0,903	0,819	0

Os resultados da Análise Discriminante aplicada aos dois grupos formados em 2000, para o estado de São Paulo, são apresentados na Tabela 4.12. As variáveis escolhidas respondem, de acordo com o lambda de Wilks, a 32% da variância entre os grupos, isso após a entrada da última variável; isso significa que o modelo seria capaz de alocar os municípios nos dois grupos com precisão de 32%. As variáveis mais significativas, segundo a função discriminante padronizada, são “Renda na indústria” e “Cana na economia”. Outras variáveis referentes à renda dos trabalhadores nos setores também foram incluídas no modelo, o que indica a forte relação dos indicadores de renda com a lógica de formação dos grupos.

Tabela 4.12 - Resultados da Análise Discriminante para a classificação em dois grupos, para São Paulo, em 2000

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Receita total	0,679	37,49	36,33	0,326
Renda na indústria	0,744	-2,36	-2,18	-0,621
Cana/km²	0,696	-8,87	-9,25	0,397
Cana na economia	0,696	460,15	467,95	-0,577
Renda no comércio	0,680	644,55	639,60	0,232
Renda agricultura	0,681	201,40	198,62	0,227
Renda nos serviços	0,679	559,84	555,66	0,212

As funções de classificação indicam a tendência de alocação dos municípios em cada um dos grupos, em função dos indicadores incluídos no modelo. A “Produção de cana (em valores monetários) por quilômetro quadrado” indica tendência de classificação do municípios maiores produtores no Grupo 1, de melhores resultados. Já a variável “Participação da cana-de-açúcar na economia municipal” indica tendência de classificação no Grupo 2, de piores resultados, ou seja, a grande dependência econômica do município na produção de cana deve ser resultar, preferencialmente, em piores indicadores sócio-econômicos.

As variáveis referentes à contribuição da renda dos diferentes setores na economia mostram que o Grupo 1 tem, na média, os municípios nos quais a contribuição dos setores de serviços, comércio e agricultura são mais relevantes. A maior relevância do setor industrial na economia tende a impactar negativamente os indicadores, levando à classificação no Grupo 2.

As correlações entre os indicadores de qualidade de vida escolhidos neste trabalho e os indicadores de analisados na Análise Discriminante (de importância econômica dos setores, assim como de produção de cana-de-açúcar) são apresentadas na Tabela 4.13. A maior importância da produção de cana-de-açúcar na economia de um município está negativamente correlacionada com o resultado de sete dos oito indicadores, com a exceção do índice “L de Theil”. As correlações marcadas em verde são estatisticamente significativas e, como se pode observar, a renda da indústria apresenta uma correlação positiva com os valores de IDH, assim

como a renda do setor de serviços. A “Produção de cana por quilômetro quadrado” tem apenas uma correlação estatisticamente significativa: com L de Theil (quanto maior a concentração de cana, menor o indicador de distribuição de renda, ou seja, melhor o resultado quanto a esse aspecto).

Na Figura 4.7 é apresentada a dispersão entre a participação da cana na economia e o IDH, para o municípios considerados canavieiros em 2000. É clara a correlação negativa entre os indicadores, como apresentado na Tabela 4.13.

Tabela 4.13 – Correlação entre indicadores econômicos e sociais – São Paulo 2000

Variável	L de Theil	Pessoas pobres	Mortalidade de infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,280	0,471	0,169	0,599	-0,566	-0,161	-0,302	-0,431
Cana/km ²	-0,4278	-0,131	-0,033	0,136	-0,133	0,019	0,087	0,140
Participação da indústria	-0,075	-0,279	-0,147	-0,267	0,238	0,138	0,161	0,243
Participação dos serviços	0,2178	-0,131	0,019	-0,207	0,210	-0,020	0,083	0,176
Participação da agropecuária	-0,149	0,551	0,186	0,624	-0,586	-0,172	-0,326	-0,553

É importante destacar que a correlação não identifica relações não lineares entre as variáveis e apenas representa uma proporção matemática. A dependência real das variáveis não é diretamente comprovada por sua correlação, ou seja, aqui não se pretende concluir que há relação entre causa e efeito, mas sim identificar-se tendências.

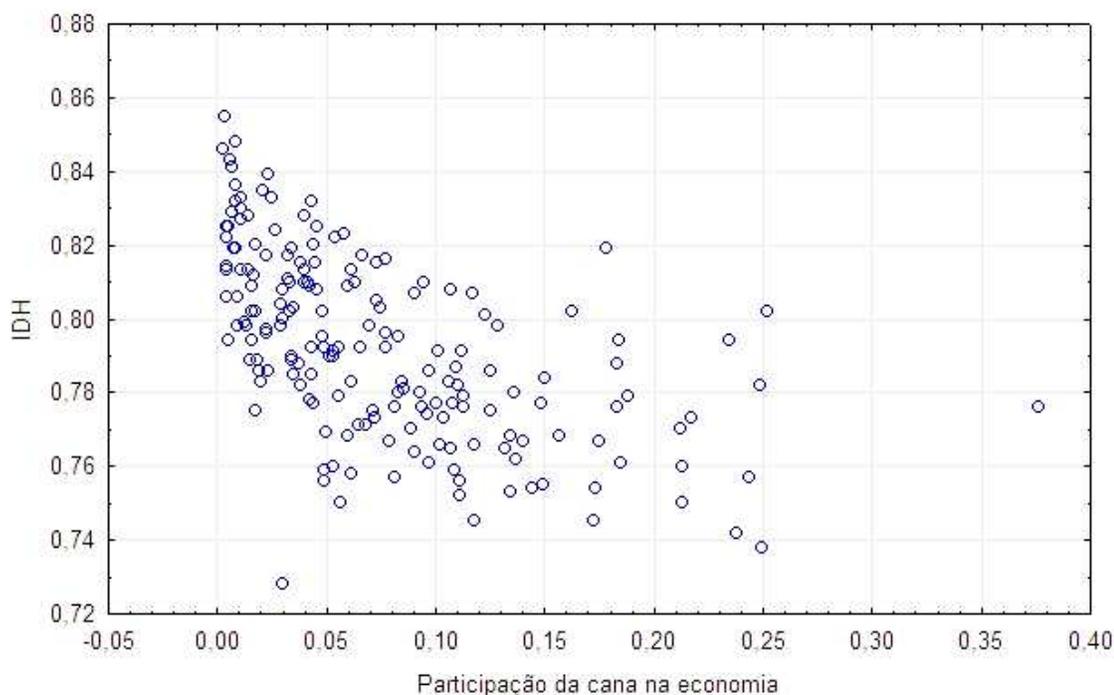


Figura 4.7 – Dispersão entre “IDH” e “Participação da cana na economia” no estado de São Paulo, no ano de 2000.

O caso de Ribeirão Preto é adequado para um aprofundamento da análise. Em 2000, a produção de cana no município estava entre as 10 maiores do estado de São Paulo, mas a participação do cultivo da cana na sua economia era irrisória, apenas 0,37%. Já seu IDH era o maior do grupo de municípios canavieiros e o sexto no estado de São Paulo naquele ano. Claramente a economia de Ribeirão Preto depende bastante da atividade canavieira, mas o plantio da cana não deve ser fator muito relevante; a diversificação da atividade econômica, e o fato de ser município polo de uma região, são fatores que não estão sendo identificados na análise aqui feita. Em adição, mesmo quando a importância do setor canavieiro não era tão relevante, o município já tinha melhores indicadores em relação aos seus vizinhos, o que reforça a necessidade de se analisar a evolução dos indicadores dentro de uma perspectiva histórica.

Na Figura 4.8 é apresentada a distribuição geográfica dos clusters, em 2000, com a indicação das cidades de Piracicaba, Ribeirão Preto e a capital, São Paulo.

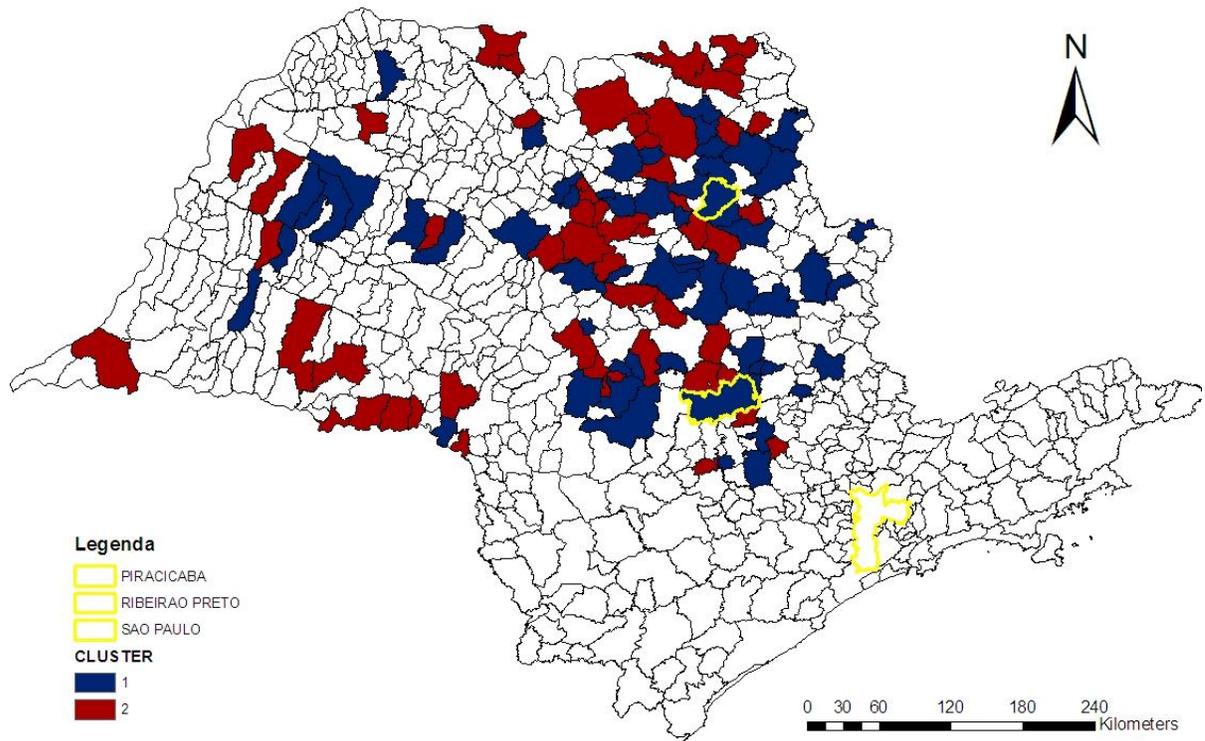


Figura 4.8 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de São Paulo, em 2000

Em 2000, é notável o aumento do número de municípios canavieiros e o aumento do número de municípios no grupo de piores indicadores. Porém, os valores médios de todos os indicadores dos municípios no Grupo 2, em comparação com 1991 e com exceção do índice “L de Theil”, melhorou consideravelmente; deve ser destacada a queda de 77% do indicador “Mortalidade infantil” no grupo de municípios com os piores indicadores.

Foram identificados sete municípios que sempre estiveram no grupo de municípios com piores indicadores, e trinta e seis que sempre estiveram no grupo de melhores indicadores. Esses municípios deveriam ser analisados com mais detalhes no esforço de entender o que mais influencia bons ou maus resultados.

Finalmente, na Tabela 4.14 são apresentadas as médias e os desvios padrão dos indicadores que refletem as participações dos setores na economia; também é apresentada informação sobre a população rural nos dois grupos de municípios canavieiros, em cada ano. Foi feito o teste *t* de hipótese para a verificação da diferença entre os grupos, com rigor estatístico; os resultados que são estatisticamente diferentes são marcados em verde, e na tabela são apresentadas as médias dos valores não transformados. Pode-se ver que em todos os anos os municípios com piores indicadores são os que têm maior população rural, e isso ocorre em todos os anos. Exceto em

1991, o grupo de municípios nos quais a importância relativa da agropecuária e da cana na economia são maiores, têm piores indicadores. Já quanto à “Participação da indústria” na economia, e exceto em 1991, o grupos de municípios com melhores indicadores apresenta maiores valores do parâmetro.

Tabela 4.14 – Médias das participações econômicas dos setores e da população rural em São Paulo, em municípios canavieiros e vários anos.

Variável	1970		1980		1991		2000	
	Grupo melhor	Grupo pior						
Participação cana	0,11±0,10	0,33±0,43	0,16±0,16	0,24±0,22	0,04±0,04	0,03±0,07	0,05±0,05	0,10±0,06
Participação indústria	0,39±0,16	0,27±0,17	0,43±0,18	0,32±0,15	0,24±0,24	0,22±0,15	0,29±0,13	0,24±0,13
Participação serviços	0,36±0,13	0,28±0,08	0,34±0,10	0,32±0,12	0,47±0,47	0,48±0,10	0,61±0,10	0,59±0,10
Participação agropecuária	0,23±0,14	0,44±0,16	0,22±0,16	0,34±0,17	0,27±0,27	0,28±0,19	0,08±0,07	0,16±0,08
Porcentagem população rural	0,31±0,12	0,52±0,17	0,23±0,12	0,36±0,15	0,13±0,13	0,25±0,12	0,09±0,06	0,14±0,08

Da Análise Discriminante conclui-se que, em geral, os indicadores que refletem a importância da cana na economia e a maior concentração da atividade têm relevância na separação dos municípios em dois grupos. Quando entra como variável relevante no modelo discriminante, quanto maior o parâmetro “Participação da cana na economia” mais provável a classificação dos municípios no grupo que têm piores indicadores sociais. Já as variáveis “Cana por área” e “Cana por habitante” não têm o mesmo comportamento no modelo discriminante. A conclusão é que os piores resultados sócio-econômicos estão associados à maior dependência econômica dos municípios da atividade canavieira, e não à produção de cana em si. Os parâmetros relativos às contribuições do estado na arrecadação municipal, quando disponíveis, também têm importância na separação dos grupos. É importante levar em conta que a análise discriminante não fornece uma comparação entre os coeficientes das funções de classificação, ou seja, não há comprovação estatística que as funções são diferentes.

A inserção dos parâmetros associados à participação relativa dos setores na economia indica que quanto mais diversificada a estrutura econômica, ou seja, maior a participação da indústria e dos serviços, e menor a participação da cana, mais provável que o município esteja no grupo dos que têm melhores indicadores.

A conclusão acima é reforçada pela análise das correlações entre os indicadores socioeconômicos e os parâmetros explorados nos modelos discriminantes. Regra geral, quanto maior a “Participação da cana-de-açúcar” na economia, piores os indicadores socioeconômicos. Além

disso, quanto maior as participações da indústria e dos serviços na economia, melhores são os indicadores socioeconômicos. Como comentado anteriormente, os municípios canavieiros com piores indicadores socioeconômicos são, em geral, aqueles que têm maior fração da população vivendo no meio rural.

4.2 Alagoas

A análise feita para o estado de Alagoas é similar à que foi feita para os municípios canavieiros do estado de São Paulo. Ao contrário do que foi feito quando da análise quando da comparação entre o grupo de municípios canavieiros e não canavieiros, descrita no Capítulo 3, o município de Maceió não foi excluído do grupo canavieiro. Uma primeira razão é que o objetivo aqui não é a comparação com os municípios não canavieiros e, em princípio, a consideração de Maceió no grupo analisado não deveria causar maiores problemas.

Maceió não faz parte do grupo de municípios que cobriram 90% da produção de cana em Alagoas em 1970, mas faz parte do grupo a partir de 1980. Portanto, em 1970 o problema não se colocava. Para os três anos a partir de 1980, quando Maceió poderia ou não ser incluído no grupo, foram feitos testes para a verificação dos clusters a serem formados, considerados municípios com melhores e piores indicadores. Observou-se que não houve significativa diferença no número de municípios em cada grupo, nem dos municípios que eram alocados em cada grupo.

Adicionalmente, foi analisado o poder de discriminação dos modelos ajustados quando da inclusão de Maceió, e verificaram-se bons resultados quando de sua consideração no grupo. Assim, optou-se por manter Maceió no grupo de municípios analisados.

4.2.1 1970

O estado de Alagoas tinha 94 municípios no ano de 1970 e destes 23 foram considerados canavieiros naquele ano, com base no método estipulado neste trabalho (municípios responsáveis por 90% da produção estadual de cana-de-açúcar).

A separação dos municípios em clusters gerou dois grupos que são estatisticamente diferentes para quatro das oito variáveis (transformadas) utilizadas. Em 1970, nenhum município do grupo aqui chamado canavieiro tinha instalações sanitárias no estado de Alagoas. O Grupo é entendido como o de municípios com piores indicadores, como mostrado na Tabela 4.15.

Tabela 4.15 – Resultados do Cluster em 2 grupos para Alagoas em 1970

	Grupo 1	Grupo 2	p
Amostra	16	7	-
L de Theil	0,321	0,310	0,667
Pessoas Pobres	88,798	93,575	0
Mortalidade infantil	191,635	199,538	0,113
Analfabetos	67,250	74,014	0,019
IDH	0,272	0,233	0
Esperança de vida	48,682	47,878	0,114
Iluminação elétrica	0,212	0,099	0,001
Instalações sanitárias	0	0	-

O procedimento anteriormente descrito para o estado de São Paulo, baseado na Análise Discriminante, foi aplicado aos dois grupos de municípios canavieiros de Alagoas, no ano de 1970. O resultado do modelo discriminante, apresentado na Tabela 4.16, tem baixo poder de discriminação (apenas 6,25% de eficiência na separação) e nele apenas duas variáveis são incluídas.

No estado de Alagoas a variável “Participação da cana na economia” tem contribuição distinta do que no estado de São Paulo: entre os municípios canavieiros, quanto maior a importância relativa da cana, mais provável a alocação no grupo dos que têm melhores indicadores (Grupo 1). Da mesma forma, os municípios com maior participação da indústria na economia têm melhores resultados.

Tabela 4.16- Resultados da Análise Discriminante com dois grupos para Alagoas em 1970

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Participação da indústria	0,937	-48,297	-53,759	-0,730
Participação da cana na economia	0,937	14,482	11,232	-0,729

Apesar da classificação aparentemente positiva da participação da cana na economia, a sua correlação ainda é negativa, quando estatisticamente significativa, em relação aos indicadores sociais, como mostra a Tabela 4.17.

Tabela 4.17 – Correlação entre indicadores econômicos e sociais – Alagoas 1970

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,573	0,375	0,128	0,431	-0,421	-0,137	-0,465	-0,482
Cana/km ²	-0,278	0,151	-0,093	0,184	-0,163	0,089	0,076	-0,250
Participação da indústria	0,183	-0,521	-0,351	-0,476	0,545	0,355	0,538	0,476
Participação dos serviços	0,183	-0,521	-0,351	-0,476	0,545	0,355	0,538	0,476
Participação da agropecuária	-0,183	0,521	0,351	0,476	-0,545	-0,355	-0,538	-0,476

O gráfico de dispersão entre “IDH” e “Participação da cana na economia” não adiciona qualquer informação, já que o número reduzido de municípios compromete a visualização de claras tendências.

Na Figura 4.9 é apresentada a dispersão geográfica dos municípios canavieiros em Alagoas, em 1970, e seus respectivos grupos, definidos pela Análise de Cluster.

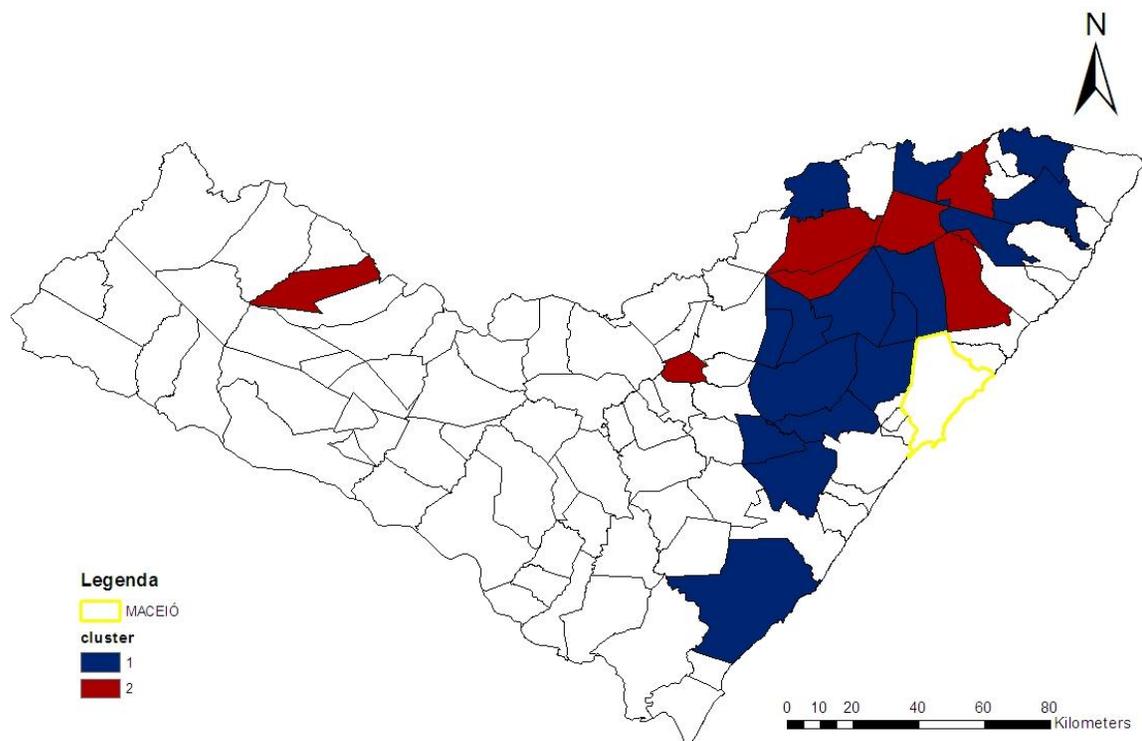


Figura 4.9 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1970.

4.2.2 1980

Em 1980 os municípios canavieiros no estado de Alagoas eram 24, e da Análise de Cluster só três foram alocados no grupo de melhores indicadores (Grupo 1) e 21 no grupo de piores indicadores; há maior concentração de renda nos municípios do Grupo 1. Em 1980 Maceió passou a fazer parte do grupo de municípios maiores produtores de cana, como anteriormente mencionado, e é um dos três do grupo com melhores resultados. Os resultados da Análise Cluster na classificação dos dois grupos são apresentados na Tabela 4.18. Conclui-se que não discernimento estatístico entre os dois grupos quanto às variáveis “Mortalidade infantil” e “Esperança de vida”, e para as demais se pode afirmar que os grupos são estatisticamente distintos.

Tabela 4.18 – Resultados do Cluster em dois grupos para Alagoas, em 1980

	Grupo 1	Grupo 2	p
Amostra	3	21	-
L de Theil	0,506	0,315	0
Pessoas Pobres	54,060	71,775	0
Mortalidade infantil	147,353	155,494	0,521
Analfabetos	41,300	66,461	0,002
IDH	54,060	71,775	0
Esperança de vida	53,500	52,617	0,540
Iluminação elétrica	0,743	0,369	0
Instalações sanitárias	0,070	0,002	0

O resultado da aplicação da Análise Discriminante aos clusters formados indica que as variáveis inseridas ao modelo respondem por 56,1% da divisão dos grupos; a variável “Participação da agropecuária” é a que mais discrimina entre os clusters formados, seguida da “Participação da cana na agricultura” e “Cana por habitante”. Os resultados são apresentados na Tabela 4.19.

Da análise das funções de classificação conclui-se que quanto maior a “Participação da agropecuária” na economia, maior a probabilidade de alocação do município no grupo dos que têm piores indicadores. A mesma conclusão vale para a variável “Participação da cana na agricultura”. Já para as outras duas variáveis que entram no modelo, vale raciocínio oposto.

Tabela 4.19 - Resultados da análise discriminante para dois grupos de municípios, em Alagoas, em 1980.

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Participação da agropecuária	0,862	41,627	106,017	-2,055
Cana/km ²	0,395	0,043	-0,017	0,559
Participação da cana na agricultura	0,510	9,807	18,765	-1,287
Cana/habitante	0,438	-9,936	-20,440	1,205

Para 1980, as correlações entre os indicadores e algumas variáveis de predição também foram calculadas, e os resultados são apresentados na Tabela 4.20. A correlação entre a variável “Participação da cana na economia” e os indicadores sociais também é inversa, ou seja, quanto maior a participação relativa da cana, pior os indicadores.

Tabela 4.20– Correlação entre indicadores econômicos e sociais – Alagoas 1980

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,438	0,519	0,328	0,530	-0,572	-0,324	-0,459	-0,567
Cana/km ²	-0,071	0,046	0,069	0,159	-0,171	-0,066	0,060	-0,155
Participação da indústria	0,542	-0,522	-0,045	-0,743	0,650	0,035	0,608	0,689
Participação dos serviços	0,542	-0,522	-0,045	-0,743	0,650	0,035	0,608	0,689
Participação da agropecuária	-0,542	0,522	0,045	0,743	-0,650	-0,035	-0,608	-0,689

Na Figura 4.10 é mostrada a dispersão espacial dos dois grupos formados. Comparado com a figura para 1970, vários municípios que estavam no grupo de melhores indicadores naquele ano não tiveram melhoria acentuada em seus indicadores.

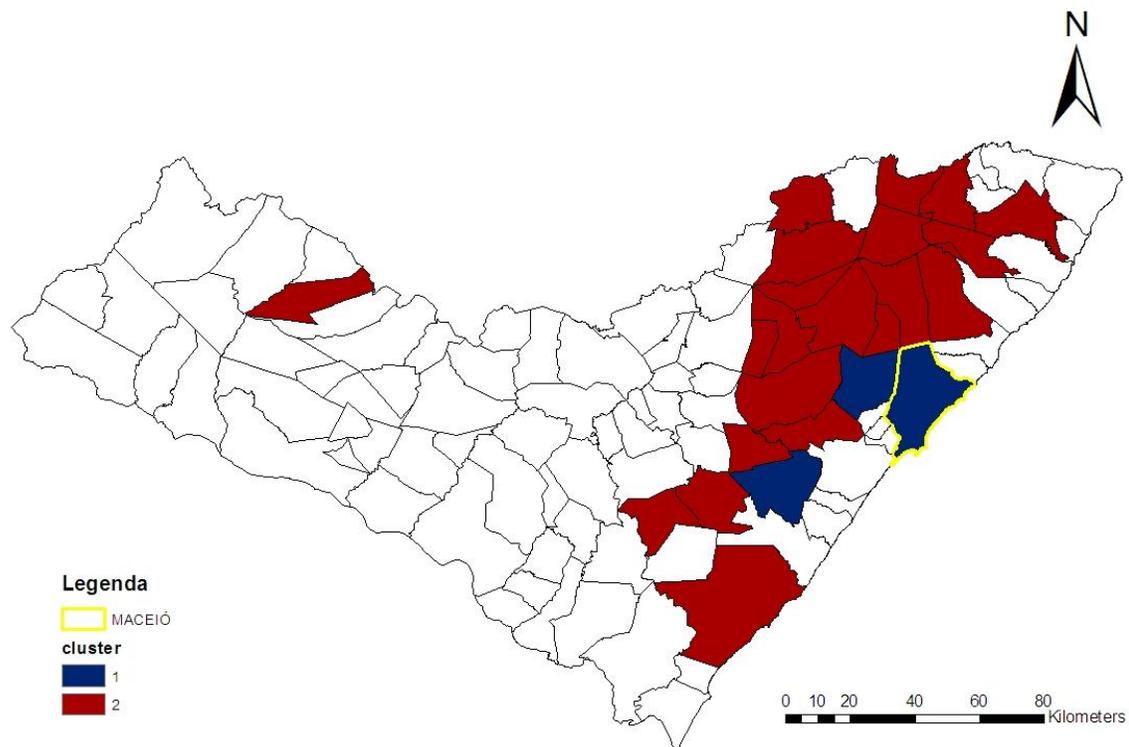


Figura 4.10 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1980

4.2.3 1991

No ano de 1991 existiam 28 municípios considerados canaveiros pelo critério adotado, sendo que 12 foram alocados no grupo de melhores indicadores, o Grupo 1, e 16 no grupo de piores indicadores, o Grupo 2. Das variáveis utilizadas no modelo, “L de Theil” e “Instalações sanitárias” (por pequena margem) não apresentam diferença estatisticamente significativa entre os grupos. A Tabela 4.21 reproduz os resultados obtidos.

Tabela 4.21– Resultados do Cluster em dois grupos para Alagoas, em 1990

	Grupo 1	Grupo 2	p
Amostra	12	16	-
L de Theil	0,436	0,358	0,076
Pessoas Pobres	70,709	80,381	0,004
Mortalidade infantil	62,939	83,952	0,
Analfabetos	49,611	59,656	0,010
IDH	0,531	0,458	0
Esperança de vida	60,244	55,958	0
Iluminação elétrica	0,785	0,675	0,042
Instalações sanitárias	0,055	0,003	0,051

Para aplicar a Análise Discriminante no ano de 1991, outras variáveis estavam disponíveis. Além das variáveis de participação econômica e da produção de cana-de-açúcar no município (relativas à sua área e ao número de habitantes), as variáveis referentes a finanças públicas e de renda nos setores da economia também foram utilizadas. Na Tabela 4.22 pode-se ver que as variáveis incluídas no modelo explicam 54,2% da divisão entre os grupos e que a variável “Transferência estadual” tem a maior relevância na divisão, seguida por “Cana/habitante” e “Participação da agropecuária”.

Tabela 4.22- Resultados da análise discriminante com dois grupos para Alagoas, em 1991

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Renda no comércio	0,492	176,704	166,124	0,511
Participação da agropecuária	0,499	82,839	95,379	-0,674
Transferência estadual	0,554	-1,485	-2,419	0,875
Cana/habitante	0,482	13,927	22,931	-0,839
Cana/km²	0,454	-0,129	-0,195	0,481
Renda nos serviços	0,468	31,792	35,071	-0,479
Renda na indústria	0,457	39,195	42,543	-0,396

Já a análise das funções de classificação indica algumas contradições. Enquanto o coeficiente da variável “Renda no comércio” é maior para o Grupo 1 (o de melhores indicadores), o coeficiente das variáveis “Renda nos serviços” e “Renda na indústria” é maior para a função de classificação do Grupo 2, de piores indicadores. Se analisadas os seus coeficientes das funções de classificação, conclui-se que as médias da renda nesses dois últimos setores são maiores no Grupo 2.

Da análise das correlações entre os indicadores e as variáveis de predição, apresentada na Tabela 4.23, nota-se que a “Participação da cana na economia” municipal apresenta proporcionalidade negativa com os indicadores sociais (exceto o L de Theil), ou seja, quando a participação da cana cresce, o desempenho desses indicadores piora. Por exemplo, o número de analfabetos cresce com a participação da cana, assim como o número de pessoas pobres e a mortalidade infantil. Deve ser lembrado que a correlação representa uma proporcionalidade e o estabelecimento de relação causa e efeito, além de não captar as relações não lineares entre as variáveis.

Tabela 4.23– Correlação entre indicadores econômicos e sociais – Alagoas 1991

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,518	0,438	0,455	0,574	-0,608	-0,468	-0,316	-0,252
Cana/km ²	-0,160	0,007	0,028	0,046	-0,041	-0,051	0,216	0,438
Participação da indústria	0,547	-0,643	-0,069	-0,640	0,612	0,102	0,390	0,122
Participação dos serviços	0,049	-0,031	-0,277	-0,043	0,090	0,266	-0,049	0,260
Participação da agropecuária	-0,509	0,572	0,340	0,582	-0,606	-0,357	-0,277	-0,367

Na Figura 4.11 é apresentado o mapa que indica a distribuição geográfica dos dois grupos, com a capital Maceió demarcada com borda vermelha. Nota-se que, comparativamente a 1980, há maior dispersão dos municípios que têm melhores indicadores.

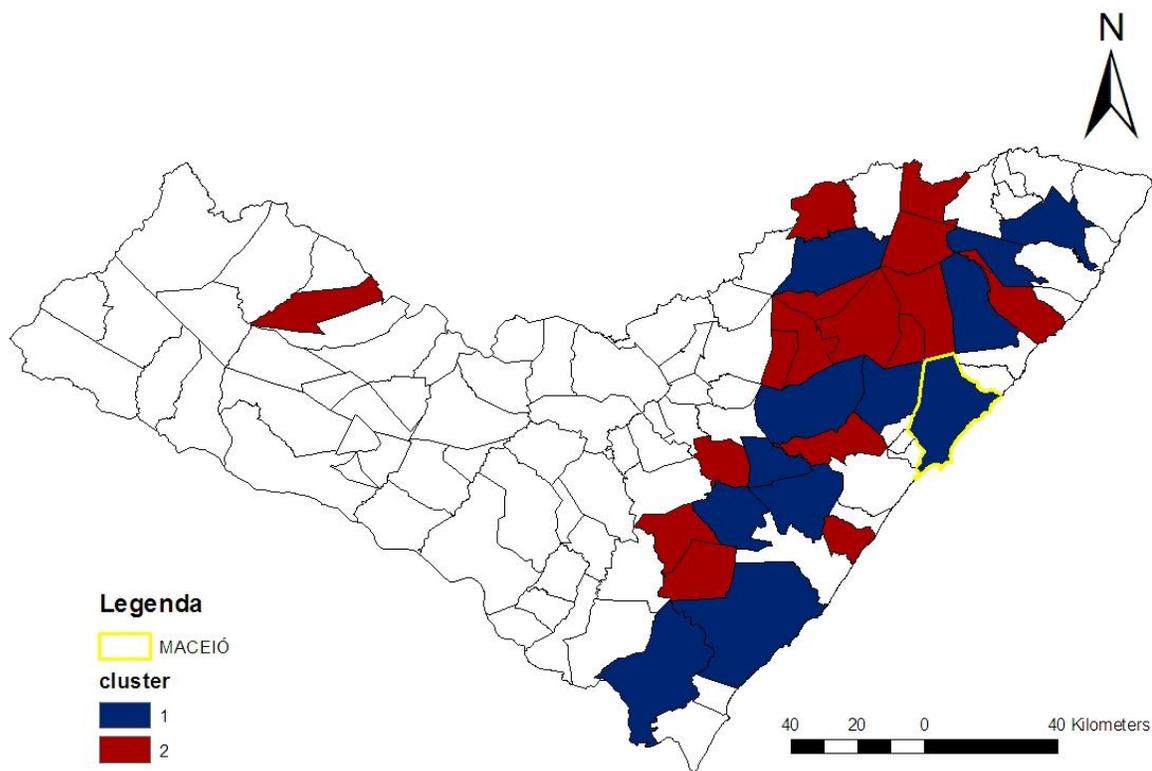


Figura 4.11 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 1991

4.2.4 2000

Em 2000, segundo o critério adotado, 30 municípios foram alocados no grupo canavieiro no estado de Alagoas, inclusive Maceió, da Análise Cluster, 17 foram classificados no Grupo 1, por terem melhores indicadores, e 13 no grupo de piores indicadores, o Grupo 2. A variável que não há diferença estatística entre os grupos é “L de Theil”. Na Tabela 4.24 são apresentados os valores das médias de cada grupo e o resultado da análise de variância.

Tabela 4.24 – Resultados do Cluster em dois grupos para Alagoas, em 2000

	Grupo 1	Grupo 2	p
Amostra	17	13	-
L de Theil	0,514	0,469	0,249
Pessoas pobres	68,457	78,057	0
Mortalidade infantil	42,932	62,373	0
Analfabetos	37,279	46,531	0
IDH	0,622	0,551	0
Esperança de vida	65,250	60,558	0
Iluminação elétrica	0,908	0,799	0
Instalações sanitárias	0,157	0,128	0

Novamente, as médias apresentadas na Tabela 4.24 foram calculadas a partir das variáveis sem transformação, porém a análise foi feita com as variáveis transformadas.

Da Análise Discriminante, que tem seus resultados apresentados na Tabela 4.25, identifica-se que a variável “Cota-parte do fundo de participação – municipal” é a que mais contribui para discriminação entre os grupos, seguida da variável “Cana/km²” e “Renda na indústria”. O modelo discriminante explica 56,7% da divisão entre os grupos. Os coeficientes das funções de classificação indicam que as variáveis relacionadas à produção da cana-de-açúcar colaboram mais para a classificação no grupo de municípios com melhores indicadores, já que o coeficiente da função do Grupo 1 é maior que o coeficiente para o Grupo 2.

Tabela 4.25- Resultados da análise discriminante com dois grupos para Alagoas, em 2000

Variáveis	Wilks-lambda	Função de classificação 1	Função de classificação 2	Função discriminante
Receita total	0,414	-789,16	-783,64	0,482
Cana/km²	0,492	6,19	4,82	-0,707
Participação dos serviços	0,456	435,27	420,45	-0,552
Renda na indústria	0,496	8,99	8,78	-0,663
Cota federal	0,473	2057,52	2029,47	-1,509
Cana/habitante	0,432	717,75	709,16	-0,593

Na Figura 4.12 é apresentada a distribuição geográfica dos dois grupos de municípios canavieiros em Alagoas, em 2000. Comparado com o mapa de 1970 (Figura 4.9), nota-se que os municípios com piores indicadores voltaram a ter concentração muito parecida em 2000.

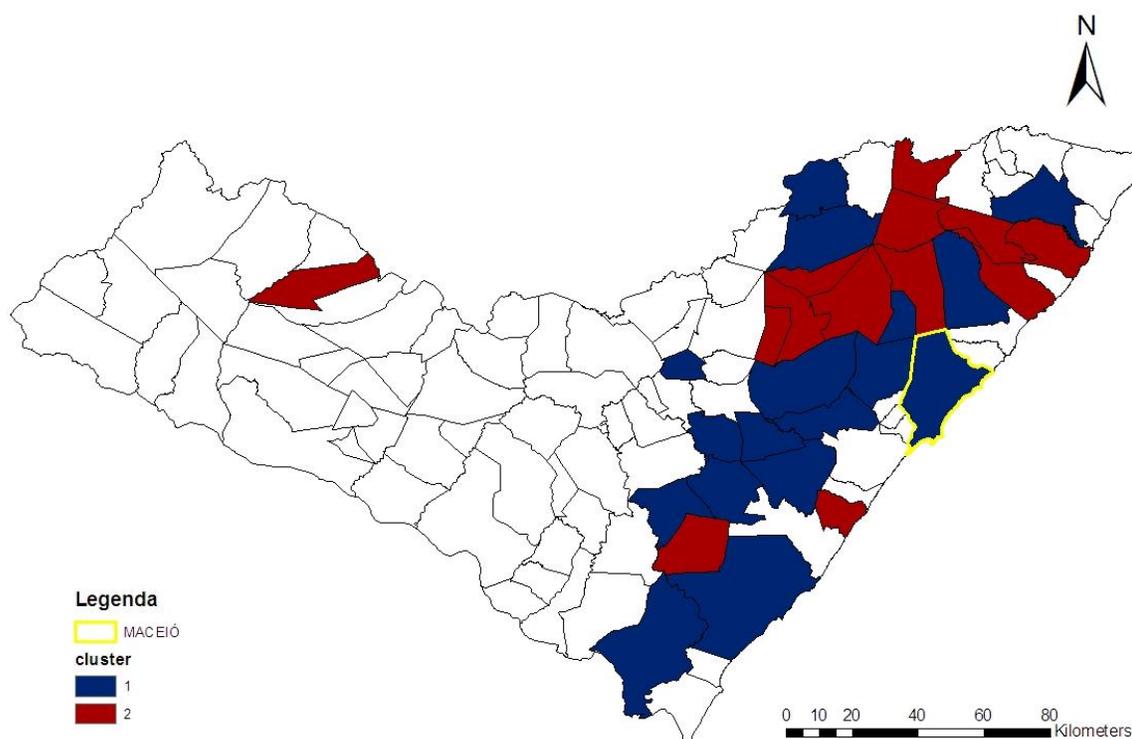


Figura 4.12 - Distribuição dos municípios nos dois grupos de indicadores de qualidade de vida no estado de Alagoas, em 2000

Na Tabela 4.26 apresenta as correlações entre os indicadores econômicos e sociais e variáveis de predição. Novamente, a variável “Participação da cana na economia”, quando estatisticamente significativa, apresenta correlação negativa com os indicadores sociais, assim como a “Participação da agropecuária”. As participações da indústria e dos serviços, por sua vez, apresentam correlação positiva, com exceção do L de Theil.

Tabela 4.26 – Correlação entre indicadores econômicos e sociais – Alagoas 2000

Variável	L de Theil	Pessoas pobres	Mortalidade infantil	Analfabetos	IDH	Esperança de vida	Iluminação elétrica	Instalações sanitárias
Cana na economia	-0,330	0,545	0,294	0,599	-0,532	-0,294	-0,194	-0,180
Cana/km ²	-0,323	-0,048	-0,518	-0,163	0,263	0,539	0,240	0,027
Participação da indústria	-0,113	-0,279	-0,412	-0,327	0,314	0,419	-0,030	0,234
Participação dos serviços	0,531	-0,428	0,019	-0,427	0,376	-0,038	0,478	-0,011
Participação da agropecuária	-0,297	0,630	0,418	0,680	-0,625	-0,411	-0,343	-0,237

Ao contrário do observado e concluído para São Paulo, da Tabela 4.27 pode-se concluir que no estado de Alagoas a relação entre a participação da cana na economia local e o estágio de desenvolvimento, neste caso analisado pelo distinto perfil dos grupos, não é tão direta. Somente nos anos de 1980 e 1991 existe diferença estatística entre os dois grupos, e quanto maior a “Participação da cana na economia”, piores os indicadores. O mesmo pode ser concluído quanto à variável “Participação da agropecuária”. Só em 1980 pode-se concluir que quanto maior a “Participação da indústria” e a “Participação dos serviços”, melhores os indicadores, o que corrobora com a hipótese de que a maior diversificação da economia auxilia na melhora dos indicadores. Estatisticamente, só em 1980 é possível concluir que os municípios canavieiros com maior população rural têm piores indicadores. Cabe notar que comparação das médias usando o teste *t* não foi feita com o rigor estatístico necessário, o que pode implicar o aumento do erro tipo I, que implica rejeitar a hipótese de igualdade, quando na verdade são iguais. Portanto, a rigor a diferença entre os dois grupos de municípios canavieiros em Alagoas pode ser ainda menor do que o que foi acima apresentado.

Tabela 4.27 – Médias das participações econômicas e da população rural em Alagoas

Variável	1970		1980		1991		2000	
	Grupo melhor	Grupo pior						
Participação cana	0,50±0,18	0,41±0,14	0,27±0,29	0,54±0,19	0,14±0,06	0,24±0,11	0,15±0,07	0,18±0,07
Participação indústria	0,29±0,09	0,25±0,05	0,41±0,10	0,26±0,06	0,41±0,16	0,38±0,06	0,21±0,11	0,16±0,16
Participação serviços	0,29±0,09	0,25±0,05	0,41±0,10	0,26±0,06	0,31±0,08	0,23±0,16	0,51±0,11	0,48±0,08
Participação agropecuária	0,41±0,18	0,49±0,10	0,18±0,20	0,47±0,12	0,28±0,12	0,38±0,13	0,28±,11	0,35±0,14
Porcentagem população rural	0,71±0,13	0,79±0,09	0,34±0,28	0,64±0,12	0,45±0,19	0,51±0,15	0,38±0,19	0,38±0,14

Também em Alagoas os indicadores relativos à produção de cana-de-açúcar também tiveram importância nos modelos discriminantes e, conclui-se que importantes para explicar a existência de municípios com melhores e piores indicadores. Ao contrário do verificado no estado de São Paulo, a variável “Participação da cana na economia” é menos relevante.

Outras variáveis importantes nos modelos discriminantes são aquelas que refletem o perfil das finanças públicas, isso quando disponíveis. Nesses casos, nos modelos ajustados, essas variáveis são as mais importantes na diferenciação entre os grupos. As variáveis de renda também tiveram sua importância na diferenciação entre os grupos, porém, em alguns casos, o grupo de melhor desempenho não apresenta maior renda.

De forma geral, as correlações entre os indicadores e as variáveis econômicas de predição resultam em conclusões similares em Alagoas e no estado de São Paulo, quando estatisticamente significativos. A importância relativa das atividades canieira e agropecuária são inversamente proporcionais aos indicadores, e a importância relativa das atividades de serviços e industrial são diretamente proporcionais aos indicadores sociais. Portanto, a conclusão é que nos municípios canieiros, quanto maior a diversificação da atividade econômica, melhores são os indicadores socioeconômicos.

5. Conclusões e Recomendações

5.1. Conclusões

O objetivo desta dissertação era estudar aspectos socioeconômicos relacionados à atividade canavieira nos principais municípios produtores nos estados de São Paulo e Alagoas (principais produtores em suas respectivas regiões). Para isso, foram utilizados os dados disponibilizados pelo IPEADATA, que têm origem nos censos realizados pelo IBGE nos anos de 1970, 1980, 1991 e 2000.

Os indicadores sociais utilizados nesta dissertação foram escolhidos em função dos objetivos da avaliação feita, e são exatamente iguais ou similares aos indicadores sociais mais importantes, segundo a bibliografia utilizada como referência. Também foram consideradas a facilidade de acesso aos indicadores, o que permite que análises similares possam ser feitas sem restrições, a disponibilidade em nível municipal, o que permite a comparação entre localidades específicas, e a periodicidade, o que permite a análise da evolução temporal.

Esta dissertação foi desenvolvida visando dar continuidade ao trabalho de Oliveira (2011), que tratou do mesmo tema. O principal objetivo de avanço em relação aquele trabalho foi posto na análise com maior rigor estatístico, tanto para que se tenha maior precisão dos resultados quanto para que sejam definidos procedimentos metodológicos que possam ser replicados.

Os principais métodos aplicados, grande parte deles estatísticos, são descritos a seguir:

- Análise e tratamento dos dados (indicadores), para verificação do cumprimento ou não de condições necessárias para a aplicação dos métodos estatísticos, principalmente quanto à normalidade da distribuição e à homogeneidade. Quando os dados originais não atenderam as condições, o que em geral ocorreu, foram aplicadas funções de transformação.
- Teste de hipótese t , na comparação dos grupos de municípios canavieiros e não canavieiros, através da comparação dos valores médios dos indicadores, com o intuito de identificar se há diferenças significativas entre os grupos, e se há vantagem ou desvantagem dos municípios produtores de cana.
- Análise de Cluster: A Análise de Cluster foi empregada para separar, de forma não tendenciosa, os municípios canavieiros em grupos de municípios com melhores e piores indicadores sociais.
- Análise Discriminante: Foi empregada com o intuito de identificar quais fatores explicam o maior ou menor estágio de desenvolvimento dos municípios canavieiros, tendo por base, para caracterizar o desenvolvimento, os indicadores socioeconômicos considerados. As variáveis utilizadas na Análise Discriminante são econômicas, de participação dos setores na economia do município, de finanças públicas e, para estudar a presença da

atividade sucroenergética, foram usados também indicadores de produção de cana-de-açúcar.

- Gráficos radar: Os gráficos radar foram feitos para facilitar a análise da evolução dos indicadores sociais ao longo do tempo. Foram feitas três comparações: primeiro, uma comparação direta entre municípios canavieiros e não canavieiros, depois uma comparação entre municípios que se tornaram canavieiros em 1980 e municípios semelhantes que nunca se tornaram canavieiros e, finalmente, entre grupos de municípios, consideradas regiões específicas, comparando municípios semelhantes nos quais a cana se tornou importante e outros nos quais a cana nunca teve importância.
- Mapas dos estados, com destaque para a distribuição de indicadores: na medida do possível, para facilitar a análise e ter-se noção de aspectos geográficos, foram feitos mapas que permitem a visualização da dispersão ou da concentração de indicadores dos municípios.

No que diz respeito aos resultados obtidos, os mesmos devem ser analisados de acordo com os objetivos relacionados acima.

Na comparação entre os municípios canavieiros e não canavieiros no estado de São Paulo, os municípios que apresentaram representativa produção de cana-de-açúcar têm, em todos os anos estudados, melhores indicadores sociais e econômicos do que os municípios semelhantes nos quais a cana não têm importância. Esta conclusão deve ser relativizada, já que, pelos testes estatísticos aplicados, grande parte das variáveis estudadas não apresentaram normalidade mesmo depois de aplicadas as transformações indicadas. Quando as variáveis não têm distribuição normal, a probabilidade de se cometer o erro tipo I aumenta. O erro tipo I corresponde rejeitar a hipótese de igualdade, quando na verdade são iguais. Ou seja, quando se diz que os municípios canavieiros e não canavieiros são diferentes, por terem diferentes médias dos indicadores, tendo por base o teste t , pode-se estar cometendo o erro do tipo I. No caso, a conclusão de que os municípios canavieiros de São Paulo têm, para todos os anos, melhores indicadores do que os congêneres não canavieiros, deve ser relativizada.

No caso do estado de Alagoas, os municípios canavieiros têm indicadores superiores aos municípios não canavieiros em todos os anos, porém não há diferença estatística em todas as situações. Para Alagoas, neste trabalho foram observadas algumas diferenças em relação aos resultados de Oliveira (2011), o que pode ser explicado pelas transformações que foram aplicadas para atingir normalidade das variáveis. No entanto, cabe a ressalva de que os testes de normalidade aqui aplicados não têm bons resultados no caso de amostras com menos de 30 indivíduos, como é o caso da aplicação em Alagoas.

Como conclusão da primeira parte, se não é possível afirmar categoricamente que os municípios canavieiros de São Paulo e Alagoas têm, na média, melhores indicadores socioeconômicos do

que municípios similares não canavieiros, é possível concluir o oposto, ou seja, que os municípios canavieiros não têm piores indicadores do que os não canavieiros.

O estudo da evolução dos indicadores de 1970 a 2000 permite concluir que há tendência de equalização dos municípios considerados canavieiros e não canavieiros no que diz respeito aos indicadores considerados. Em São Paulo, os municípios canavieiros apresentam melhores indicadores do que os congêneres e, com o passar dos anos, essa vantagem dos municípios canavieiros é perdida. Por um lado, há expansão do número de municípios canavieiros, com incorporação de municípios que tinham menor grau de desenvolvimento. Mas a hipótese mais provável para a equalização é que a atividade canavieira, que era fator diferencial em 1970, passou a ter menor importância relativa, com o desenvolvimento econômico do país e com políticas públicas voltadas a disseminação dos serviços de saúde e de educação.

Também em Alagoas se observa a tendência de equalização dos dois grupos de municípios, o que leva a concluir que do ponto de vista socioeconômico a tendência é a redução das vantagens relativas dos municípios canavieiros. Há que se ressaltar, entretanto, que os municípios canavieiros não têm vantagem em relação aos não canavieiros em todos os anos, e para todos os indicadores. O uso dos Gráficos Radar foi o recurso utilizado para visualização dos resultados, mas é inadequado para uma análise comparativa mais detalhada, seja pela baixa resolução, seja por não ser possível verificar a eventual redução da dispersão dos resultados. Então, outros métodos estatísticos, como séries temporais, podem ajudar no entendimento das questões não estáticas que influenciam o desenvolvimento dos municípios.

Uma das questões que se pretendeu responder é se os municípios que se tornaram canavieiros em algum instante, por exemplo, entre 1970 e 1980, quando da primeira expansão da produção de cana, tiveram vantagem ou desvantagem em relação a municípios similares nos quais a cana jamais teve importância econômica. Essa análise só feita para São Paulo, pois em Alagoas não houve substancial variação dos municípios canavieiros entre 1970 e 2000. Definidos os grupos objeto de comparação, constatou-se que os novos municípios canavieiros (em 1980 e 1991) já tinham melhores indicadores em 1970. Uma possível explicação é que a cana tenha substituído atividades agrícolas que já resultavam vantagens desses municípios. De qualquer forma, com o passar do tempo a tendência foi a aproximação dos resultados dos dois grupos, ou seja, verifica-se que a importância da atividade canavieira não trouxe vantagens significativas ao longo do tempo, em relação aos municípios congêneres não canavieiros.

O procedimento de análise também foi aplicado na comparação de pequenos grupos de municípios, geograficamente próximos, na tentativa de homogeneizar a base de comparação e se comparar municípios para os quais aspectos locais podem ser relevantes. Nesse caso, como os grupos eram de poucos municípios (em alguns casos, dois ou três), não foi possível realizar um estudo estatístico rigoroso, e a análise feita é considerada apenas um teste. Os resultados mostram que, em geral, há tendência de aproximação dos grupos quanto aos indicadores. Porém, em alguns casos o grupo dos municípios que nunca se tornaram canavieiros tem indicadores

superiores, o que indica que análises comparativas devem ser feitas com atenção, e os grupos objeto de comparação devem ser criteriosamente definidos.

Finalmente, os resultados das Análises de Cluster e Discriminante permitiram alcançar algumas conclusões importantes a respeito das diferenças que existem entre os municípios canavieiros, do ponto de vista dos indicadores socioeconômicos aqui considerados.

Uma primeira conclusão é que a separação dos municípios em dois grupos, pela Análise de Cluster, é lógica no que diz respeito aos indicadores socioeconômicos. Sempre foi possível a identificação de um grupo de municípios com melhores indicadores – em todos os casos, exceto o L de Theil – e outro de municípios com piores indicadores.

Já os modelos ajustados pela Análise Discriminante foram invariavelmente pouco eficientes considerado o objetivo de distinção dos grupos, embora não fosse este o propósito nesta dissertação. A hipótese é que outros fatores além dos aqui testados, e alguns inclusive não quantificáveis ou não facilmente expressos em números, afetam o desenvolvimento dos municípios (aqui expresso pelos indicadores socioeconômicos considerados). Tais fatores podem ser históricos, geográficos, políticos, culturais e demográficos. Assim, na continuidade da pesquisa é recomendável a análise específica de alguns casos, inclusive com levantamento de campo.

Por outro lado, a Análise Discriminante foi aplicada nesta dissertação com o objetivo de se identificar fatores comuns entre os municípios canavieiros que têm melhores e piores indicadores, em São Paulo e em Alagoas, nos anos analisados. Em São Paulo e Alagoas, de uma forma geral, os municípios que têm melhores indicadores têm a atividade econômica mais diversificada, ou seja, quanto maior a participação relativa de outros setores na economia (indústria e serviços), além da agropecuária, melhores os resultados.

Especificamente quanto às variáveis que indicam a importância relativa da atividade canavieira, conclui-se que a presença da cana-de-açúcar não impacta negativamente os indicadores de qualidade de vida do município. Entretanto, a maior dependência da economia local da atividade canavieira está associada, em geral, à menor qualidade de vida.

Também como regra geral para os dois estados, e obviamente, os municípios canavieiros com maior receita têm melhores indicadores. Da mesma forma, municípios que proporcionalmente recebem mais transferências estaduais e federais têm melhores indicadores.

Para compreender melhor a relação dos indicadores sociais e aspectos específicos dos municípios relacionados, por exemplo, com a estrutura da economia local, foram calculadas as correlações entre as variáveis. Tais resultados também indicam tendências gerais que reforçam as conclusões acima mencionadas: quanto maior a importância da agricultura na economia local, e mais especificamente da cana-de-açúcar, piores os indicadores sociais; e quanto maior a participação relativa dos setores de serviços e indústria na economia local, melhores os indicadores sociais.

Infelizmente, na análise feita nesta dissertação não foi possível incluir os dados de 2010, pelo fato de não terem sido disponibilizados a tempo. Com isso não foi possível avaliar a tendência mais recente e os impactos da expansão da atividade canavieira na primeira década do século XXI. Incluir 2010 à análise é, claro, muito importante.

Outro aperfeiçoamento que deve ser realizado na análise que foi feita nesta dissertação é a consideração do impacto de outras atividades relacionadas ao setor canavieiro, além da produção de cana. Evidentemente, precisa ser analisado o eventual impacto da existência de unidades industriais (usinas), e tal avaliação pode ser feita com alguma facilidade. Já é menos simples, em princípio, mas igualmente relevante, a avaliação dos impactos de outras atividades relacionadas ao setor, como a prestação de serviços e a indústria de máquinas e componentes.

5.2. Sugestões

Algumas sugestões podem ser feitas para trabalhos futuros no mesmo tema, levando em consideração as lacunas deixadas nesta pesquisa e as limitações da metodologia empregada. São elas:

- Para que possam ser considerados aspectos regionais que muito possivelmente impactam os indicadores de municípios fisicamente próximos, a metodologia conhecida como “*Propensity score*” deve ser utilizada, como mencionada por Chagas (2010);
- Alternativamente, as municipalidades e seus indicadores devem ser analisados com o emprego de metodologias da estatística espacial, para que possam ser consideradas mais explicitamente a influência e a distribuição espacial dos municípios;
- A análise deve ser complementada com a consideração das municipalidades que têm usinas instaladas, bem como outras atividades econômicas relacionadas ao setor sucroenergético;
- A análise apresentada nesta dissertação deve ser feita para o ano de 2010, a partir da liberação dos dados censitários pelo IBGE;
- A análise de evolução dos indicadores deve ser feita com o emprego de análise de dados em painel.
- Pesquisa de campo para entender as variáveis não mensuráveis que influenciam o desenvolvimento do município.

6. Referências bibliográficas

ACTION. **Distribuição Normal**. Disponível em <www.portalaction.com.br>. Acesso em: 05 de Dezembro de 2011.

BACCARIN, José Giacomo et al. Avanço da Mecanização da Colheita de Cana-de-açúcar e Alteração na Composição da Ocupação Sucroalcooleira no Estado de São Paulo, Entre 2007 e 2010. In: CONGRESSO SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 49., 2011, Belo Horizonte. **Grupo de Pesquisa: Evolução e Estrutura da Agropecuária no Brasil**. Belo Horizonte: Sober, 2011. p. 1 - 15.

BARBOSA, Maria Luiza. Social responsibility and benefits. In: UNICA. **Sugarcane's Energy**. São Paulo: Berlendis, 2008. p. 199-236.

BELLEN, Hans Michael Van. **Indicadores de Sustentabilidade: uma análise comparativa**. Rio de Janeiro: FGV, 2005.

CAMARGO, Alceu Salles; TONETO, Rudinei. Indicadores socioeconomicos e a cana-de-açúcar no estado de São Paulo. **1º Workshop do Observatório do Setor Sucroalcooleiro**, Piracicaba, n. , p.20-34, 01 abr. 2008.

CAMPEÃO, Patrícia. Apresentação Oral - Desenvolvimento Rural, Territorial e Regional. In: CONGRESSO SOBER, 47., 2009, Porto Alegre. **Influência do setor sucroalcooleiro no Índice de Desenvolvimento Humano Municipal (IDH-M)**. Porto Alegre: Sociedade Brasileira de Economia, Administração e Sociologia Rural, 2009. p. 1 - 15.

CAUMO, Alessandra Juliana et al. O corte manual da cana-de-açúcar sob uma perspectiva de gênero: um estudo de caso no município de Mirandópolis - SP. In: CONGRESO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 49., 2011, Belo Horizonte. **Grupo de Pesquisa 9: Políticas Sociais para o Campo**. Belo Horizonte: Sober, 2011. p. 1 – 19

CHAGAS, André Luiz Squarize. **Impacto da produção de cana-de-açúcar sobre as condições sociais das regiões canavieiras**. Disponível em: <www.fipe.org.br>. Acesso em: 01 jul. 2011.

COLANTONIO, Andrea. **Social Sustainability: An Exploratory Analysis of its Definition, Assessment Methods, Metrics and Tools**. Oxford: Oisd, 2007.

FERNANDES, Carlândia Brito Santos et al. O mercado de trabalho formal no setor sucroalcooleiro no Brasil. In: CONGRESSO SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 49., 2011, Belo Horizonte. **Grupo de Pesquisa 10: Desenvolvimento Rural, Territorial e Regional**. Belo Horizonte: Sober, 2011. p. 1 - 15.

GODOY, Ricardo A. et al. Do Markets Worsen Economic Inequalities?: Kuznets in the Bush. **Human Ecology**, Santa Barbara, v. 32, n. 3, p.340-364, 1 jun. 2004.

HAIR, Joseph F. et al. **Multivariate data analysis**. 7. ed. Cornell University: Pearson Prentice Hall, 2010.

HART, Maureen. **Guide to Sustainable Community Indicators**. North Andover: Hart Environmental Data, 1999.

HOFFMANN, Rodolfo. Income of people engaged in Brazil's sugar cane agribusiness. In: UNICA. **Sugarcane**. São Paulo: Berleandis, 2008. p. 199-236.

HOFFMANN, Rodolfo; OLIVEIRA, Fabíola Cristina Ribeiro de. Apresentação oral. In: SOBER, 46., 2008, Rio Branco. **Remuneração e características das pessoas ocupadas na agroindústria canieira no Brasil de 2002 a 2006**. Rio Branco: Sociedade Brasileira de Economia, Administração e Sociologia Rural, 2008. p. 1 - 20.

IBGE. **Matriz Insumo Produto**. Disponível em: <ftp://ftp.ibge.gov.br/Contas_Nacionais/Matriz_de_Insumo_Produto/>. Acesso em: 10 jan. 2012.

IEA. **Energy balances of non-OECD countries**. Paris: International Energy Agency, 2010.

IPCC. **Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change**. Geneva: Ipcc 104, 2007.

IPCC. **Special Report on Renewable Energy Sources and Climate Change Mitigation. The Fifth Assessment Report**. Cambridge: Cambridge University Press, 2011.

IPEADATA. **Dados sociais e regionais do Brasil**. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso em: 10 jan. 2012.

JANNUZZI, Paulo de Martino. **Indicadores sociais no Brasil**. Campinas: Alínea, 2001.

KAUFMAN, Leonard; ROUSSEEUW, Peter J.. **Finding Groups in Data: An Introduction to Cluster Analysis**. Bruxelas: Wiley-interscience, 1990.

KRZANOWSKI, Wojtek. **Statistical Principles and Techniques in Scientific and Social Research**. Exeter: Oxford University Press, 2007.

MAGGINO, Filomena. **The state of the arts in indicators construction**. Firenze: Università Degli Studi Firenze, 2007.

MEADOWS, Donella. **Indicators and information systems for sustainable development**. Hartland Four Corners: The Sustainable Institute, 1998.

MENDONÇA, Maria Luisa. **A OMC e os efeitos destrutivos da indústria da cana no Brasil**. Disponível em: <www.social.org.br/cartilha_rede_omc_novo_formato.pdf>. Acesso em: 01 jul. 2011.

MORAES, Marcia Azanha Ferraz Dias de et al. Externalidades sociais dos combustíveis. In: UNICA. **Etanol e Bioeletricidade: A cana-de-açúcar no futuro da matriz energética**. São Paulo: Luc Projetos de Comunicação, 2010. p. 48-74., 2010. p. 18-43.

- MORAES, Márcia Azanha Ferraz Dias de. Indicadores do Mercado de Trabalho do Sistema Agroindustrial da Cana-de-Açúcar do Brasil no Período 1992-2005. **Estudos Econômicos**, São Paulo, v. 37, n. 4, p.875-902, 01 out. 2007.
- MORAES, Márcia Azanha Ferraz Dias de. Number and Quality of Jobs in the sugarcane agribusiness. In: UNICA. **Sugarcane's Energy**. São Paulo: Berlendis, 2008. p. 199-236.
- NEVES, Marcos Fava; TROMBIN, Vinicius Gustavo; CONSOLI, Matheus. O mapa sucroenergetico do Brasil. In: UNICA. **Etanol e Bioeletricidade: A cana-de-açúcar no futuro da matriz energética**. São Paulo:
- NOVAES, José Roberto Pereira. Campeões de produtividade: dores e febres nos canaviais paulistas. **Estudos Avançados**, São Paulo, v. 59, n. 21, p.167-177, 01 abr. 2007.
- OIT. **Convenção sobre trabalho forçado ou obrigatório**. Genebra, 1930.
- OLIVEIRA, Janaína Garcia. **Estudo dos impactos socioeconômicos regionais do setor sucroalcooleiro**. 2011. 211 f. Tese (Doutorado) - Unicamp, Campinas, 2011.
- ORTIZ, Lúcia. **Despoluindo Incertezas: Impactos Territoriais da Expansão das Monoculturas Energéticas no Brasil e Replicabilidade de Modelos Sustentáveis de Produção e Uso de Biocombustíveis**. Belo Horizonte: Instituto Vitae Civilis, 2007.
- PRESCOTT-ALLEN, Robert. **The wellbeing of nations: a country-by-country index of quality of life and the environment**. Washington: Island Press, 2001.
- RAYKOV, Tenko; MARKOULIDES, George A.. **An Introduction to applied multivariate analysis**. Nova Iorque: Routledge, 2008.
- RAZALI, Nonadiah Mohd; WAH, Yap Bee. Power comparisons of Shapiro-Wilk, Kolmogorov-Smornov, Lilliefors and Anderson-Darling tests. **Journal Of Statistical Modeling And Analytics**, Malasia, p. 21-33. 1 jan. 2011.
- RENCHER, Alvin C.. **Methods of Multivariate Analysis**. 2. ed. Brigham Young University: Wiley-interscience, 2002.
- REPÓRTER BRASIL. **O Brasil dos agrocombustíveis: Impactos das lavouras sobre a terra, o meio e a sociedade**. São Paulo, 2010.
- SILVA, Roseli. Setor sucroalcooleiro no estado de São Paulo: mensurando impactos socioeconômicos. **1º Workshop do Observatório do Setor Sucroalcooleiro**, Piracicaba, n., p.1-16, 01 abr. 2008.
- SIRKIN, R. Mark. **Statistics for the social sciences**. 3. ed. California: Sage, 2006.
- STEVENS, James P.. **Applied multivariate statistics for the social sciences**. 5. ed. University Of Cincinnati: Routledge, 2009
- TABACHNICK, Barbara C.; FIDELL, Linda S.. **Using Multivariate Analysis**. 3. Ed. Northridge: Pearson, 2007.

TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. Cluster Analysis: Basic Concepts and Algorithms. Em: TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. Michigan State University: Michigan State University, 2006. p. 487-588.

TEIXEIRA, Monica La Porte; FREITAS, Rosa Maria Vieira de. Acidentes do Trabalho Rural no Interior Paulista. **São Paulo em Perspectiva**, São Paulo, n. , p.81-90, 01 abr. 2003.

TINSLEY, Howard E.a.; BROWN, Steven D. (Ed.). **Handbook of Applied Multivariate Statistics and Mathematical Modeling**. Eua: Academic Press, 2000.

UNICA. **Relatório de Sustentabilidade 2010**. São Paulo, 2011. 128 p.

VIEIRA, Gilberto. **Avaliação do custo, produtividade e geração de emprego no corte de cana-de-açúcar, manual e mecanizado, com e sem queima prévia**. 2003. 127 f. Dissertação (Mestrado) - Unesp, Botucatu, 2003.

WONNACOTT, Ronald J.; WONNACOTT, Thomas H.. **Econometria**. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

WONNACOTT, Thomas H.; WONNACOTT, Ronald J.. **Estatística aplicada à economia e à administração**. Rio de Janeiro: Livros Técnicos e Científicos, 1981.

Apêndice A – Transformação dos indicadores socioeconômicos

Pelo fato de que os dados dos indicadores originalmente não atendiam as condições necessárias à aplicação do teste t , nesta dissertação foram feitas transformações em todas as variáveis, seguindo os procedimentos descritos por Tabachnick e Fidel (2007) e Stevens (2009). A transformação feita aos dados do IDH do ano de 1991 para o estado de Alagoas é aqui tomado como exemplo, por ter se ajustado a distribuição normal após a aplicação da transformação pela raiz quadrada. Os dados de IDH inicialmente apresentavam *skewness* levemente positiva de 0,4062 e *kurtosis* também positiva de 0,2005, o que indica a utilização de raiz quadrada como método de transformação. Após a transformação, os parâmetros acima mencionados passaram a ser 0,1765 e -0,0028, respectivamente, que é uma melhoria significativa. Os resultados indicam que os parâmetros transformados têm distribuição normal, segundo o teste W de *Shapiro-Wilk*. Na Figura A1, são apresentadas as distribuições do IDH antes e depois da transformação.

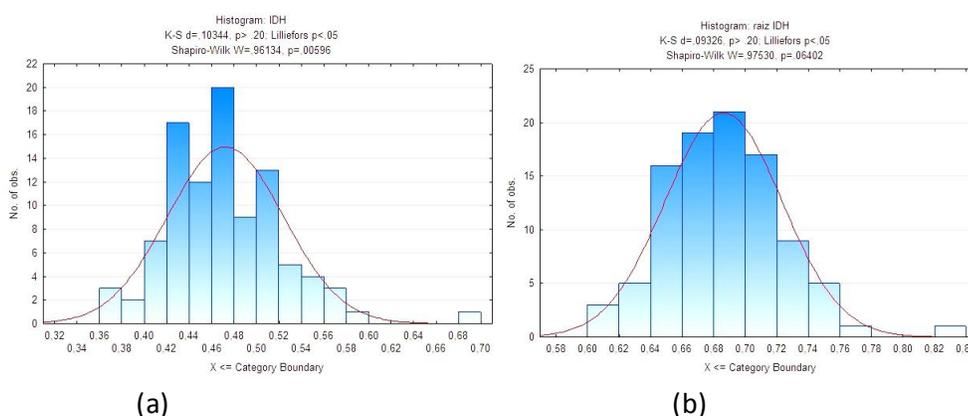


Figura A1 – (a) Histograma da variável IDH antes da transformação e; (b) depois da transformação (raiz do IDH).

É importante notar que os valores Z de *skewness* e de *kurtosis*, calculados a partir das equações 6 e 8, apresentadas na seção de revisão metodológica, também podem ser usados como testes de normalidade. A um nível de significância de 1%, valores de z que excedem o valor crítico de $\pm 2,58$ devem ser considerados como indicativos de distribuição não normal em termos daquela característica (*skewness* ou *kurtosis*) (HAIR et al., 2010). Assim, com relação a *skewness* e a *kurtosis*, no caso do estado de São Paulo, cinco das variáveis transformadas podem ser consideradas como tendo distribuição normal, sendo as exceções as variáveis “Instalações sanitárias”, “Iluminação elétrica” e “Mortalidade infantil”.

Outra suposição requerida pelo teste t é a de homogeneidade de variância (homoscedasticidade). O teste de Levene é usado para testar a suposição. Quando significativo, a hipótese de variâncias homogêneas deve ser rejeitada e, para esta dissertação, o nível de significância do teste foi definido em 5%.

A violação das suposições de normalidade e homogeneidade de variância não impede a aplicação do teste *t*, mas nesse caso os resultados devem ser analisados com cuidado. Apesar de robusto, quando a variável apresenta *kurtosis* positiva o valor teste tende a ser menor, o que impede a rejeição da hipótese testada. O oposto acontece quando da ocorrência de *kurtosis* negativa. No caso de *skewness*, o teste não é afetado (STATSOFT, 2011). Stevens (2009) comenta que o teste pode ser considerado robusto no caso de violação da homogeneidade de variância quando o tamanho dos grupos for parecido, próximo a 1,5. Porém, quando os grupos têm número de observações/indivíduos diferentes (neste caso, municípios) e a diferença de variância ocorre em pequenos grupos, o teste pode ser considerado pouco rigoroso, o que significa que a rejeição da hipótese pode ser falsa.

São Paulo

1970

Na Tabela A1 são mostradas as transformações feitas para a análise comparativa entre os municípios canavieiros e não canavieiros. Os testes de normalidade de *Shapiro-Wilk* e de *skewness* e *kurtosis* também são apresentados.

Tabela A1 – Resultado das transformações dos indicadores de bem-estar 1970 no estado de São Paulo

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Raiz quadrada	0,991	0	0,993	0,019	-1,234	-1,612
Pessoas pobres	Inversão e raiz	0,967	0	0,993	0,028	-0,530	-2,559
Mortalidade infantil	Logaritmo	0,965	0	0,988	0	1,493	-2,695
Analfabetos	Raiz quadrada	0,983	0	0,997	0,644	1,141	-0,060
IDH	Raiz quadrada	0,989	0	0,996	0,321	1,677	-0,013
Esperança de vida	Inversão e raiz	0,981	0	0,991	0	-0,525	-1,821
Iluminação elétrica	$\sqrt{\log \frac{1+x}{1-x}}$	0,964	0	0,98	0	2,684	-1,066
Instalações sanitárias	Raiz quadrada	0,858	0	0,861	0	0,692	-6,938

Os indicadores “Analfabetos” e “IDH” são os únicos que atingiram a normalidade depois das transformações aplicadas, de acordo com o teste W de *Shapiro-Wilk*.

Na Tabela A2 é mostrado que, no caso de São Paulo, e para 1970, apenas a variável “L de Theil” transformada (raiz de L de Theil) não apresenta homogeneidade de variância.

Tabela A2 – Resultado do teste de Levene para 1970, no estado de São Paulo

Variável	F	p
L de Theil	0,596	0,440
Pessoas pobres	26,850	0
Mortalidade infantil	7,981	0,004
Analfabetos	6,091	0,013
IDH	7,718	0,005
Esperança de vida	4,692	0,030
Iluminação elétrica	11,358	0
Instalações sanitárias	12,093	0

1980

Para o ano de 1980 a distribuição dos indicadores tem comportamento distinto. Por exemplo, a variável “Domicílios com iluminação elétrica” tem uma distribuição concentrada em valores entre 0,8 e 1, o que interfere na escolha das transformações para a adaptação à normalidade.

As transformações utilizadas para os indicadores de 1980 são apresentadas na Tabela A3. As transformações não trouxeram as variáveis para a distribuição normal, exceto no caso da variável “Analfabetos”. Mesmo segundo os valores de *skewness* e *kurtosis* a distribuição das variáveis está fora da normalidade (exceto “Analfabetos” e “Esperança de vida”), o que, teoricamente, prejudica a análise da aplicação do teste *t*.

Tabela A3 – Resultado das transformações dos indicadores de 1980, no estado de São Paulo

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Logaritmo	0,929	0	0,990	0,001	0,004	2,929
Pessoas pobres	Raiz quadrada	0,975	0	0,987	0	-1,184	-3,422
Mortalidade infantil	Raiz quadrada	0,991	0	0,991	0,002	-4,414	-0,075
Analfabetos	Raiz quadrada	0,981	0	0,996	0,364	1,231	1,128
IDH	Inversão e logaritmo	0,90	0	0,919	0	9,049	0,856
Esperança de vida	Raiz quadrada	0,991	0,002	0,992	0,004	0,995	-2,041
Iluminação elétrica	Inversão e logaritmo	0,857	0	0,894	0	11,006	4,344
Instalações sanitárias	-	0,919	0	-	-	0,729	-5,893

Os resultados dos testes de Levene para homogeneidade de variância também indicam que poucas variáveis são homogêneas. De acordo com o que é apresentado na Tabela A4, metade delas não é homogênea, sendo elas “L de theil”, “Mortalidade infantil”, “Analfabetos” e “Esperança de vida”.

Tabela A4 – Resultados do teste *t* para 1980 no estado de São Paulo

Variável	F	<i>p</i>
L de Theil	0,057	0,811
Pessoas pobres	11,301	0
Mortalidade infantil	1,010	0,315
Analfabetos	1,614	0,204
IDH	64,527	0
Esperança de vida	1,405	0,236
Iluminação elétrica	63,957	0
Instalações sanitárias	12,539	0

1991

Para o ano de 1991 foram feitas as transformações indicadas na Tabela A5. As transformações foram bem sucedidas apenas para as variáveis “Analfabetos” e “Esperança de vida”; as demais variáveis tiveram melhora nos índices de normalidade, mas não é possível dizer que tenham distribuição normal.

Tabela A5 – Transformações dos indicadores de 1991 no estado de São Paulo

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Raiz quadrada	0,972	0	0,991	0,017	1,908	4,059
Pessoas pobres	Raiz quadrada	0,918	0	0,979	0	4,087	-0,484
Mortalidade infantil	Raiz quadrada	0,946	0	0,959	0	6,937	4,934
Analfabetos	Raiz quadrada	0,984	0	0,997	0,735	0,505	0,196
IDH	Inversão e logaritmo	0,866	0	0,885	0	-12,446	10,502
Esperança de vida	Inversão e raiz quadrada	0,964	0	0,995	0,277	1,091	-0,298
Iluminação elétrica	Inversão e raiz quadrada	0,480	0	0,503	0	35,381	100,361
Instalações sanitárias	Inversão e raiz quadrada	0,922	0	0,939	0	6,065	-0,896

Os resultados do teste de homogeneidade da variância (teste de Levene), apresentados na Tabela A6, mostram que apenas as variáveis “L de theil” e “Esperança de vida” são homogêneas.

Tabela A6 – Resultados do teste de Levene para 1991 no estado de São Paulo

Variável	F	<i>p</i>
L de Theil	3,089	0,079
Pessoas pobres	16,781	0
Mortalidade infantil	9,241	0,002
Analfabetos	6,465	0,011
IDH	21,888	0
Esperança de vida	0,749	0,387
Iluminação elétrica	30,787	0
Instalações sanitárias	35,963	0

2000

As transformações feitas para o ano de 2000, e os resultados alcançados, são apresentados na Tabela A7. No caso, apenas a variável “Analfabetos” atingiu a condição de distribuição normal. Algumas variáveis alcançam um valor de *kurtosis* significativa (“Mortalidade infantil”, “Analfabetos” e “Instalações sanitárias”), o que permite as conclusões do teste *t* aplicado.

Tabela A7 – Transformações dos indicadores de 2000 no estado de São Paulo

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Logaritmo	0,922	0	0,983	0	5,488	4,416
Pessoas pobres	Raiz quadrada	0,899	0	0,973	0	6,930	5,541
Mortalidade infantil	Logaritmo	0,915	0	0,992	0,002	0,476	2,472
Analfabetos	Raiz quadrada	0,990	0	0,997	0,366	-0,759	-1,651
IDH		0,981	0	-	-	-3,068	8,669
Esperança de vida	Inversão e raiz quadrada	0,975	0	0,991	0,001	-0,719	2,784
Iluminação elétrica	-	0,435	0	-	-	-63,593	265,915
Instalações sanitárias	Inversão e logaritmo	0,888	0	0,930	0	9,029	0,946

Em relação à homogeneidade de variância, apenas as variáveis “L de theil”, “Mortalidade infantil” e “Esperança de vida” podem ser consideradas homogêneas, como mostram os resultados do teste de Levene apresentados na Tabela A8.

Tabela A8 – Resultados do teste de Levene para 2000 no estado de São Paulo

Variável	F	<i>p</i>
L de Theil	0,337	0,561
Pessoas pobres	20,036	0
Mortalidade infantil	1,919	0,166
Analfabetos	17,276	0
IDH	12,312	0
Esperança de vida	1,368	0,242
Iluminação elétrica	40,873	0
Instalações sanitárias	77,502	0

Alagoas

1970

As transformações feitas para o estado de Alagoas tiveram melhores resultados em comparação àquelas realizadas para os indicadores do estado de São Paulo. Das oito variáveis utilizadas na comparação entre municípios canavieiros e não canavieiros, cinco atingiram a normalidade, ou já eram normais antes de transformadas, como apresentado na Tabela A9. Se considerados os

valores de Z de *skewness* e *kurtosis*, a variável “Pessoas pobres” também atingiu normalidade depois da transformação.

Tabela A9 – Transformações dos indicadores de 1970 no estado de Alagoas

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Logaritmo	0,923	0	0,986	0,430	1,444	-0,057
Pessoas pobres	Inversão e logaritmo	0,847	0	0,972	0,042	-0,969	-1,357
Mortalidade infantil	-	0,978	0,123	-	-	0,337	0,234
Analfabetos	Inversão e raiz quadrada	0,938	0	0,987	0,510	0,664	1,879
IDH	Logaritmo	0,883	0	0,962	0,009	3,131	4,079
Esperança de vida	-	0,977	0,103	-	-	0,498	0,305
Iluminação elétrica	Logaritmo	0,756	0	0,988	0,614	-0,235	-0,269
Instalações sanitárias	Raiz quadrada	0,157	0	0,170	0	23,582	70,685

Para a variável “Instalações sanitárias” não é possível a análise, já que apenas três municípios, entre os canavieiros, possuíam instalações sanitárias no ano de 1970, sendo eles Maceió (não inclusa na análise), Penedo e Delmiro Gouveia.

Da aplicação do teste de Levene para verificação da homogeneidade de variância, três variáveis resultaram não homogêneas – “Pessoas Pobres”, “Analfabetos” e “Esperança de vida”, como mostrado na Tabela A10.

Tabela A10 – Resultado dos testes de Levene para 1970 no estado de Alagoas

Variável	F	<i>p</i>
L de Theil	3,479	0,065
Pessoas pobres	9,575	0,002
Mortalidade infantil	0,073	0,786
Analfabetos	15,903	0
IDH	1,575	0,212
Esperança de vida	14,562	0
Iluminação elétrica	0,947	0,332
Instalações sanitárias	0,002	0,956

1980

Também no ano de 1980, as transformações apresentaram melhores resultados do que para o estado de São Paulo no mesmo ano. Os resultados para as transformações são apresentados na Tabela A11. A conclusão é que apenas duas variáveis não têm distribuição normal após a transformação (“L de Theil” e “IDH”), e que a transformação foi bem sucedida em alguns casos nos demais casos.

Tabela A11 – Transformações dos indicadores de 1980 no estado de Alagoas

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Raiz quadrada	0,849	0	0,909	0	4,782	2,983
Pessoas pobres	Inversão e raiz quadrada	0,955	0,002	0,989	0,697	-0,301	-0,404
Mortalidade infantil	-	0,985	0,412	-	-	1,100	-0,043
Analfabetos	Inversão e raiz quadrada	0,958	0,004	0,990	0,751	-0,344	1,815
IDH	Inversão e raiz quadrada	0,870	0	0,958	0,004	-0,301	-0,404
Esperança de vida	-	0,990	0,727	-	-	-0,246	-0,063
Iluminação elétrica	Logaritmo	0,921	0	0,984	0,313	-0,471	-1,464
Instalações sanitárias	-	0,212	0	-	-	26,267	97,572

Os resultados do teste de homogeneidade de variância, no entanto, não foram tão bem sucedidos. Apenas as variáveis transformadas “Analfabetos”, “Instalações sanitárias” e “IDH” podem ser consideradas homogêneas pelo teste de Levene, cujos resultados são apresentados na Tabela A12.

Tabela A12 – Resultado do teste de Levene para 1980 no estado de Alagoas

Variável	F	<i>p</i>
L de Theil	4,313	0,040
Pessoas pobres	4,123	0,045
Mortalidade infantil	7,224	0,008
Analfabetos	2,189	0,142
IDH	1,213	0,273
Esperança de vida	6,317	0,013
Iluminação elétrica	5,439	0,021
Instalações sanitárias	0,429	0,514

1991

Na Tabela A13 são apresentadas as transformações das variáveis aplicadas para os dados de 1991. De acordo com os índices de *skewness* e *kurtosis*, apenas as variáveis “Instalações sanitárias” e “Pessoas Pobres” não atingiram a normalidade. É importante esclarecer que não foi aplicada transformação à variável “Instalações sanitárias”, pela grande distorção causada: como se trata de um grande conjunto de baixos valores, a normalidade torna-se impossível de ser atingida com as transformações indicadas por Tabachnick e Fidel (2007) e Stevens (2009).

Tabela A13 – Transformações dos indicadores de 1991 no estado de Alagoas

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Raiz quadrada	0,958	0,003	0,977	0,098	1,581	-0,055
Pessoas pobres	Inversão e raiz quadrada	0,889	0	0,959	0,004	-3,296	5,283
Mortalidade infantil	-	0,978	0,110	-	-	-0,149	-1,460
Analfabetos	Inversão e raiz quadrada	0,960	0,005	0,991	0,822	-0,253	1,285
IDH	Raiz quadrada	0,961	0,005	0,975	0,064	2,517	2,661
Esperança de vida	Raiz quadrada	0,973	0,045	0,975	0,064	0,979	-1,223
Iluminação elétrica	Inversão e raiz quadrada	0,969	0,023	0,971	0,030	0,144	-2,078
Instalações sanitárias	-	0,270	0	-	-	18,804	44,954

A Tabela A14 apresenta os resultados do teste de Levene para o estado de Alagoas, em 1991. No caso, conclui-se que apenas a variável “Instalações sanitárias” não é homogênea.

Tabela A14– Resultado do teste de Levene para 1991 no estado de Alagoas

Variável	F	<i>p</i>
L de Theil	0,210	0,647
Pessoas pobres	1,097	0,297
Mortalidade infantil	0,002	0,956
Analfabetos	0,257	0,613
IDH	1,293	0,258
Esperança de vida	0,485	0,487
Iluminação elétrica	0,401	0,527
Instalações sanitárias	34,796	0

E para o ano de 2000, as transformações feitas levaram cinco das oito variáveis apresentadas na Tabela A15 à normalidade quando analisados os resultados do teste de Shapiro–Wilk, exceto as três a seguir mencionadas: “L de Theil”, “Instalações sanitárias” e “Iluminação elétrica”. As duas últimas são variáveis com alta concentração de valores nos extremos da distribuição, sendo que o indicador de domicílios com “Instalações elétricas” tem concentração em altas porcentagens, e o indicador de domicílios com “Instalações sanitárias” tem concentração em baixas porcentagens. Por sua vez, se analisados os valores dos índices de *skewness* e *kurtosis*, “Pessoas pobres” e “Iluminação elétrica” não são variáveis normalmente distribuídas.

Tabela A15 – Transformações dos indicadores de 2000 no estado de Alagoas

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>P</i>	Valor Z skewness final	Valor Z Kurtosis final
L de Theil	Logaritmo	0,939	0	0,974	0,044	1,986	-0,490
Pessoas pobres	Inversão e raiz quadrada	0,902	0	0,975	0,055	1,101	2,784
Mortalidade infantil	Logaritmo	0,940	0	0,981	0,169	1,143	-0,538
Analfabetos	Inversão e raiz quadrada	0,941	0	0,991	0,762	0,800	-0,221
IDH	-	0,979	0,124	-	-	2,186	2,197
Esperança de vida	Inversão e raiz quadrada	0,969	0,018	0,987	0,476	-0,390	-0,304
Iluminação elétrica	Inversão e raiz quadrada	0,918	0	0,930	0	4,070	1,388
Instalações sanitárias	Logaritmo	0,620	0	0,958	0,003	0,634	-2,068

De acordo com os resultados do teste de Levene, apresentados na Tabela A16, constata-se que a hipótese de homogeneidade da variância de todas as variáveis não deve ser rejeitada.

Tabela A16 – Resultado do teste de Levene para 2000 no estado de Alagoas

Variável	F	<i>p</i>
L de Theil	0,106	0,745
Pessoas pobres	0,038	0,844
Mortalidade infantil	0,092	0,762
Analfabetos	0,445	0,506
IDH	0,194	0,659
Esperança de vida	0,086	0,768
Iluminação elétrica	2,208	0,140
Instalações sanitárias	0,017	0,893

Apêndice B – Procedimento para escolha do número de clusters

Neste apêndice são apresentados os procedimentos adotados para a escolha do número de clusters, em cada estado e em cada ano. O que é apresentado, como exemplo, é o procedimento aplicado à formação de cluster de municípios canavieiros em São Paulo, em 1991.

Na Tabela B1 são apresentados os resultados obtidos para quando foi imposta a formação de três clusters.

Tabela B1 – Resultados da Análise Cluster em três grupos, para São Paulo, em 1991

	Grupo 1	Grupo 2	Grupo 3	p
Membros	74	55	19	-
L de Theil	0,414	0,384	0,479	0,000
Pessoas pobres	12,450	15,910	30,062	0,000
Mortalidade infantil	20,399	27,939	28,738	0,000
Analfabetos	12,255	15,205	17,358	0,000
IDH	0,782	0,7412	0,7015	0,000
Esperança de vida	71,087	68,164	67,942	0,000
Iluminação elétrica	0,993	0,991	0,939	0,000
Instalações sanitárias	0,840	0,779	0,479	0,000

Já na Tabela B2 são apresentados os resultados obtidos para quando foi imposta a formação de quatro clusters.

Tabela B2 – Resultados da Análise Cluster com quatro grupos, para São Paulo, em 1991

	Grupo 1	Grupo 2	Grupo 3	Grupo 4	p
Membros	68	48	4	28	0
L de Theil	0,411	0,381	0,476	0,454	0
Pessoas pobres	11,477	15,144	40,871	24,881	0
Mortalidade infantil	20,510	28,630	38,024	23,970	0
Analfabetos	12,044	15,089	18,350	16,296	0
IDH	0,784	0,743	0,637	0,727	0
Esperança de vida	71,041	67,904	65,418	69,587	0
Iluminação elétrica	0,994	0,992	0,828	0,975	0
Instalações sanitárias	0,852	0,796	0,324	0,593	0

A divisão dos municípios canavieiros em quatro clusters resulta na formação de um cluster que tem melhores resultados em sete dos oito indicadores (o Grupo 1), com exceção de “L de Theil”,

e um grupo com piores resultados para todos os indicadores (o Grupo 3). Esta classificação dos municípios foi descartada pelo baixo número de municípios no Grupo 3, com apenas quatro na amostra.

Os resultados da Análise Discriminante em três grupos são apresentados na Tabela B3. No caso da formação de três grupos, existem duas funções discriminantes para as quais, a princípio, não é possível identificar entre quais grupos elas discriminam; para tanto é necessário analisar a dispersão de suas raízes.

Tabela B3 - Resultados da Análise Discriminante para a classificação em três grupos, para São Paulo, em 1991

Variáveis	Wilks-lambda	Função de class. 1	Função de class. 2	Função de class. 3	Função discrim. 1	Função discrim. 2
Participação da agricultura na economia	0,905	39,166	37,988	40,871	0,684	0,211
Transferência estadual	0,937	-0,033	-0,140	-0,060	1,139	-1,232
Cana/km ²	0,916	0,657	0,740	0,518	-0,591	-0,237
Cota-parte municipal	0,907	4,197	4,435	4,184	-0,831	0,554

O resultado da Análise Discriminante aplicada para a divisão de municípios canavieiros em três grupos é ruim, pois há 0% de acerto na alocação dos municípios no grupo dos que têm piores resultados (Grupo 3) e 72,9% de acerto na alocação no grupo de melhores resultados (Grupo 1), de acordo com os resultados obtidos do software *STATISTICA*.

A decisão por dividir os municípios em apenas dois grupos, que foi a decisão tomada, foi também baseada na distância euclidiana entre o centroide dos grupos. A distância euclidiana é a medida de dissimilaridade, que é um resultado dado pelo software. Apesar de ter distâncias maiores entre os componentes dos grupos, como mostrado na Tabela B4, deve-se levar em conta que o Grupo 3 (de municípios com piores resultados) tem apenas quatro indivíduos na amostra, e que não se identifica facilmente, entre os Grupos 2 e 4, qual é o que tem municípios com piores indicadores.

Tabela B4 – Distâncias Euclidianas para os clusters – divisão em 2, 3 e 4 grupos

Número de clusters	Distância Euclidiana					
	1-2	1-3	2-3	2-4	3-4	1-4
2	1,17	-	-	-	-	-
3	0,87	1,72	1,23	-	-	-
4	0,91	3,17	2,65	0,86	2,23	1,18

Para visualizar a distribuição espacial dos grupos, um mapa do estado de São Paulo foi feito no software ArcGIS. E, com intuito de comparação, o mapa quando da consideração de três clusters é apresentado na Figura B1. Na comparação entre os mapas, se nota que todos os municípios que estão no grupo de pior desempenho quando da divisão em três clusters estão no grupo de pior desempenho quando da divisão em dois clusters. Nesse sentido, não há vantagem em terem-se três clusters.

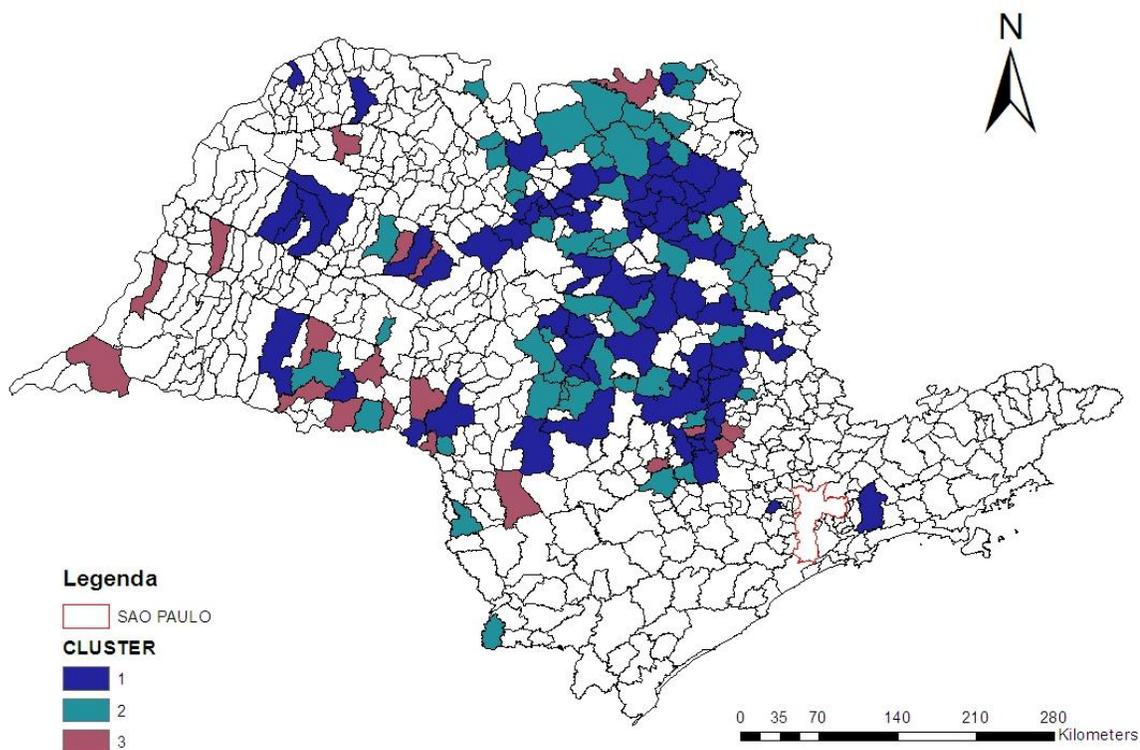


Figura B1- Separação dos municípios em três clusters considerados indicadores de qualidade de vida no estado de São Paulo em 1991. Grupo 1, melhores resultados; Grupo 3, piores resultados.

Apêndice C – Transformação das variáveis *predictors*

Como a aplicação do método de Análise Discriminante requer normalidade das variáveis, homogeneidade de variância e linearidade, as variáveis utilizadas como *predictors* foram transformadas para que atingissem ou se aproximassem das condições requeridas. As transformações feitas são apresentadas a seguir, de acordo com o estado e o ano.

São Paulo

As variáveis foram transformadas seguindo os padrões de transformação apresentados por Tabachnick e Fidell (2007) e Stevens (2009). Evitou-se o uso do logaritmo para não haver perda de informação no caso dos municípios que têm valores de alguma variável igual à zero, em um dado ano. De acordo com Hair et al. (2010), o valor crítico a um nível de significância de 1% é 2,58 para os testes para *skewness* e *kurtosis*.

1970

A Tabela C1 apresenta os resultados das transformações para as variáveis utilizadas como *predictors* na Análise Discriminante para o ano de 1970. As variáveis “Participação da agropecuária” e “Participação dos serviços” atingiram normalidade, além da variável “Participação da indústria”, que não recebeu transformação. A variável “Cana/km²” teve resultado estatisticamente significativo em relação a kurtosis a 1%.

Tabela C1 – Resultado das transformações feitas para realização da Análise Discriminante – São Paulo 1970

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Participação da agropecuária	Raiz quadrada	0,966	0,01878	0,988	0,582	-0,898	-0,676
Participação cana na economia	Raiz quadrada	0,491	0	0,860	0	8,230	18,116
Cana/km ²	Raiz quadrada	0,795	0	0,920	0	3,359	-0,089
Participação dos serviços	Raiz quadrada	0,951	0,001	0,983	0,287	1,330	-0,416
Cana/hab	Raiz quadrada	0,746		0,934	0	4,147	3,0115
Participação da Indústria	-	0,977	0,115	-	-	1,110	-1,184

Já a Tabela C2 apresenta o resultado do teste de Levene, que indica que apenas a variável “Participação da cana na economia” não apresenta homogeneidade de variância. Lembra-se que quando o teste de Levene é estatisticamente significativo, a hipótese de que existe homogeneidade de variância deve ser ignorada.

Tabela C2 – Resultado do teste de Levene para aplicação da Análise Discriminante – 1970, no estado de São Paulo

Variável	F	p
Participação da agropecuária	2,313	0,131
Participação cana na economia	4,099	0,045
Cana/km²	0,037	0,846
Participação dos serviços	2,395	0,125
Cana/hab	1,189	0,278
Participação da indústria	0,459	0,499

Para testar linearidade, Tabachnick e Fidell (2007) recomendam a análise das dispersões entre as variáveis objeto de análise, em procedimento gráfico como o apresentado na Figura C1. Variáveis lineares e normais apresentam dispersão oval. A sequência das variáveis na figura mostra é a mesma apresentada na Tabela C1. Na Figura C1 são apresentadas as dispersões das variáveis consideradas na Análise Discriminantes, em São Paulo, em 1970. Pode-se observar que, com raras exceções, não há linearidade entre elas, e um padrão de relação é de difícil percepção.

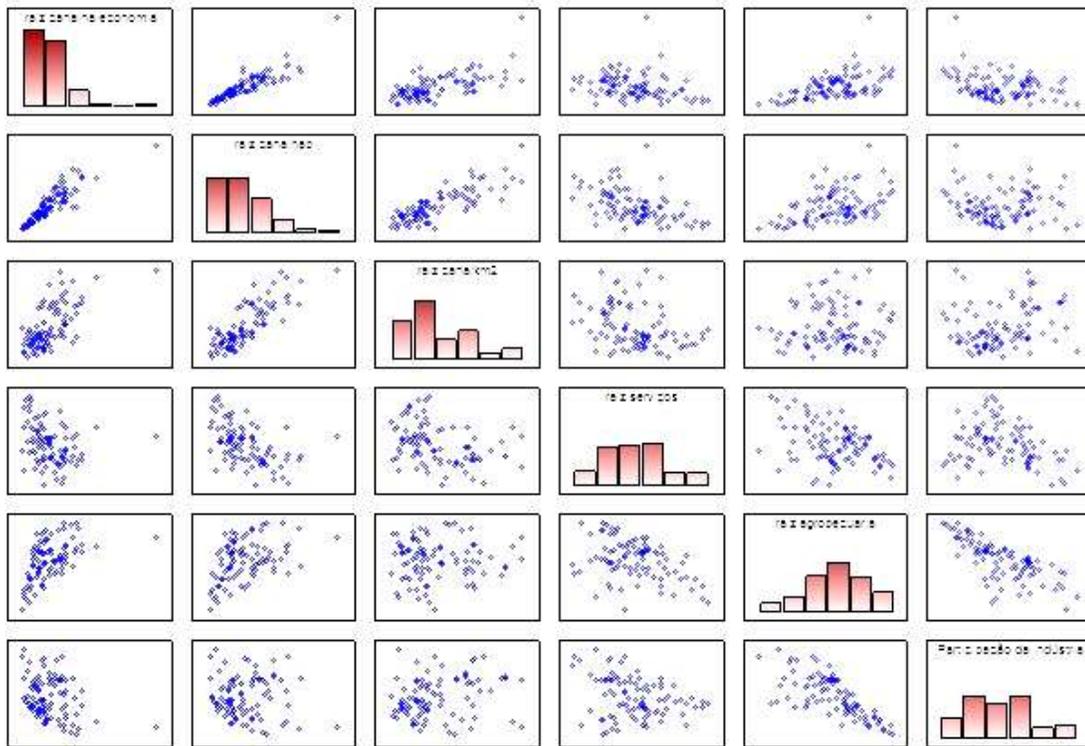


Figura C1 – Dispersões entre as variáveis *predictors* empregadas para 1970, em São Paulo.

1980

A Tabela C3 apresenta os resultados das transformações feitas para realização da Análise Discriminante no estado de São Paulo, no ano de 1980. Os resultados indicam que, segundo o teste de *Shapiro –Wilk*, apenas a variável “Participação dos serviços” atingiu a normalidade, embora as variáveis “Participação da agropecuária” e “Cana/habitante” também tenham valores significativos a 1% para os testes de *skewness* e *kurtosis*.

Tabela C3 – Resultado das transformações para realização da Análise Discriminante – São Paulo, 1980

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Participação da agropecuária	Logaritmo	0,934	0	0,963	0,003	-2,419	-0,509
Participação da cana na agropecuária	Raiz quadrada	0,760	0	0,922	0	5,328	6,616
Cana/hab	Raiz quadrada	0,874	0	0,969	0,010	2,173	-0,689
Participação dos serviços	Logaritmo	0,944	0	0,995	0,962	-0,199	0,7195

Na Figura C2 são apresentadas as dispersões das variáveis consideradas na Análise Discriminantes, em São Paulo, em 1980. As conclusões são as mesmas de 1970.

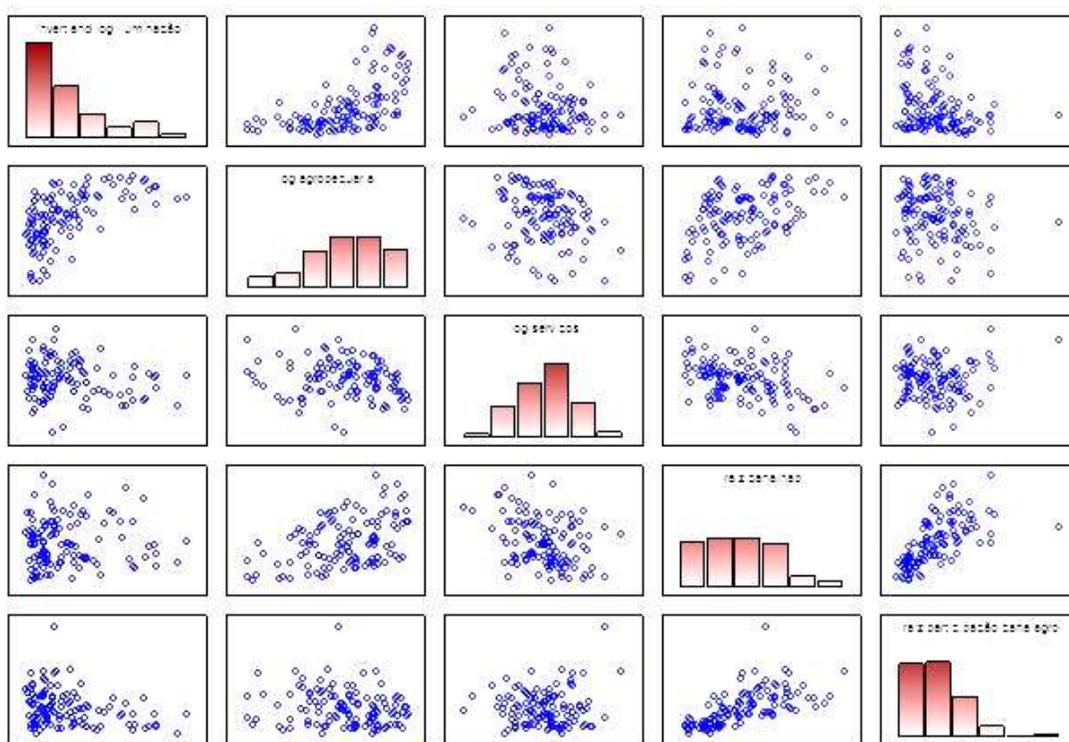


Figura C2– Dispersões entre as variáveis *predictors* empregadas para 1980, em São Paulo.

1991

A Tabela C4 mostra os resultados das transformações feitas para realização da Análise Discriminante com os dados de 1991, no estado de São Paulo, apenas para os municípios considerados canaveiros.

Tabela C4 – Resultado das transformações para realização da Análise Discriminante – São Paulo, 1991

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Renda serviços	Logaritmo	0,830	0	0,978	0,020	2,149	4,418
Cana/hab	Raiz quadrada	0,586	0	0,822	0	6,642	3,385
Cana na economia	Raiz quadrada	0,481	0	0,811	0	8,275	9,482
Transferência estadual	Raiz quadrada	0,303	0	0,599	0	17,750	38,250
Cota-parte municipal	Raiz quadrada	0,716	0	0,866	0	5,044	4,439
Cana na agricultura	Raiz quadrada	0,330	0	0,958	0	3,613	1,069

O teste de Levene relacionado à homogeneidade de variância indica que apenas a variável “Transferência estadual” (transformada) não apresenta a característica de homogeneidade, como mostrado na Tabela C5.

Tabela C5 – Resultado do teste de Levene para as variáveis transformadas utilizadas na Análise Discriminante – 1991, no estado de São Paulo

Variável	F	<i>p</i>
Renda serviços	1,306	0,254
Cana/habitante	2,116	0,147
Cana na economia	0,269	0,604
Transferência estadual	4,149	0,043
Cota-parte municipal	1,868	0,173
Cana na agricultura	1,857	0,175

Na Figura C3 são apresentadas as dispersões das variáveis consideradas na Análise Discriminantes, em São Paulo, em 1991. As dispersões entre as variáveis não indicam linearidade entre elas, e um padrão de relação é de difícil percepção.

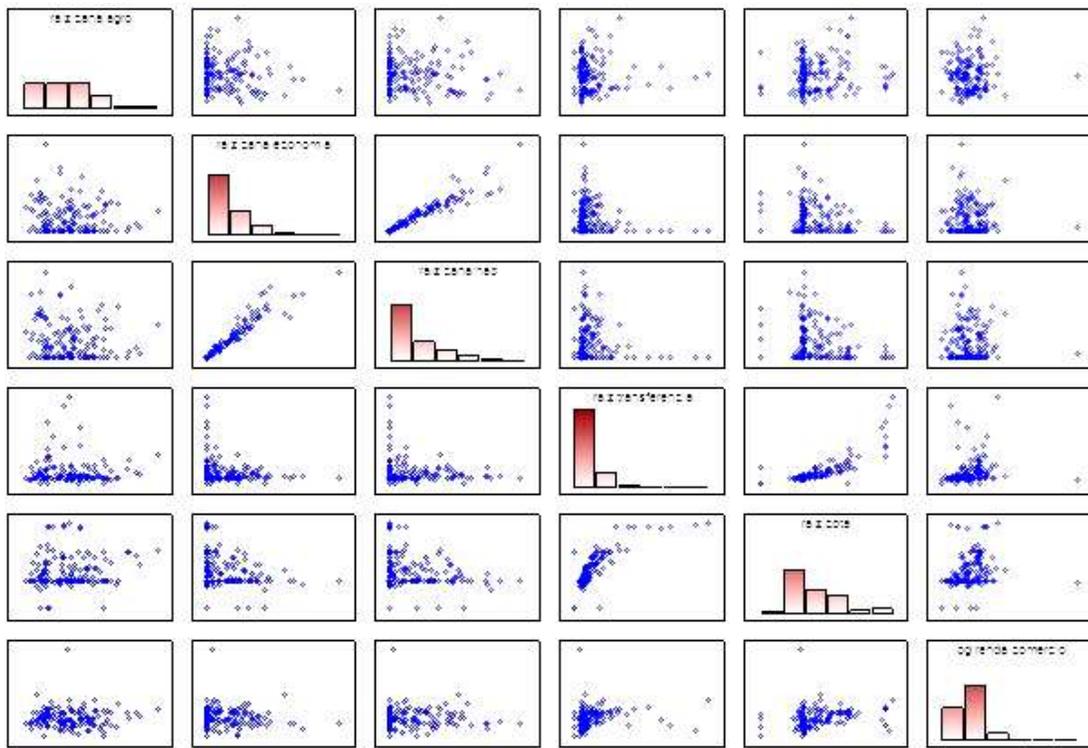


Figura C3 – Dispersões entre as variáveis *predictors* empregadas para 1991, em São Paulo.

2000

Os resultados das transformações das variáveis usadas na Análise Discriminante, em São Paulo, em 2000, são apresentados na Tabela C6.

Tabela C6– Resultado das transformações para aplicação da Análise Discriminante – São Paulo, 2000

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Receita total	Logaritmo	0,367	0	0,954	0	4,578	4,861
Renda na indústria	Raiz quadrada	0,952	0,00001	0,958	0	-1,950	10,828
Cana/km²	Raiz quadrada	0,931	0	0,978	0,006	1,221	-2,190
Cana na economia	Raiz quadrada	0,880	0	0,976	0,003	2,105	-1,095
Renda no comércio	Logaritmo	0,926	0	0,982	0,020	1,375	5,057
Renda agricultura	Logaritmo	0,906	0	0,984	0,040	2,586	1,127
Renda nos serviços	Logaritmo	0,973	0,001	0,992	0,444	0,092	1,117

Os valores *Z* de *kurtosis* e *skewness* indicam que algumas variáveis (“Renda na agricultura”, “Cana na economia”, “Renda nos serviços” e “Cana/km²”) podem ser consideradas normais. Se levados em consideração os Lambdas de *Wilk* resultantes da transformação, apenas a variável “Renda nos serviços” é normal.

Os resultados do teste de Levene para verificar a homogeneidade das variáveis inclusas no modelo discriminante, em 2000, para o estado de São Paulo, são apresentados na Tabela C7.

Tabela C7 – Resultado do teste de Levene para as variáveis aplicadas na Análise Discriminante – 2000, no estado de São Paulo

Variável	F	<i>p</i>
Receita total	13,035	0
Renda na indústria	2,651	0,105
Cana/km²	0,608	0,436
Cana na economia	0,058	0,809
Renda no comércio	0	0,986
Renda agricultura	0,226	0,634
Renda nos serviços	0,903	0,343

Apenas a variável “Receita total” não apresenta homogeneidade de variância. A linearidade também foi testada usando-se o procedimento gráfico de análise de dispersão, como apresentado na Figura C4. Novamente, relações não lineares não podem ser detectadas na análise das dispersões.

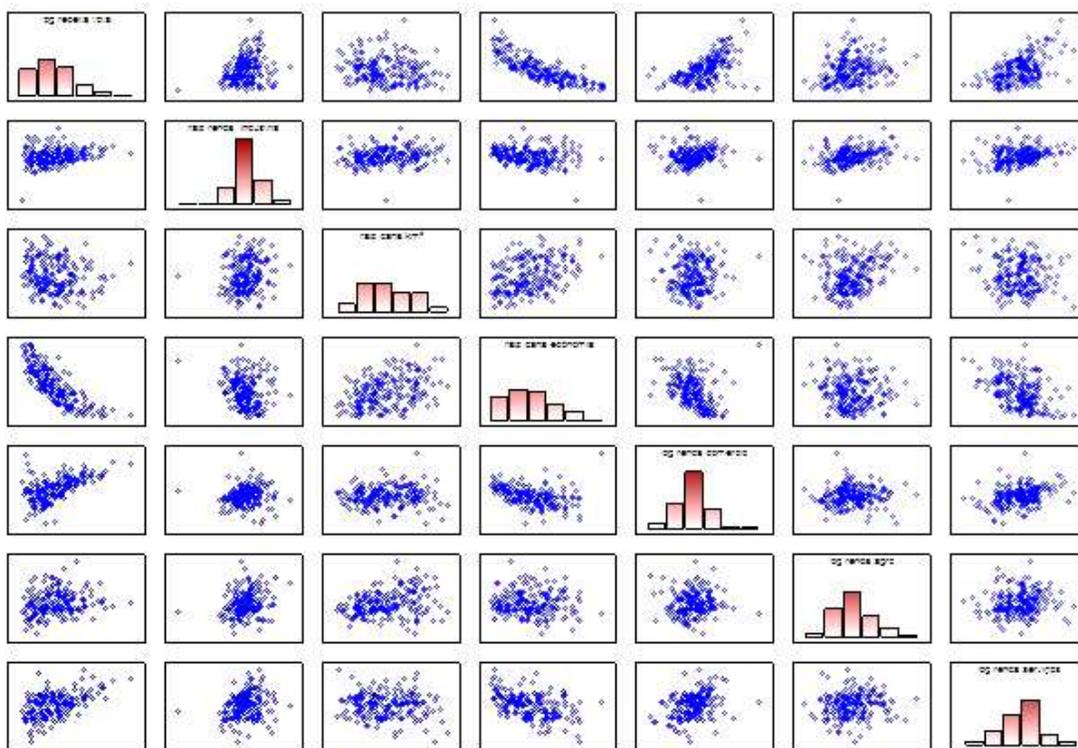


Figura C4 – Dispersões entre as variáveis *predictors* empregadas para 2000, em São Paulo.

Alagoas

1970

Para Alagoas, em 1970, as variáveis usadas na Análise Discriminante são apresentadas na Tabela C8, assim como os resultados das transformações aplicadas. Nenhuma das duas variáveis atingiu a condição de normalidade.

Tabela C8 – Resultado das transformações realizadas para aplicação da Análise Discriminante – Alagoas, 1970

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Participação da indústria	Logaritmo	0,774	0	0,905	0,032	2,835	3,305
Participação da cana na economia	-	0,908	0,037	-	-	2,581	4,147

De acordo com os resultados do teste de Levene apresentados na Tabela C9, nenhuma variável apresenta homogeneidade de variância.

Tabela C9 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante – 1970, no estado de Alagoas

Variável	F	p
Participação da cana na economia	0,197	0,661
Participação da indústria	0,327	0,573

Para testar linearidade, o procedimento gráfico de análise das dispersões entre as variáveis analisadas foi aplicado, mas com poucos indivíduos na amostra o procedimento é praticamente inviável. A dispersão obtida é apresentada na Figura C5.

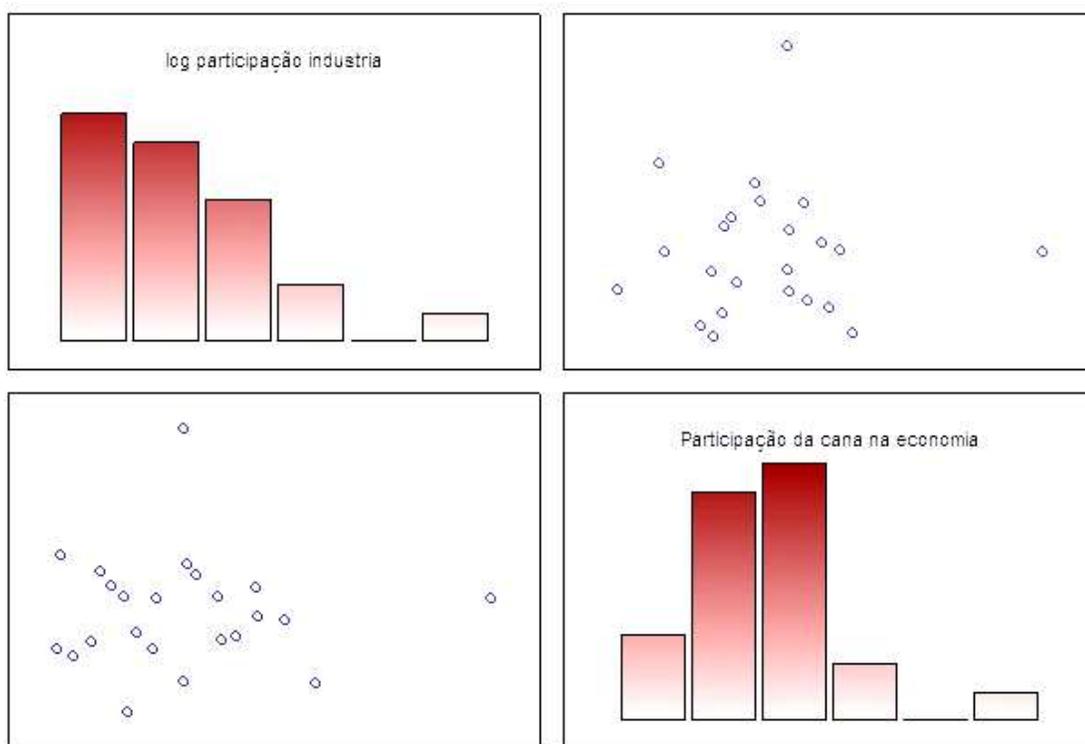


Figura C5 – Dispersões entre as variáveis *predictors* empregadas, em Alagoas, em 1970.

1980

As variáveis usadas na Análise Discriminante para o ano de 1980 no estado de Alagoas, são apresentadas na Tabela C10. Nenhuma variável necessitou de transformação para atingir a condição requerida de normalidade.

Tabela C10 – Resultado da análise das variáveis empregadas na Análise Discriminante – Alagoas, 1980

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Participação da agropecuária	-	0,908	0337	-	-	-1,954	1,441
Cana/km ²	-	0,968	0631	-	-	0,998	-0,099
Participação da cana na agricultura	-	0,946	0,224	-	-	1,530	0,034
Cana/habitante	-	0,968	0,637	-	-	-0,348	-0,357

Por outro lado, de acordo com os resultados do teste de Levene, apresentados na Tabela C11, nenhuma variável apresenta homogeneidade de variância.

Tabela C11 – Resultado do teste de Levene para as variáveis aplicadas na Análise Discriminante – 1980, no estado de Alagoas

Variável	F	<i>p</i>
Participação da agropecuária	0,391	0,537
Cana/km ²	4,225	0,051
Participação da cana na agricultura	2,085	0,162
Cana/habitante	1,259	0,273

A Figura C6 apresenta o resultado do procedimento adotado para análise da dispersão entre as variáveis *predictors*. Como comentado, é difícil visualizar tendência com um número de indivíduos na amostra.

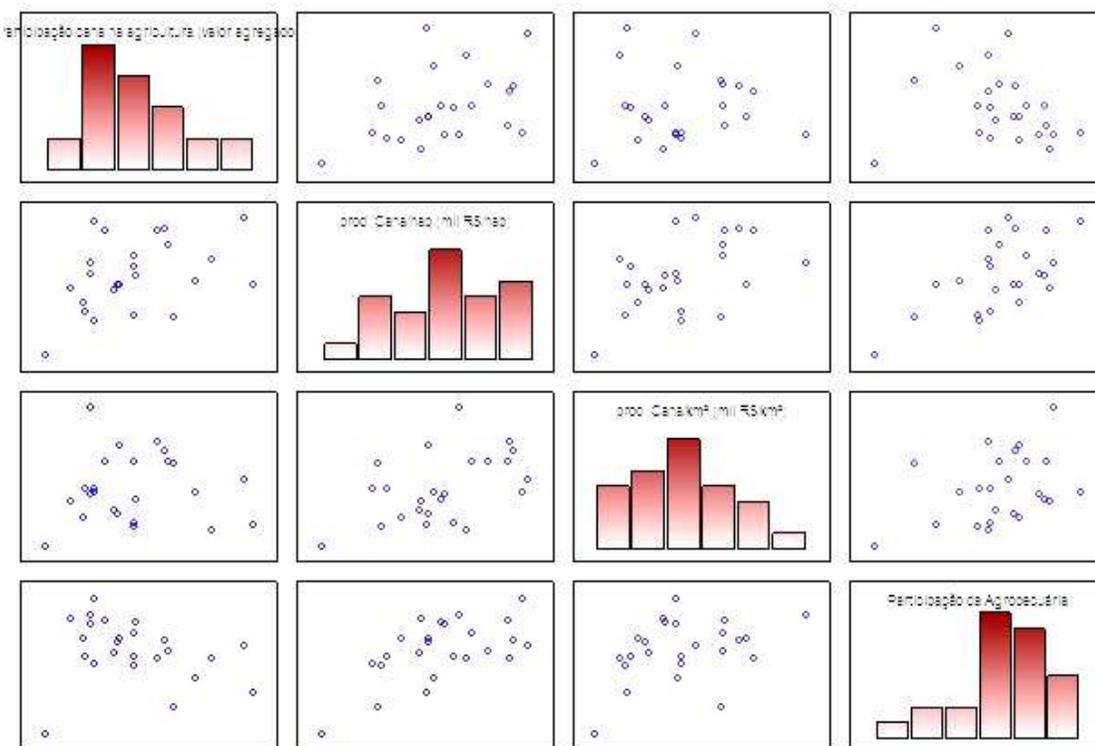


Figura C6– Dispersões entre as variáveis *predictors*, para 1980, em Alagoas.

1991

As variáveis usadas na Análise Discriminante para o ano de 1980, no estado de Alagoas, os procedimentos e os resultados das transformações, quando aplicadas, são apresentados na Tabela C12. As variáveis “Participação da agropecuária”, “Cana/habitante” e “Cana/km²” apresentaram normalidade sem transformação, e as variáveis “Renda no comércio” e “Transferência estadual” atingiram normalidade segundo o testes de *Shapiro-Wilk*. A variável “Renda na indústria” teve resultado estatisticamente significativo a 1% para o teste Z de *kurtosis*.

Tabela C12 – Resultado das transformações aplicadas às variáveis para realização da Análise Discriminante – Alagoas, 1991

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Participação da agropecuária	-	0,985	0,959	-	-	-0,512	0,362
Renda na indústria	Logaritmo	0,612	0	0,876	0,004	3,073	2,373
Cana/km ²	-	0,935	0,083	-	-	1,749	0,298
Cana/habitante	-	0,965	0,468	-	-	1,156	-0,019
Renda no comércio	Logaritmo	0,885	0,007	0,948	0,215	1,449	-0,316
Transferência estadual	Raiz quadrada	0,739	0	0,932	0,071	5,231	6,834
Renda nos serviços	Logaritmo	0,417	0	0,901	0,020	2,909	4,268

E, de acordo com os resultados do teste de Levene apresentados na Tabela C13, nenhuma variável apresenta homogeneidade de variância.

Tabela C13 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante – 1991, no estado de Alagoas

Variável	F	<i>p</i>
Participação da agropecuária	0,099	0,755
Cana/km ²	1,869	0,184
Participação da cana na agricultura	2,140	0,156
Cana/habitante	2,057	0,163

A Figura C7 mostra a dispersão entre as variáveis *predictors* empregadas, no ano 1991. Como antes mencionado, o número de indivíduos na amostra é pequeno e tendências são de difícil visualização.

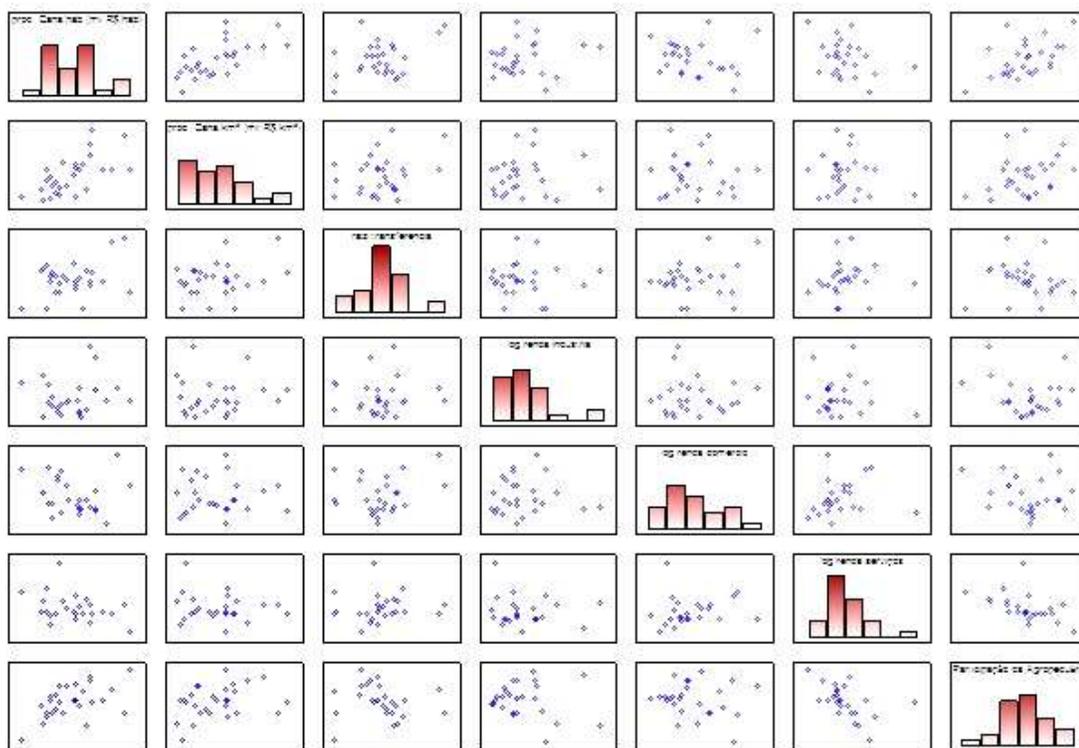


Figura C7 – Dispersões entre as variáveis *predictors* empregadas, para 1991, em Alagoas.

2000

Finalmente, as transformações realizadas para as variáveis empregadas no ano de 2000, para o estado de Alagoas, são apresentadas na Tabela C14. As variáveis “Cana/habitante” e “Cana/km²” atingiram normalidade com a transformação aplicada, já a variável “Participação dos serviços” não necessitou de transformação.

Tabela C14 – Resultado das transformações realizadas para as variáveis empregadas na Análise Discriminante – Alagoas, 2000

Variável	Transformação	W de Shapiro-Wilk inicial	<i>p</i>	W de Shapiro-Wilk final	<i>p</i>	Valor Z skewness final	Valor Z Kurtosis final
Receita total	Logaritmo	0,241	0	0,785	0	5,808	12,041
Renda na indústria	Raiz quadrada	0,778	0	0,897	0,007	0,935	3,025
Cana/km²	Raiz quadrada	0,931	0,053	0,954	0,220	0,568	-1,179
Cana/hab	Raiz quadrada	0,933	0,061	0,973	0,631	-0,484	1,896
Participação dos serviços	-	0,977	0,767	-	-	0,907	1,262
Cota federal municipal	Logaritmo	0,250	0	0,669	0	12,670	35,406

De acordo com os resultados do teste de Levene apresentados na Tabela C15, as variáveis “Receita total”, “Participação dos serviços” e “Cota federal municipal” não apresentam homogeneidade de variância.

Tabela C15 – Resultado do teste de Levene para variáveis empregadas na Análise Discriminante -1991, no estado de Alagoas

Variável	F	<i>p</i>
Receita total	5,386	0,028
Renda na indústria	2,140	0,156
Cana/km²	2,258	0,144
Cana/habitante	1,456	0,238
Participação dos serviços	6,201	0,019
Cota federal municipal	4,602	0,041

Por último, a Figura C8 mostra os resultados da análise gráfica de dispersão entre as variáveis *predictors*, que é inconclusivo.

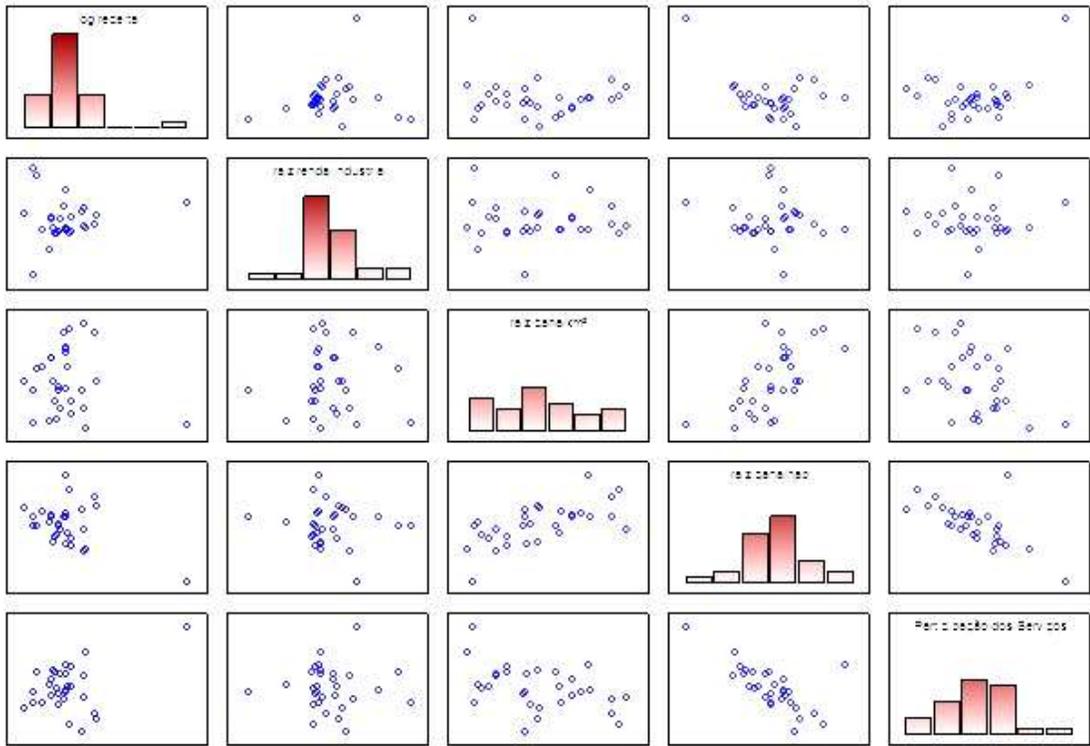


Figura C8 – Dispersões entre as variáveis *predictors* empregadas, em 1991, em Alagoas.

Apêndice D – Novos municípios canavieiros no estado de São Paulo

A localização dos municípios canavieiros no estado de São Paulo é apresentada nos mapas abaixo, na Figura D1, para os anos 1970, 1980, 1991 e 2000.

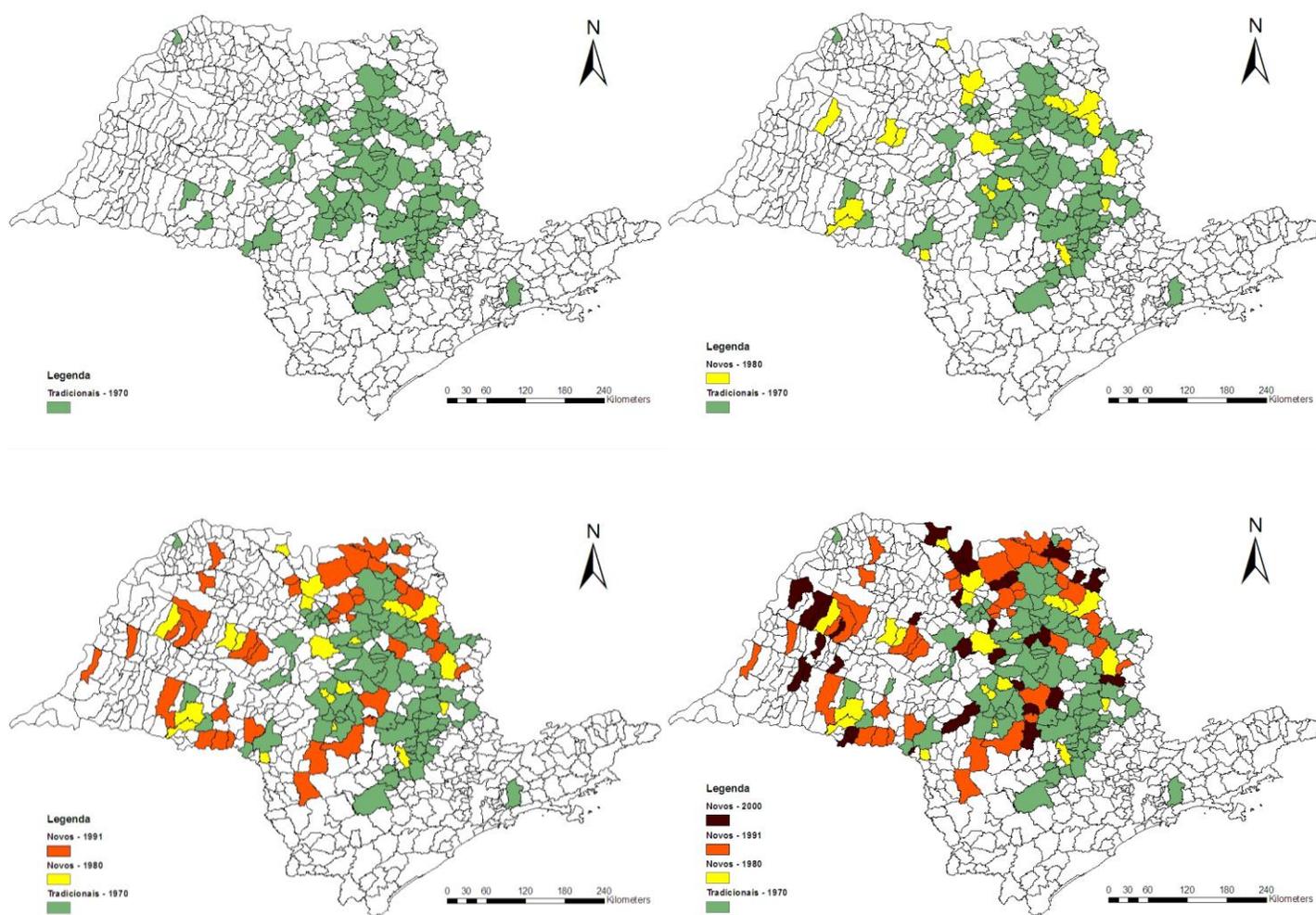


Figura D1 – Expansão da atividade canvieira nos municípios do estado de São Paulo, de 1970 a 2000.

Apêndice E – Teste de hipóteses não paramétrico de Mann-Whitney para novos canavieiros

A Tabela E1 mostra o resultado do teste de hipóteses de Mann-Whitney utilizado na análise da evolução dos indicadores dos novos municípios canavieiros, na comparação com os que nunca se tornaram canavieiros.

Tabela E1 – Teste de hipóteses não paramétrico para análise da evolução dos indicadores – São Paulo de 1970 a 2000

Variável	z	p
Analfabetos 1970	-3,104	0,001
Analfabetos 1980	-2,633	0,008
Analfabetos 1991	-2,449	0,014
Analfabetos 2000	-2,858	0,004
Esperança de vida 1970	-3,747	0
Esperança de vida 1980	-1,708	0,087
Esperança de vida 1991	-1,699	0,089
Esperança de vida 2000	-1,678	0,093
Mortalidade infantil 1970	-3,716	0
Mortalidade infantil 1980	-1,715	0,086
Mortalidade infantil 1991	-1,699	0,089
Mortalidade infantil 2000	-1,678	0,093
IDH 1970	-3,778	0
IDH 1980	-3,654	0
IDH 1991	-3,362	0
IDH 2000	-2,966	0,003
Pessoas pobres 1970	-2,915	0,003
Pessoas pobres 1980	-4,447	0
Pessoas pobres 1991	-5,190	0
Pessoas pobres 2000	-4,298	0
L de Theil 1970	-0,092	0,926
L de Theil 1980	-0,472	0,636
L de Theil 1991	-1,019	0,308
L de Theil 2000	-1,409	0,158
Iluminação elétrica 1970	-4,566	0
Iluminação elétrica 1980	-4,658	0
Iluminação elétrica 1991	-4,071	0
Iluminação elétrica 2000	-3,121	0,001
Instalações sanitárias 1970	-4,635	0
Instalações sanitárias 1980	-5,342	0
Instalações sanitárias 1991	-5,981	0
Instalações sanitárias 2000	-6,341	0

Apêndice F – Evolução dos indicadores IDH e L de Theil

São Paulo

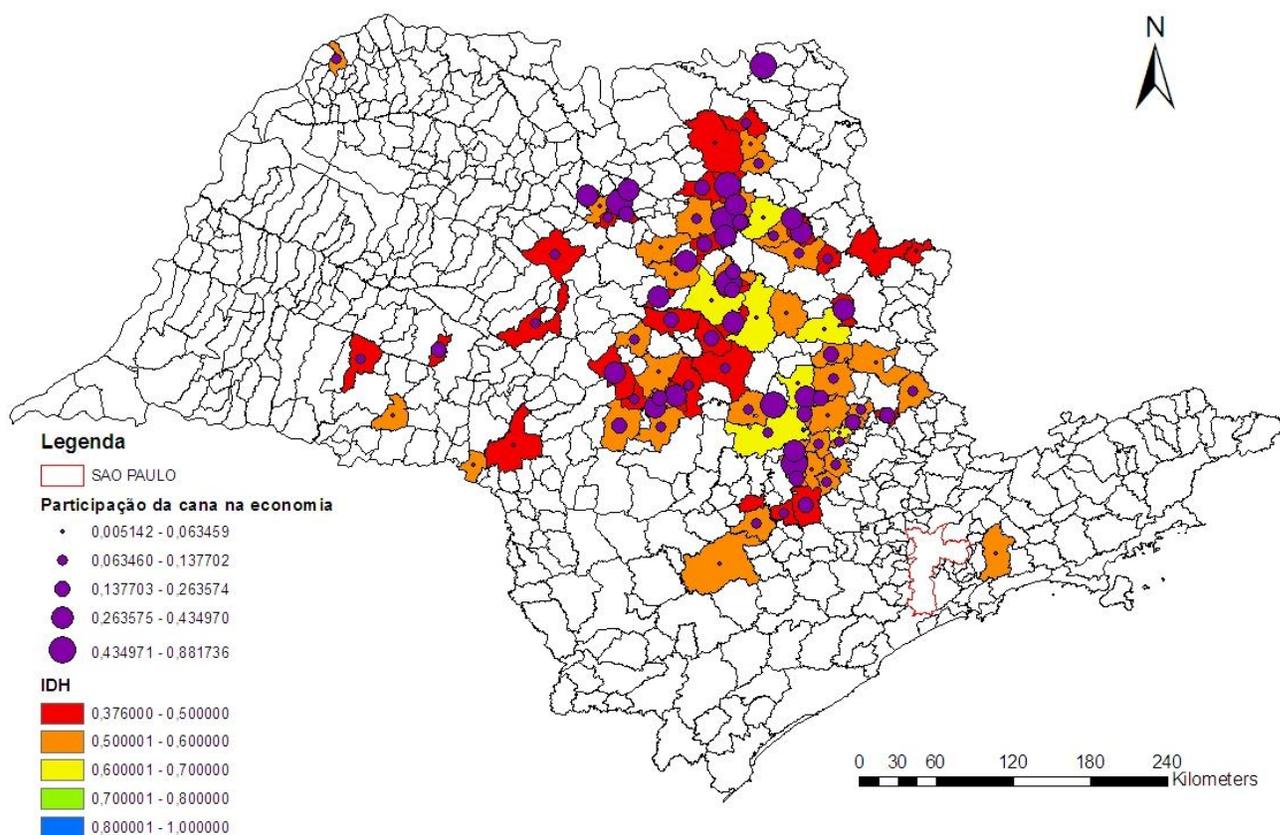


Figura F1 – IDH e participação da cana na economia – São Paulo – 1970.

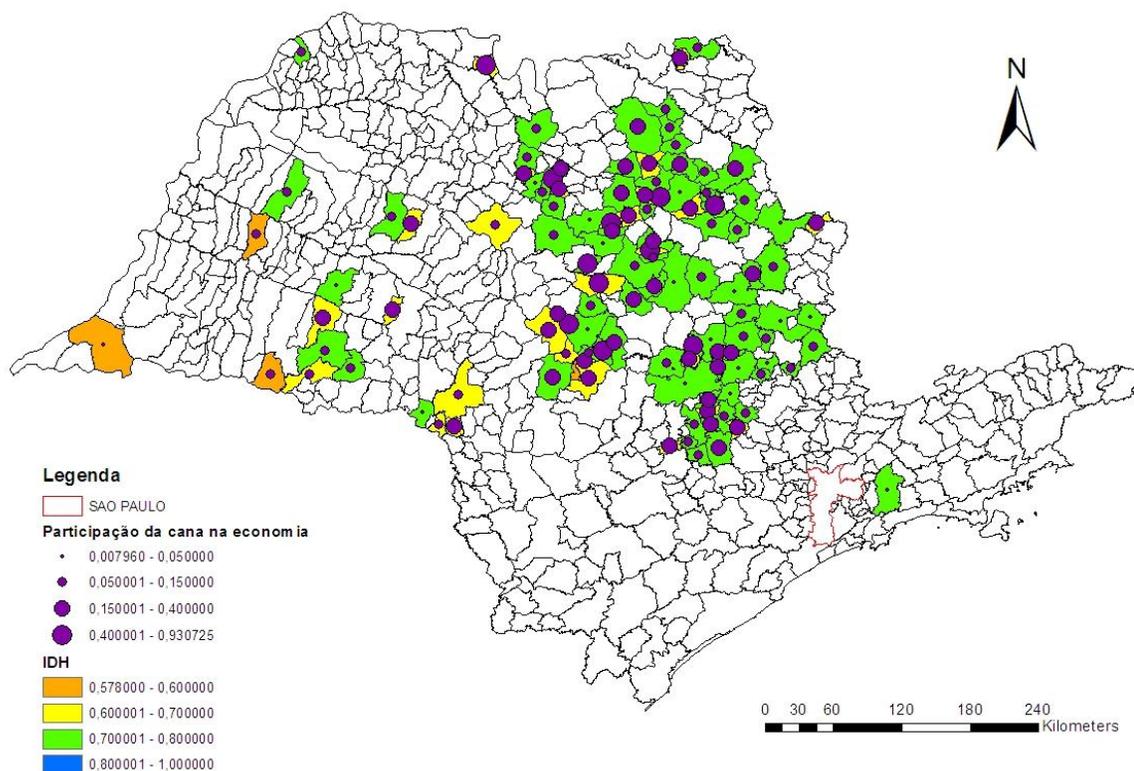


Figura F2 - IDH e participação da cana na economia – São Paulo – 1980.

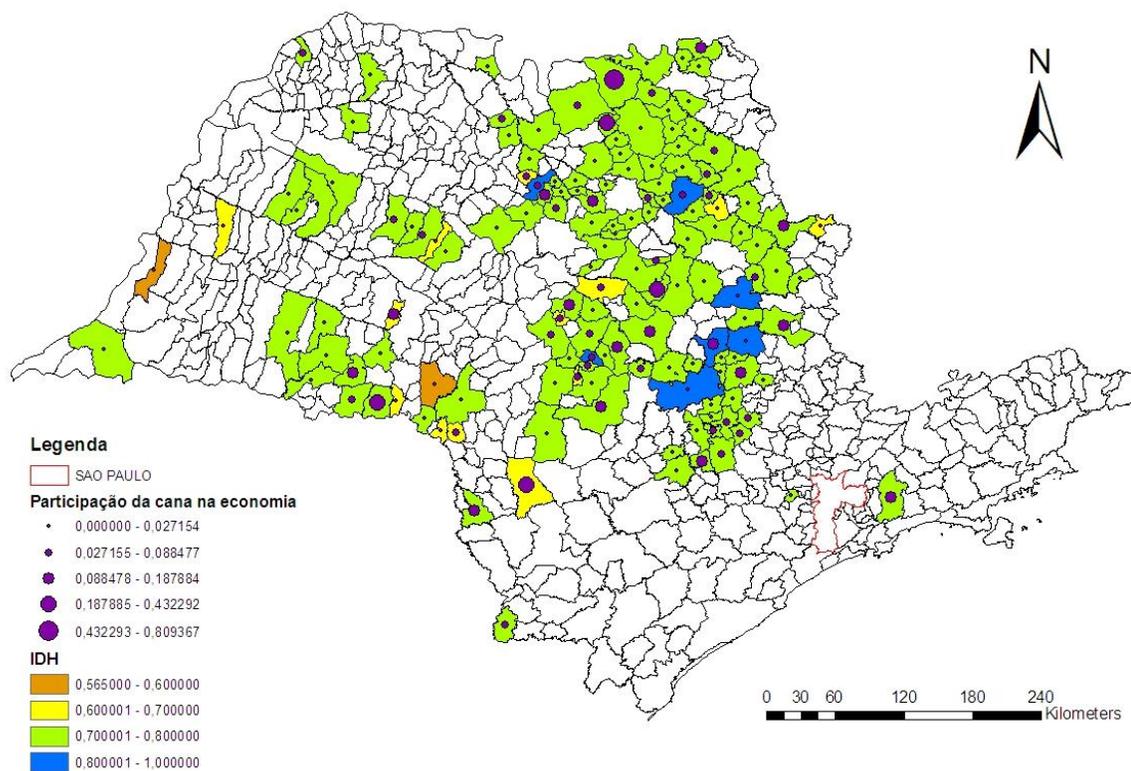


Figura F3 - IDH e participação da cana na economia – São Paulo – 1991.

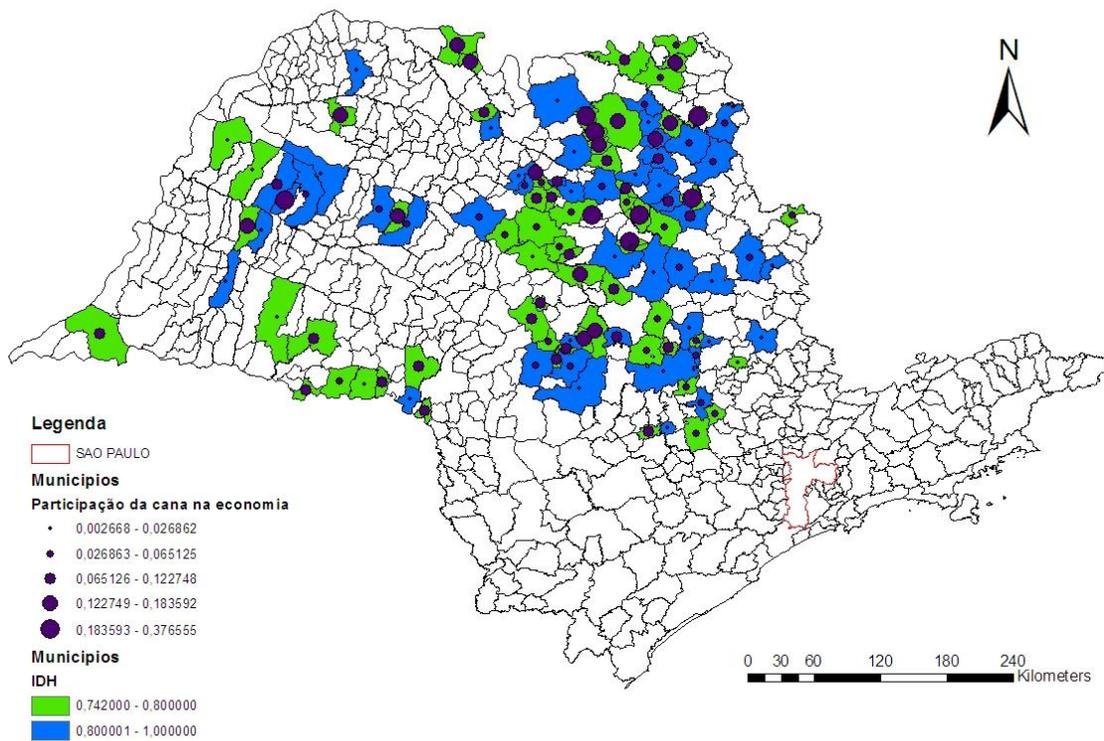


Figura F4 - IDH e participação da cana na economia – São Paulo – 2000.

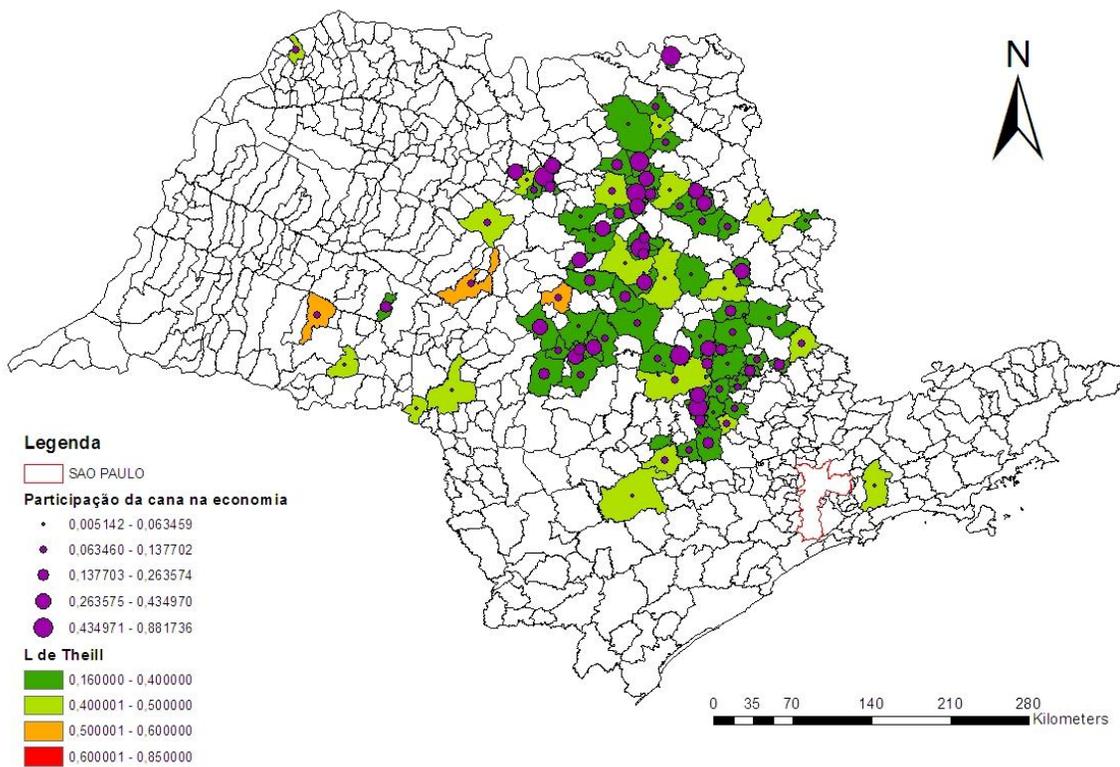


Figura F5 – L de Theil e participação da cana na economia – São Paulo – 1970.

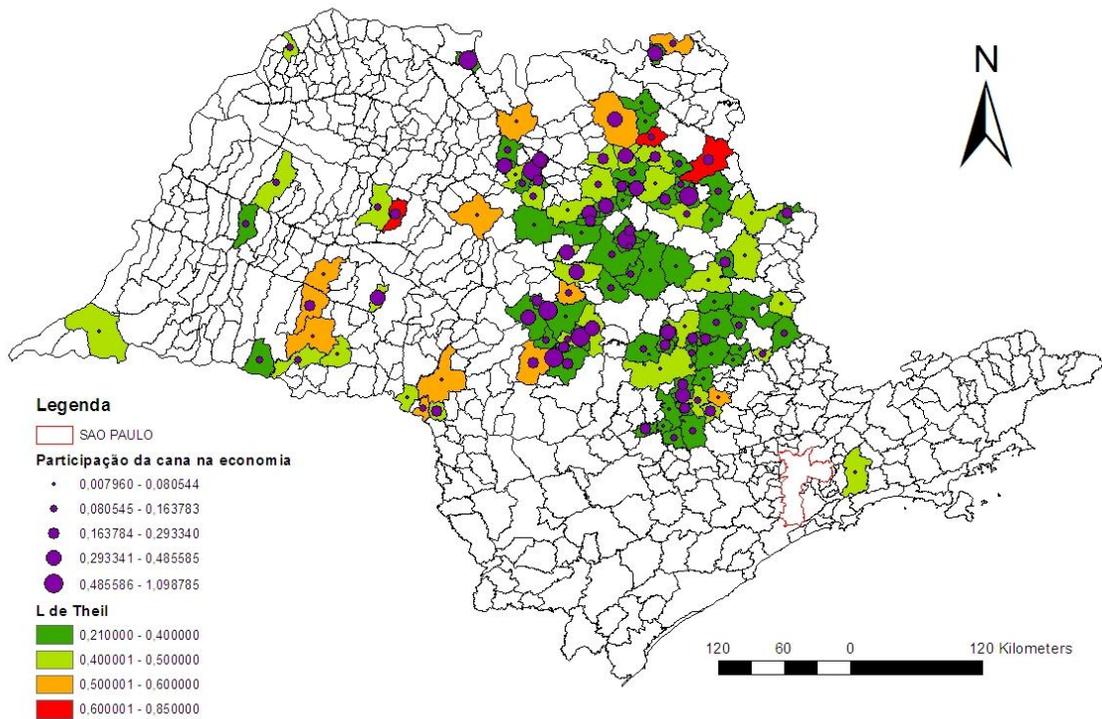


Figura F6 - L de Theil e participação da cana na economia – São Paulo – 1980.

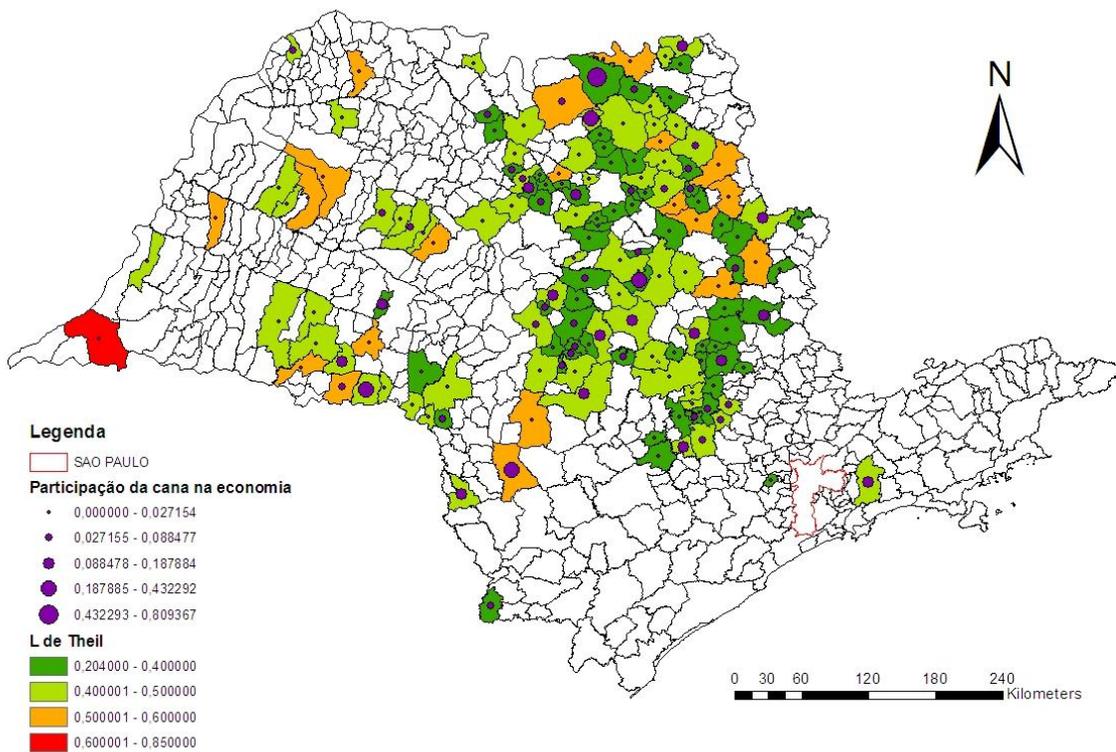


Figura F7 - L de Theil e participação da cana na economia – São Paulo – 1991.

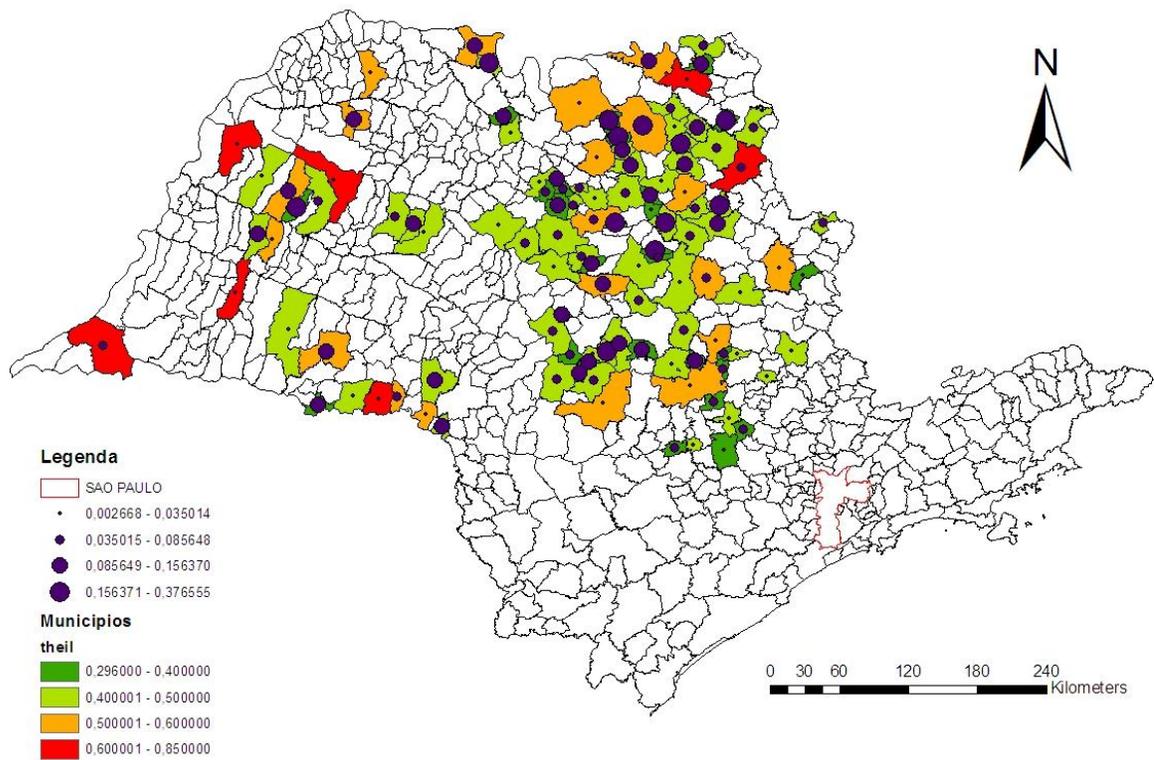


Figura F8 - L de Theil e participação da cana na economia – São Paulo – 2000.
Alagoas

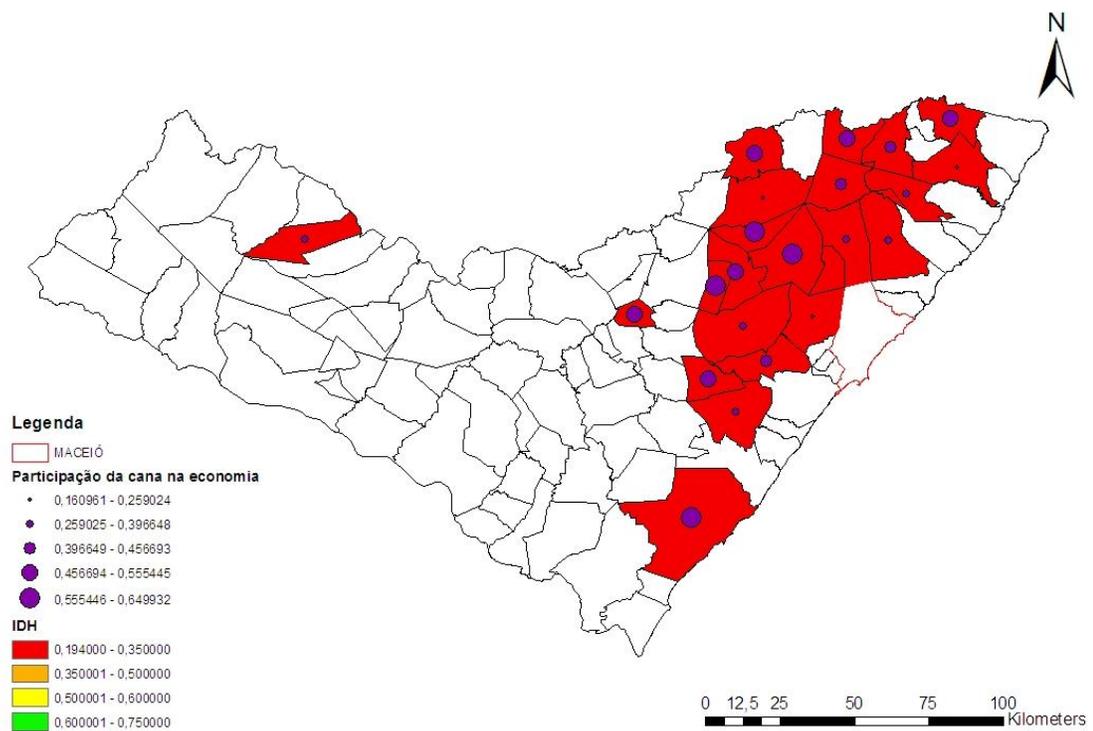


Figura F9 - IDH e participação da cana na economia – Alagoas – 1970.

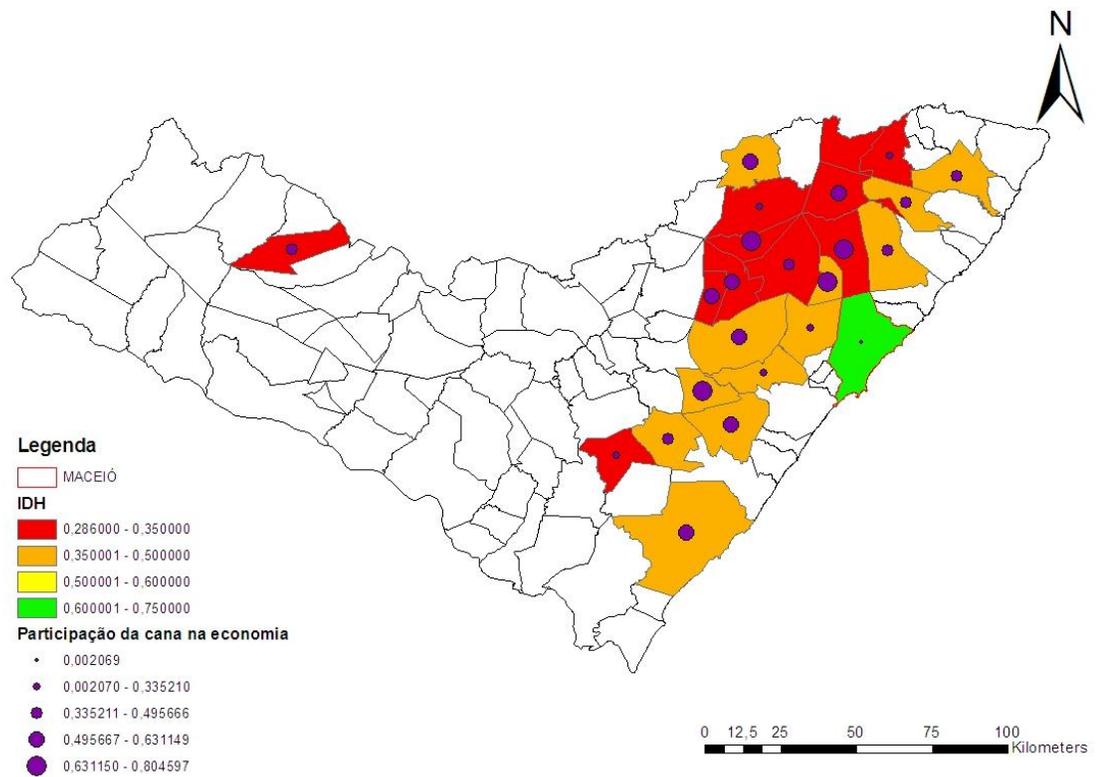


Figura F10 - IDH e participação da cana na economia – Alagoas – 1980.

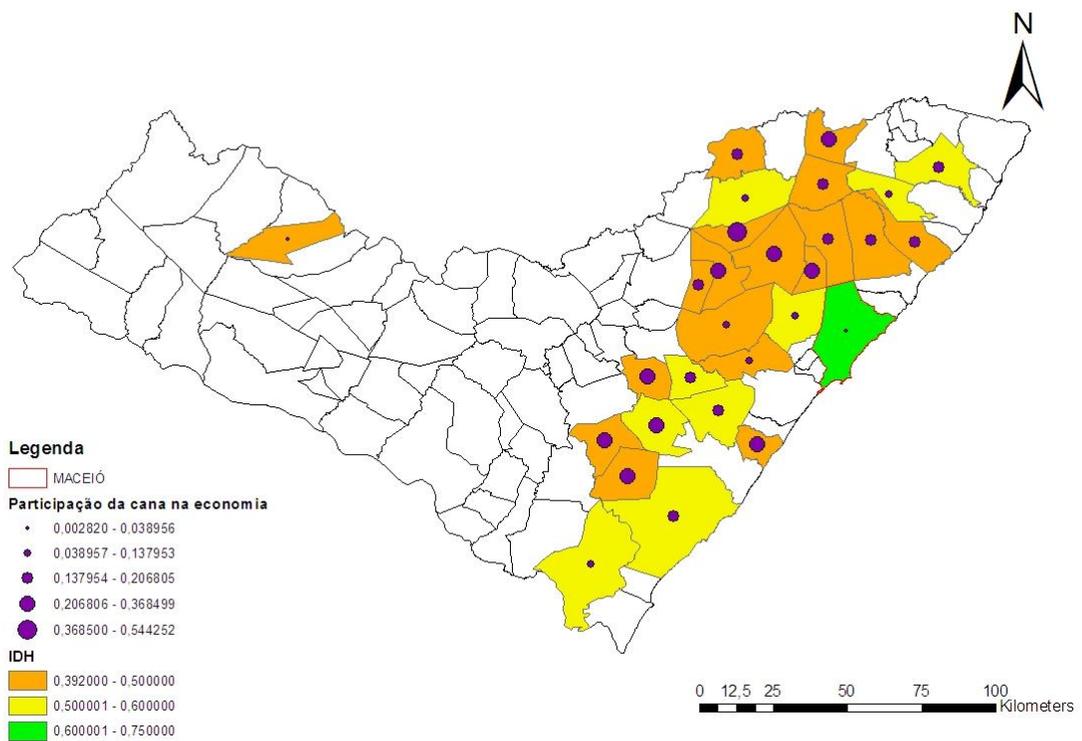


Figura F11 - IDH e participação da cana na economia – Alagoas – 1991.

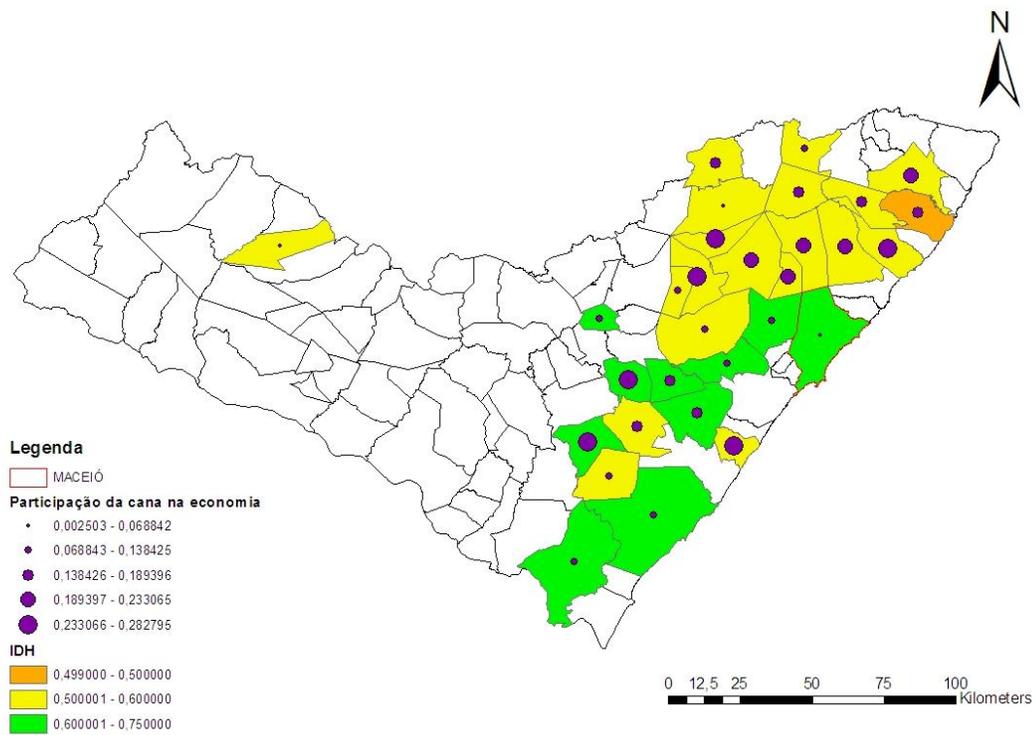


Figura F12 - IDH e participação da cana na economia – Alagoas – 2000.

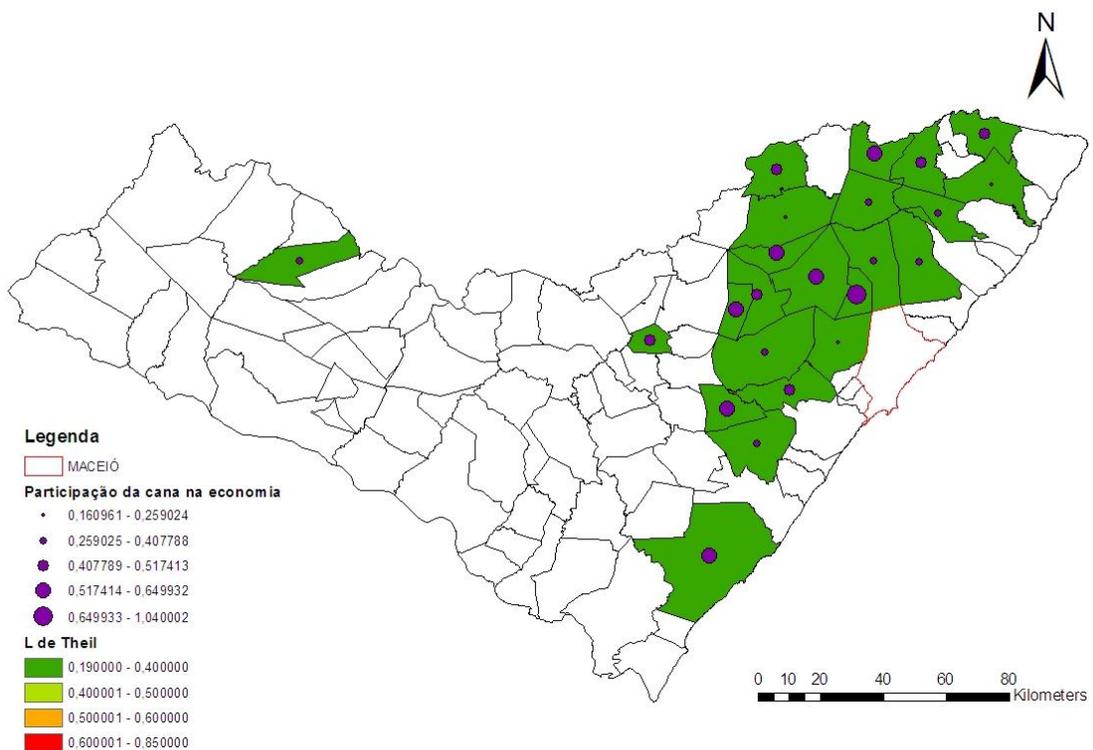


Figura F13 – L de Theil e participação da cana na economia – Alagoas – 1970.

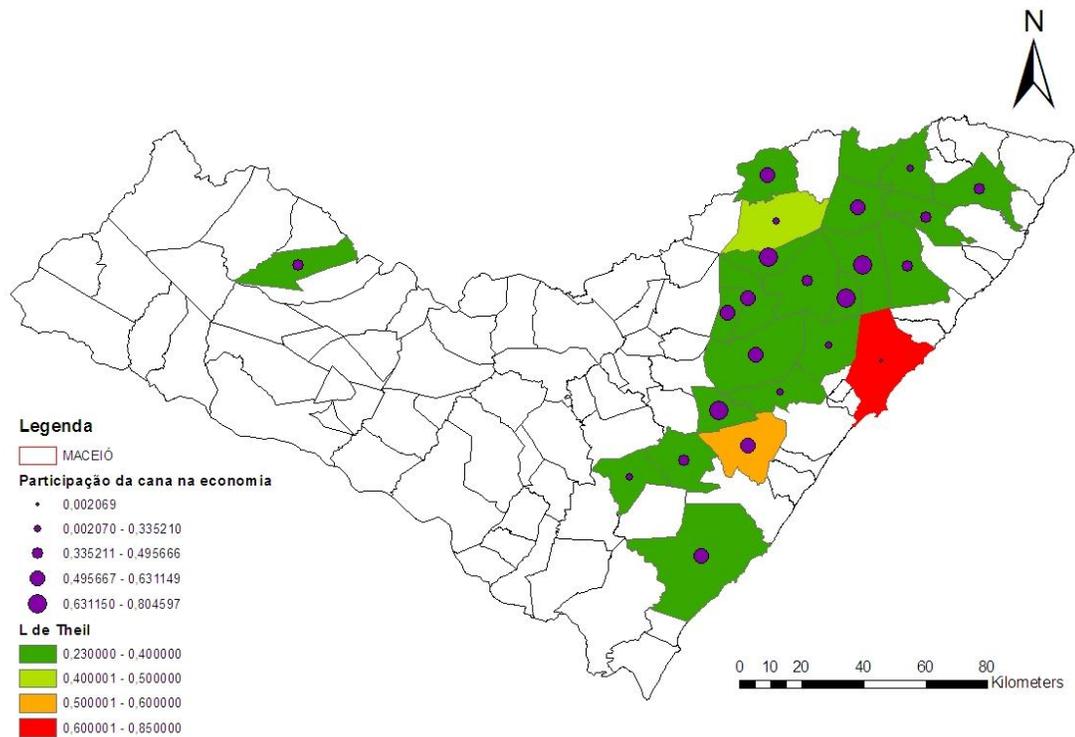


Figura F14 – L de Theil e participação da cana na economia – Alagoas – 1980.

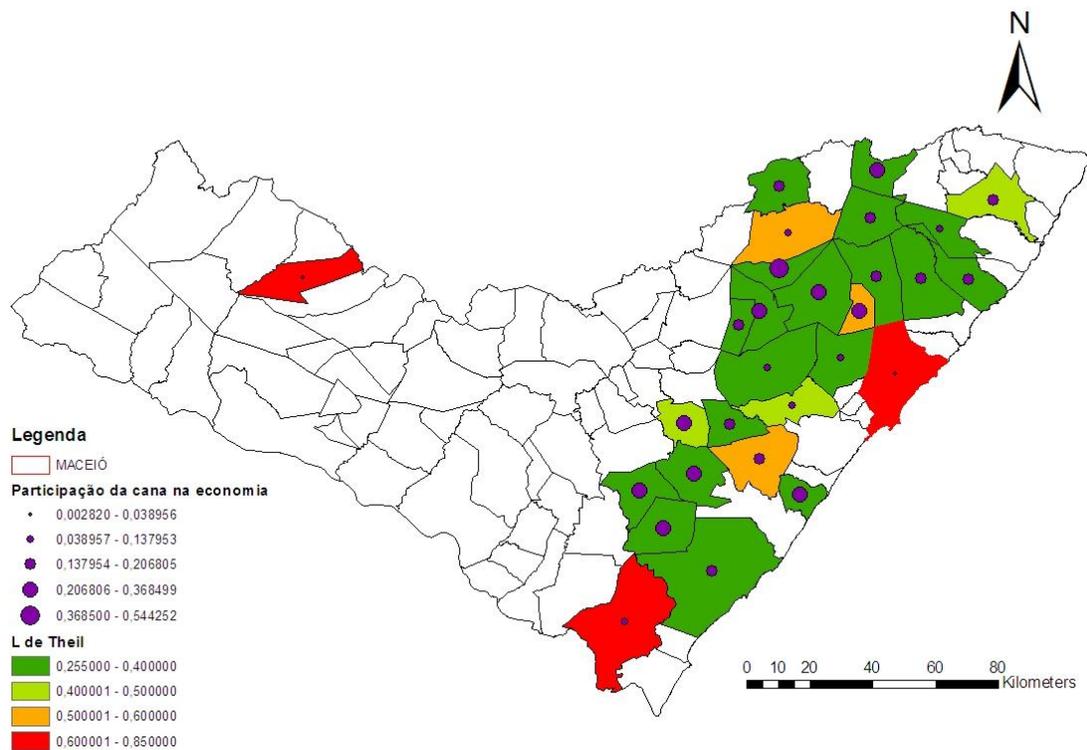


Figura F15 – L de Theil e participação da cana na economia – Alagoas – 1991.

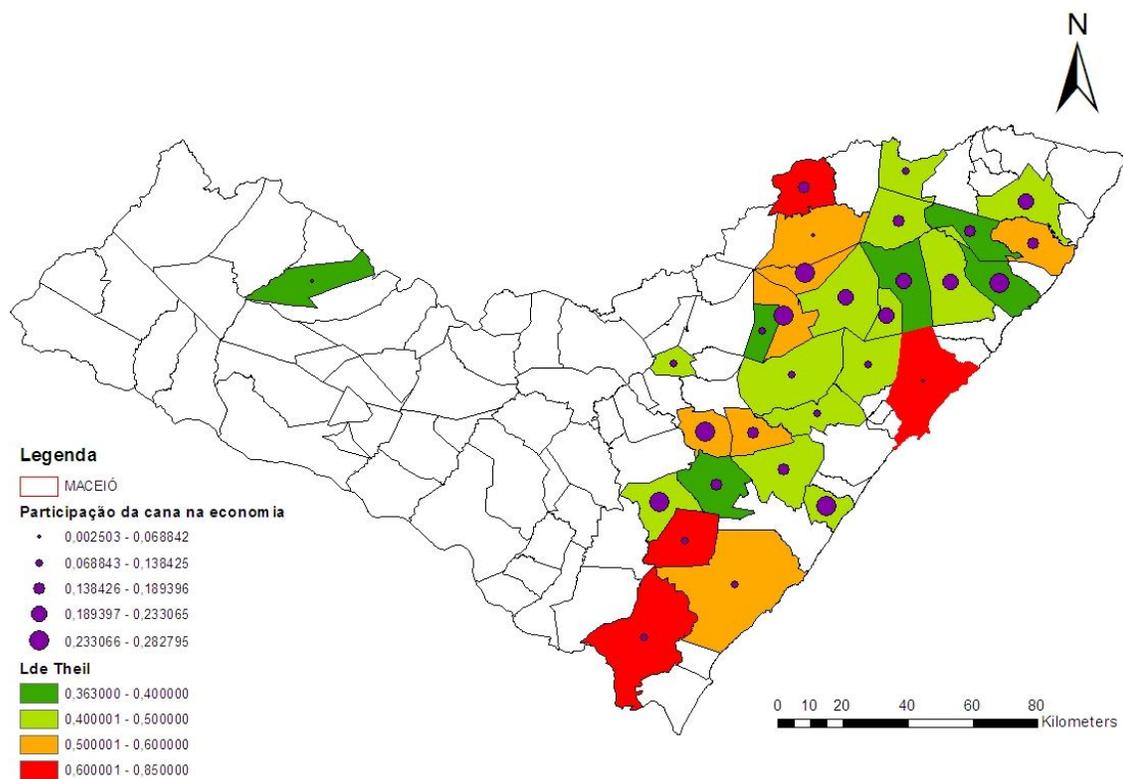


Figura F16 – L de Theil e participação da cana na economia – Alagoas – 2000.

Anexo I

Tabela An1– C Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de São Paulo entre 1970-1980 – valores sem transformação de Oliveira (2011)

Variável	1970		1980	
	Com cana	Sem cana	Com cana	Sem cana
L de Theil	0,33 ± 0,09	0,383 ± 0,092	0,396 ± 0,100	0,442 ± 0,128
Pessoas pobres	57,91 ± 11,35	70,2 ± 14,3	17,2 ± 8,4	31,6 ± 13,3
Mortalidade infantil	79,73 ± 13,89	92,93 ± 20,04	53,52 ± 11,51	59,2 ± 12,9
Analfabetos	24,70 ± 5,80	30,20 ± 8,00	19,8 ± 5,0	22,7 ± 6,0
IDH	0,50 ± 0,05	0,44 ± 0,07	0,708 ± 0,035	0,659 ± 0,066
Esperança de vida	55,45 ± 2,29	53,40 ± 3,06	60,35 ± 2,41	59,2 ± 2,5
Iluminação elétrica	74,50 ± 15,00	47,50 ± 22,90	92,3 ± 6,3	79,0 ± 16,2
Instalações sanitárias	35,30 ± 21,10	15,20 ± 17,90	54,0 ± 20,0	27,4 ± 23,6

Tabela An2 - Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de São Paulo entre 1991-2000 – valores sem transformação de Oliveira (2011)

Variável	1991		2000	
	Com cana	Sem cana	Com cana	Sem cana
L de Theil	0,41 ± 0,08	0,46 ± 0,09	0,453 ± 0,09	0,475 ± 0,09
Pessoas pobres	16,21 ± 7,77	28,8 ± 12,9	15,90 ± 5,37	22,8 ± 10,12
Mortalidade infantil	24,2 ± 5,53	29,0 ± 8,63	13,66 ± 3,58	16,18 ± 5,4
Analfabetos	14,51 ± 3,35	17,0 ± 4,6	9,96 ± 2,63	11,7 ± 3,66
IDH	0,742 ± 0,02	0,716 ± 0,04	0,793 ± 0,02	0,773 ± 0,035
Esperança de vida	69,61 ± 2,09	67,95 ± 2,85	72,57 ± 2,06	71,3 ± 2,67
Iluminação elétrica	96,00 ± 5,40	93,7 ± 7,2	97,40 ± 3,90	96,3 ± 5,3
Instalações sanitárias	74,0 ± 17,0	50,1 ± 23,0	84,40 ± 11,80	63,2 ± 21,7

Tabela An3– Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de Alagoas entre 1970-1980 – valores sem transformação de Oliveira (2011)

Variável	1970		1980	
	Com cana	Sem cana	Com cana	Sem cana
L de Theil	0,314 ± 0,054	0,342 ± 0,118	0,332 ± 0,066	0,391 ± 0,132
Pessoas pobres	89,76 ± 2,88	95,60 ± 3,36	70,61 ± 5,74	83,20 ± 6,97
Mortalidade infantil	191,93 ± 14,60	207,85 ± 29,49	153,47 ± 17,60	162,15 ± 32,
Analfabetos	69,13 ± 7,93	69,90 ± 7,72	64,58 ± 6,47	64,54 ± 8,32
IDH	0,265 ± 0,025	0,227 ± 0,035	0,357 ± 0,035	0,314 ± 0,046
Esperança de vida	48,67 ± 1,54	0,342 ± 0,118	52,83 ± 2,02	51,96 ± 3,61
Iluminação elétrica	71,39 ± 13,28	10,01 ± 9,93	40,03 ± 13,43	22,84 ± 12,88
Instalações sanitárias	0,0 ± 0,0	0,62 ± 3,04	0,71 ± 2,41	0,98 ± 4,87

Tabela An4 - Comparação das médias dos indicadores entre municípios com cana e sem cana para o estado de Alagoas entre 1970-1980 – valores sem transformação de Oliveira (2011)

Variável	1991		2000	
	Com cana	Sem cana	Com cana	Sem cana
L de Theil	0,385 ± 0,11	0,434 ± 0,108	0,496 ± 0,096	0,591 ± 0,131
Pessoas pobres	77,55 ± 6,16	82,23 ± 6,78	73,59 ± 6,48	77,71 ± 6,88
Mortalidade infantil	75,91 ± 13,01	87,34 ± 14,52	50,68 ± 13,28	55,37 ± 12,59
Analfabetos	56,74 ± 8,03	58,94 ± 7,99	42,28 ± 6,40	45,20 ± 6,25
IDH	0,483 ± 0,04	0,455 ± 0,045	0,589 ± 0,042	0,569 ± 0,042
Esperança de vida	57,59 ± 2,61	55,44 ± 2,66	63,41 ± 3,20	62,22 ± 2,87
Iluminação elétrica	71,39 ± 13,28	54,58 ± 17,04	86,42 ± 7,41	78,71 ± 11,91
Instalações sanitárias	1,75 ± 5,87	1,87 ± 7,82	13,43 ± 17,21	7,93 ± 15,81