



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**ANÁLISE ESPACIAL E ESTATÍSTICA DA METODOLOGIA DE
CONSTRUÇÃO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL:
RELATIVO À ÁREA URBANA DE SÃO JOSÉ DOS CAMPOS - SP
(CENSO IBGE 1991)**

*Patrícia C. Genovez
Norton Roberto Caetano
Roberto Duque Estrada*

Projeto FAPESP - Pesquisas em Políticas Públicas
Dinâmica Social, Qualidade Ambiental e Espaços Intra-urbanos
em São Paulo: Uma Análise Sócio-Espacial

INPE
São José dos Campos
2000

SUMÁRIO

	Pág.
CAPÍTULO 1 - INTRODUÇÃO	4
1.1 – REDUÇÃO DO UNIVERSO DE VARIÁVEIS.....	4
CAPÍTULO 2 – OBJETIVOS	6
2.1 – OBJETIVOS GERAIS.....	6
2.2 – OBJETIVOS ESPECÍFICOS.....	6
CAPÍTULO 3 – METODOLOGIA	7
3.1 – CÁLCULO DO ÍNDICE COMPOSTO DE EXCLUSÃO/INCLUSÃO SOCIAL.....	8
3.2 – ANÁLISE ESPACIAL E ESTATÍSTICA DA SENSIBILIDADE DOS INDICADORES	9
3.3 –CÁLCULO DO ÍNDICE COMPOSTO DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO.....	12
3.3.1 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTRAUTOPICO (EX_INTRAU).....	12
3.3.2 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTERUTOPICO (EX_INTERU).....	14
3.3.3 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTERVARIÁVEIS (EX_INTERVAR).....	15
3.4 - - ANÁLISE DOS RESÍDUOS.....	16
CAPÍTULO 4 – RESULTADOS	17
4.1. – REDUÇÃO DAS VARIÁVEIS INTRA-UTÓPICAS.....	18
4.1.1 – REDUÇÃO DAS VARIÁVEIS INTRA-AUTONOMIA DE RENDA.....	18
4.1.2 – REDUÇÃO DAS VARIÁVEIS INTRA-DESENVOLVIMENTO HUMANO.....	25
4.1.3 – REDUÇÃO DAS VARIÁVEIS INTRA-QUALIDADE DE VIDA.....	32
4.1.4 – REDUÇÃO DAS VARIÁVEIS INTRA-EQUIDADE.....	39
4.1.5 – ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTRA-UTÓPICO.....	44
4.2 - REDUÇÃO INTER-UTÓPICA.....	44
4.2.1 - ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTER-UTÓPICO.....	44
4.3 - REDUÇÃO INTER-VARIÁVEIS.....	51
4.3.1 - ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTER-VARIÁVEIS.....	51
44 – APRESENTAÇÃO DOS MAPAS MÍNIMOS.....	57
4.5 – ANÁLISE ESPACIAL DE ÁREAS	57
CAPÍTULO 5 – CONCLUSÕES	64
CAPÍTULO 6 – REFERÊNCIAS BIBLIOGRÁFICAS	68

CAPÍTULO 1 - INTRODUÇÃO

1.1 - REDUÇÃO DO UNIVERSO DE VARIÁVEIS

Fenômenos socioambientais mensurados através de índices compostos exigem a disponibilidade de dados muitas vezes ausentes ou coletados com baixa frequência, como os dados do Censo. A relativa composição e eficiência destes Índices está diretamente ligada às variáveis selecionadas e às possíveis interações, correlações, existentes entre as mesmas, fatos que propõe questões como:

- I. O que explicar (variável dependente)?
- II. Quais os fatores (variáveis independentes) relevantes à interpretação de tal problema?
- III. Como contornar a ausência de dados sistematizados e a baixa frequência de coleta dos mesmos?

Os métodos Estatísticos permitem discriminar as variáveis independentes mais correlacionadas à variável dependente, provendo maior confiabilidade devido a eliminação de informações redundantes proporcionando a geração de um índice composto mínimo para a exclusão/inclusão social.

A coerência deste índice mínimo será verificada, comparativamente, em relação ao índice original. A semelhança dos “Mapas Mínimos” com o original é um forte indicativo de coerência entre os índices. No entanto, as modificações numéricas e qualitativas realizadas sobre o índice original podem causar mudanças globais e locais, evidenciadas no relacionamento espacial dos dados quando analisados no território.

A aplicação de técnicas de Análise Espacial de Áreas evidenciando agrupamentos espaciais dotados de comportamentos específicos, torna-se metodologicamente importante para demonstrar se o relacionamento espacial dos dados no território manteve-se ou não semelhante ao original, após a redução das variáveis.

A validação do método proposto poderá, através da redução do índice composto, tornar-se uma alternativa para a baixa frequência de sistematização e coleta dos dados, simplificando e tornando a coleta e produção dos índices mais econômica possibilitando atualizações mais frequentes. Deste modo, o melhor acompanhamento da movimentação dos índices expressos no território poderá auxiliar o monitoramento dos resultados de políticas e intervenções no espaço intra-urbano.

Portanto, a redução das variáveis componentes de um índice composto simplifica a representação, porém, decisões sobre quais fatores devem ser incluídos, como deveriam ser ponderados e como o índice deve ser composto acarretam mudanças substanciais nos resultados (Maxwell, 1999). Os testes a serem realizados com dados referentes a 1991, 1996 e, possivelmente 2000, proverão material para o estudo e possível esclarecimento destas questões. O documento elaborado baseia-se apenas nos dados de 1991.

A utilização de dados produzidos em todo o território nacional pelo Censo, coletados por setores censitários (**Figura 1.1**), promove a possibilidade de geração e validação de um método genérico passível de aplicação em diferentes regiões socioambientais do país e provavelmente do mundo.

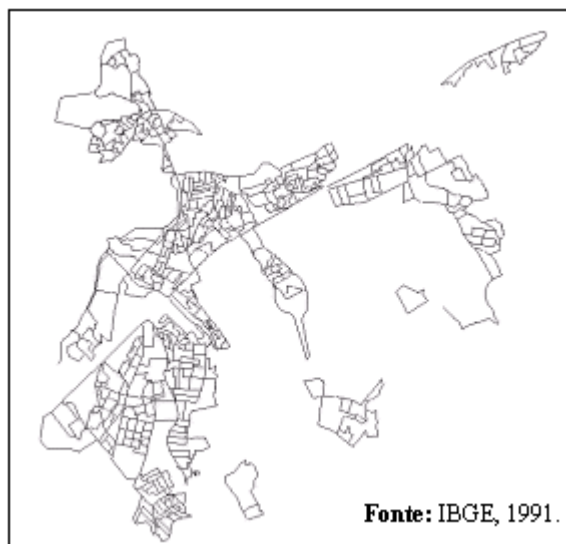


Figura 1.1 – Setores Censitários referentes a área urbana do Município de São José dos Campos – SP.

A análise das propriedades estatísticas das variáveis, a redução de redundâncias entre as mesmas, a produção de “Mapas Mínimos”, a possibilidade de análises mais frequentes (inter-censo), a geração de Políticas Públicas globais e locais, a possibilidade de um método genérico e adaptável à diferentes realidades socioambientais são algumas das possíveis contribuições deste trabalho.

CAPÍTULO 2 – OBJETIVOS

2.1 - OBJETIVOS GERAIS

O principal objetivo deste trabalho é aprimorar estudos relacionados à utilização e aprimoramento de indicadores socioeconômicos espacializados em Sistemas de Informação Geográfica (GIS), tendo por base o modelo conceitual e metodológico proposto por Sposati (1996).

Desta forma, torna-se necessária a pesquisa de métodos direcionados à investigação das variáveis disponíveis para as análises, com base nas técnicas de Estatística Multivariada e de Análise Espacial.

2.2 – OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho, derivados de análises quantitativas a serem realizadas após a aplicação da proposta metodológica de Sposati (1996) são:

- I. Analisar os possíveis problemas relativos à estrutura do Índices Compostos propostos por Sposati em São Paulo (1996) quando adaptados para outras realidades urbanas como São José dos Campos;

- II. Buscar uma metodologia de análise estatística para a redução da redundância derivada do uso de variáveis com diferenciados graus de correlação;
- III. Gerar e avaliar “Mapas Mínimos”¹ a partir da redução de variáveis;
- IV. Gerar alternativas para superar o “vazio” de informações existentes no intervalo de 10 anos da coleta do censo, possibilitando a produção de um mapa onde nem todas as variáveis estejam disponíveis.
- V. Planejamento de Políticas Públicas a partir da aplicação do método estatístico proposto à cidade como um todo e à áreas específicas da cidade, objetivando a descoberta e o entendimento das diferenças socioestruturais globais e locais da cidade.

CAPÍTULO 3 – METODOLOGIA

3.1 – CÁLCULO DO ÍNDICE COMPOSTO DE EXCLUSÃO/INCLUSÃO SOCIAL

A construção do índice composto de exclusão/inclusão social adotado por Sposati (1996) parte da escolha de quatro campos conceituais, Autonomia de Renda, Qualidade de Vida, Desenvolvimento Humano e Equidade, compostos por índices e indicadores específicos.

O **Quadro 3.1** apresenta um esquema da composição dos Índices e Indicadores com variáveis disponíveis no censo para São José dos Campos.

¹ *Mapa Mínimo* é entendido, neste contexto, como um mapa gerado a partir de uma redução estatística de variáveis realizada a partir do conceito do mapa original, o Mapa de São Paulo de 1996 (Sposati, 1996).

Quadro 3.1 - Variáveis presentes na composição do Índice de Exclusão/Inclusão Social

FONTE	INDICADORES	ÍNDICES		CAMPO	I E X C L U S Ã O / I N C L U S Ã O S O C I A L
CENSO IBGE	lexi Chefes de família abaixo da linha de Pobreza (sem Rendimento) lexi Chefe de Família na Linha de Pobreza (com ganho até 2 SM)	lex Precária Condição de Sobrevivência		lex AUTONOMIA DE RENDA DOS CHEFES DE FAMÍLIA	
	lexi sem Rendimento lexi até 0,5 SM lexi de 0,5 até 1 SM lexi de 1 a 2 SM lexi de 2 a 3 SM lexi de 3 a 5 SM lexi de 5 à 10 SM lexi de 10 a 15 SM lexi de 15 a 20 SM lexi mais de 20 SM	lex de Distribuição de Renda dos Chefes de Família			
CENSO IBGE	lexi Chefes de Família não Alfabetizados lexi Escolaridade Precária (de 1 à 3 anos de estudo) lexi de 4 a 7 anos de estudo lexi de 8 a 10 anos de estudo lexi de 11 a 14 anos de estudo lexi mais de 15 anos de estudo	lex de Desenvolvimento Educacional		lex DESENVOLVIMENTO HUMANO	
	lexi Alfabetização Precoce (com 5 a 9 anos) lexi Alfabetização Tardia (de 10 a 14 anos)	lex Estímulo Educacional			
	lexi não Alfabetizados lexi Alfabetização Precária	lex Escolaridade Precária			
	lexi População acima de 70 anos	lex Longevidade			
CENSO IBGE	lexi Precário Abastecimento de Água lexi Precário Instalação sanitária (Esgoto) lexi Precário Tratamento do Lixo	lex Qualidade Ambiental	lex Qualidade Domiciliar	lex QUALIDADE DE VIDA	
	lexi Propriedade Domiciliar				
	lexi Densidade Habitacional lexi Condições de Privacidade lexi Conforto Sanitário	Conforto Domiciliar			
	lexi Habitação Precária				
CENSO IBGE	lexi Mulheres não Alfabetizadas lexi Concentração de Mulheres Chefes de Família			lex EQUIDADE	

Os índices percentuais, gerados para cada uma das quatro Utopias definidas, são escalonados a partir de uma transformação linear em dois intervalos; de 0 a 1 associados às características positivas de inclusão e de -1 a 0, às características negativas de exclusão. Posteriormente, estes índices intermediários são somados e reescalonados de 1 à -1, abrangendo a escala da exclusão e da inclusão gerando o Índice Composto de Exclusão/Inclusão Social para o Município. O valor zero (0) é atribuído aos padrões básicos de inclusão definidos para cada indicador. Neste estudo, o padrão proposto para São Paulo será mantido para São José dos Campos².

² A construção destes referenciais deve ser um processo coletivo e organizado o que inviabiliza sua realização dentro do escopo deste trabalho que objetiva avaliar questões metodológicas em seus aspectos mais operacionais, justificando, desta forma, a manutenção dos padrões de inclusão de São Paulo como referência para São José dos Campos.

3.2 – ANÁLISE ESPACIAL E ESTATÍSTICA DA SENSIBILIDADE DOS INDICADORES

A partir dos dados provenientes do Mapa de Exclusão/Inclusão Social do Município de São José dos Campos, pretende-se chegar a novas formas de calcular a exclusão/inclusão social através de um conjunto de variáveis mínimo, de tal forma que os índices reduzidos mantenham coerência com o índice original.

O primeiro método investiga a importância parcial das variáveis em relação ao universo das quatro Utopias componentes do Índice Composto de Exclusão/Inclusão Social. Este procedimento inicia-se com a redução das variáveis intrautópicas respeitando, hierarquicamente, cada utopia separadamente. Para tal, serão utilizados sequencialmente os métodos de matriz de correlação, componentes principais e regressão linear simples e múltipla detalhados em Neter e Wasserman (1974) e Chatterjee e Price (1977).

O segundo método proposto investiga a importância parcial das quatro dimensões utópicas em relação ao Índice Composto de Exclusão/Inclusão propriamente dito. Este procedimento, embora aparentemente semelhante ao anterior, parte das Utopias anteriormente reduzidas e não das variáveis aplicando os mesmos métodos. Este experimento permite verificar qual das quatro Utopias contribui estatisticamente com relevância para o fenômeno da exclusão/inclusão social.

Finalmente, pretende-se analisar os efeitos de uma desvinculação total do universo das Utopias proposto. Para isto, as análises estatísticas partirão das variáveis intrautópicas anteriormente selecionadas, desconsiderando a hierarquia inerente às utopias. A matriz de correlação, a técnica de componentes principais e as regressões lineares serão aplicadas nas variáveis resultantes da seleção diretamente em relação o Índice de Exclusão/Inclusão Social. Este procedimento evidencia a importância relativa das variáveis propriamente ditas.

Os três métodos propostos, apresentados na **Figura 3.1**, serão denominados Índice de Exclusão/Inclusão Social Intrautópico, Interutópico e Intervariáveis. A análise comparativa

entre os novos índices e o original, com base em sua expressão no território, permite averiguar se os efeitos destas reduções revelam ou não cenários muito diferentes na distribuição espacial dos territórios da exclusão/inclusão na cidade. Algumas questões fundamentais a serem investigadas são:

- I. É possível produzir mapas, com um número menor de variáveis, mantendo a coerência com as características originalmente encontradas na expressão territorial do fenômeno quando consideradas todas as variáveis possivelmente envolvidas na produção do mesmo ?
- II. Até que ponto pode-se reduzir o universo de variáveis sem prejudicar as dimensões de entendimento da cidade representadas, neste caso, pelas Utopias Autonomia de Renda, Desenvolvimento Humano, Qualidade de Vida e Equidade ?
- III. Será que estaremos à beira de um “reducionismo” extremo que resumirá o conceito de exclusão/inclusão social à dimensões muito restritas como por exemplo renda, educação, dentre outras o que se mostraria contrário às discussões apresentadas anteriormente para o universo conceitual que definiu a exclusão/inclusão social?
- IV. Ou este “reducionismo”, mantendo coerência quando analisado espacialmente, tornar-se-á uma forma quantitativa alternativa para mensurar a exclusão/inclusão social em períodos de escassez de dados, como os períodos intercenso ?
- V. A inclusão da avaliação de áreas verdes e mortalidade infantil intra-urbana na composição do Índice de qualidade de vida reposicionaria os territórios da cidade com relação a esta dimensão? E com relação ao Índice composto final?

As três metodologias elaboradas para geração dos Índices Compostos de Exclusão/Inclusão Social Mínimos Intrautópico, Interutópico e Intervariáveis serão aplicadas, neste trabalho, para dados referentes ao Censo 1991 em São José dos Campos.

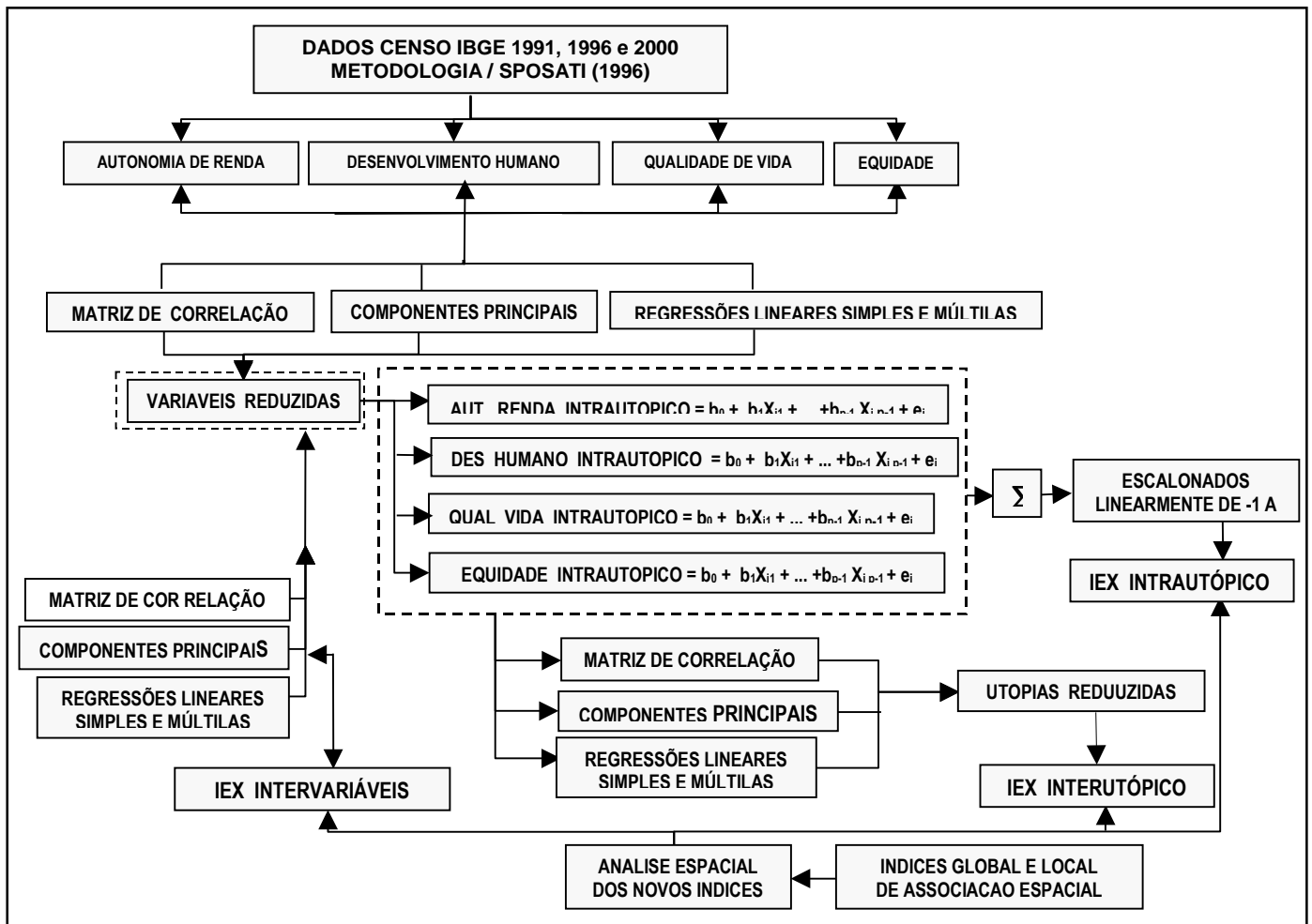


Figura 3.1 – Metodologia utilizada para a geração dos mapas Mínimos através da Redução de Variáveis.

As análises abordarão o fenômeno estudado espacialmente do nível global ao local utilizando os Índices Global e Local de Moran, o Diagrama de Espalhamento de Moran, o Box Map, o Lisa Map e o Moran Map (Câmara et. al., 2000; Fisher et. al., 1996; Bailey e Gatrell, 1995). O objetivo principal é a utilização destas técnicas como um padrão para a comparação dos resultados entre o Índice Composto de Exclusão/Inclusão Social original e os três Índices Mínimos especializados. Tal procedimento verificará se a autocorrelação espacial original entre as áreas de exclusão/inclusão será mantida após a redução do mapa original aos mapas mínimos.

3.3 –CALCULO DO ÍNDICE COMPOSTO DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO

Nesta seção serão detalhadas as três metodologias elaboradas para a geração dos Índices Compostos Mínimos a partir do Índice gerado pela metodologia de Sposati (1996).

3.3.1 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTRAUTOPICO (EX_INTRAU)

Tendo por objetivo principal a investigação da importância parcial das variáveis interiores a cada Utopia (Intrautópicas) em relação ao Índice Composto de Exclusão/Inclusão Social, realiza-se uma análise estatística visando a seleção de variáveis para cada Utopia separadamente, ou seja, todos os métodos descritos serão repetidos para cada uma das utopias, Autonomia de Renda, Desenvolvimento Humano, Qualidade de Vida e Equidade.

Para tal, primeiramente, aplica-se às variáveis contidas em cada uma das Utopias uma análise em função da Matriz de Correlação que proporcionará a eliminação das variáveis independentes (X) menos correlacionadas com a variável dependente (Y) e altamente correlacionadas entre si.

Secundariamente o método de Componentes Principais será aplicado ao conjunto de variáveis restantes, visando, através da análise de sua última componente principal eliminar uma ou mais variáveis evitando redundâncias.

Finalmente realiza-se, com as variáveis independentes restantes (X) e a variável dependente (Y) uma série de regressões lineares simples, par a par, para verificar relativamente as que possuem melhor coeficiente de regressão (r^2). Em seguida realizam-se regressões múltiplas testando as combinações possíveis entre as variáveis X, responsáveis pelos maiores coeficientes (r^2), e a variável Y (Utopia referente). Comparando-se os resultados da

inserção progressiva das variáveis no modelo, em ordem decrescente dos coeficientes (r^2), pode-se verificar o ponto no qual a inserção de variáveis no modelo provoca um aumento insignificante no coeficiente, indicando que tal ou tais variáveis contribuem irrisoriamente para o enriquecimento do modelo e podem ser excluídas do mesmo.

As Regressões Lineares, Múltiplas e Simples, aplicadas aos quatro Índices Intermediários, componentes do Índice Composto de Exclusão/Inclusão Social original (Sposati, 1996), fornecem, como resultado, quatro equações de regressão que possibilitam o cálculo de novos valores para os mesmos, a partir da substituição dos b(s) e dos valores originais das variáveis independentes restantes no modelo.

Esses quatro novos valores obtidos para os Índices são somados linearmente e posteriormente escalonados de -1 a 1 resultando no novo Índice de Exclusão/Inclusão Social Intrautópico (EX_INTRAU). A **Figura 3.2** apresenta a síntese da metodologia a ser utilizada.

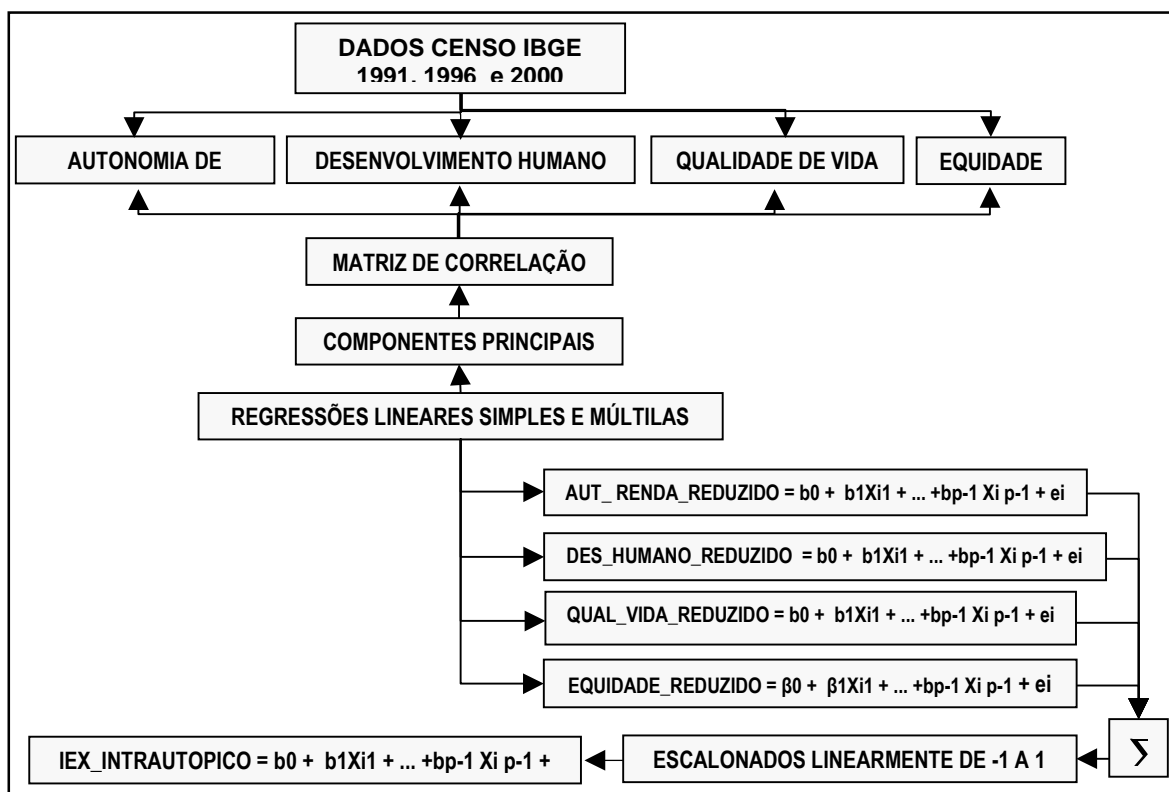


Figura 3.2 – Metodologia específica para o cálculo do Índice de exclusão/Inclusão Intrautópico.

3.3.2 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTERUTOPICO (EX_INTERU)

O Índice de Exclusão/Inclusão Social Interutópico será originado, metodologicamente, de maneira semelhante ao Índice originado no item acima, podendo ser considerado como uma continuação do mesmo. Entretanto o universo de variáveis a ser reduzido será o das novas Utopias geradas após a redução Intrautópica.

Este experimento pretende avaliar a importância entre as Utopias, ou seja, Interutopias, verificando qual das quatro Utopias contribui estatisticamente com relevância para o fenômeno da exclusão/inclusão social.

Aplicar-se-á aos quatro novos Índices gerados pela redução Intrautópica, uma análise da Matriz de Correlação, dos Componentes Principais e de Regressões Lineares Simples e Múltiplas tendo por variável dependente o Índice Original.

As Regressões Lineares, Múltiplas e Simples, aplicadas aos Índices Intrautópicos restantes, componentes do Índice Composto de Exclusão/Inclusão Social Intrautópico, fornecem, como resultado, uma nova equação de regressão que possibilita o cálculo de um novo Índice de Exclusão/Inclusão Social denominado Interutópico, resultante da substituição do(s) $b(s)$ e do(s) valor(es) do(s) Índice(s) Intrautópico(s) selecionado(s) para continuar(em) no modelo.

O novo valor gerado será posteriormente escalonado de -1 à 1 resultando no novo Índice de Exclusão/Inclusão Social Interutópico (EX_INTERU). A **Figura 3.3** apresenta a síntese da metodologia a ser utilizada.

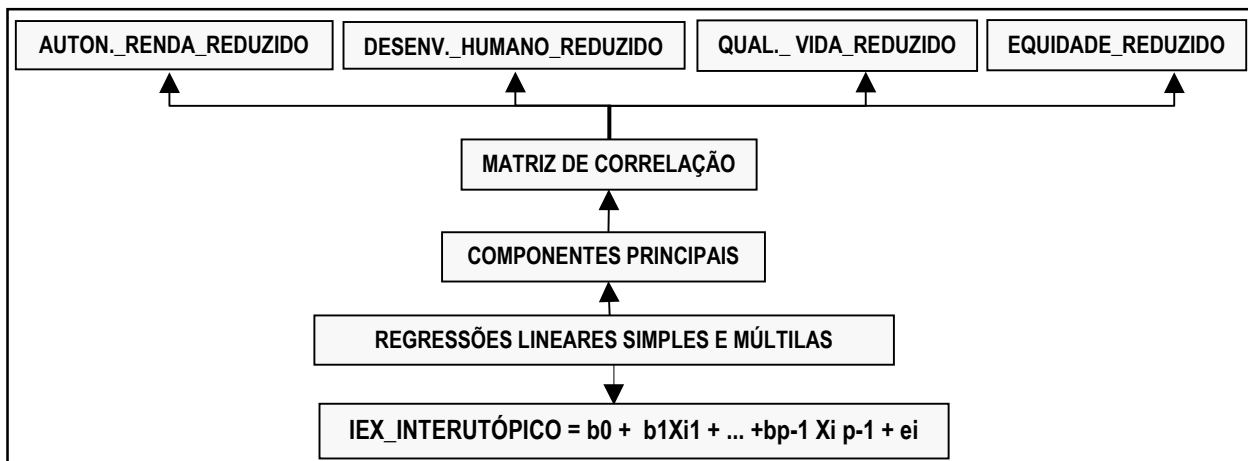


Figura 3.3 – Metodologia específica para o cálculo do Índice de exclusão/Inclusão Interutópico.

3.3.3 - CÁLCULO DO ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL INTERVARIÁVEIS (EX_INTERVAR)

Tendo por objetivo principal a investigação da importância parcial das variáveis em relação ao Índice Composto de Exclusão/Inclusão Social, realiza-se uma análise estatística visando a seleção de variáveis desconsiderando-se totalmente a hierarquia imposta pelas quatro dimensões Utópicas.

De maneira semelhante aos demais Índices calculados aplica-se às variáveis selecionadas no primeiro método, uma análise da Matriz de Correlação, dos Componentes Principais e de Regressões Lineares Simples e Múltiplas em função do Índice Original.

As Regressões Lineares, Múltiplas e Simples, relacionando diretamente, as variáveis selecionadas intrautópicas com o Índice Composto de Exclusão/Inclusão Social original (Sposati, 1996), fornecerão, como resultado, uma equação de regressão que possibilita o cálculo de um novo valor para o Índice, a partir da substituição dos $b(s)$ e dos valores originais das variáveis independentes restantes.

Este novo valor para o Índice Composto de Exclusão/Inclusão Social, obtido pela substituição dos valores das variáveis preservadas nas equações de regressão geradas, são escalonados de -1 à 1 resultando no novo Índice de Exclusão/Inclusão Social Intervariáveis (EX_INTERVAR). A **Figura 3.4** apresenta a síntese da metodologia a ser utilizada.

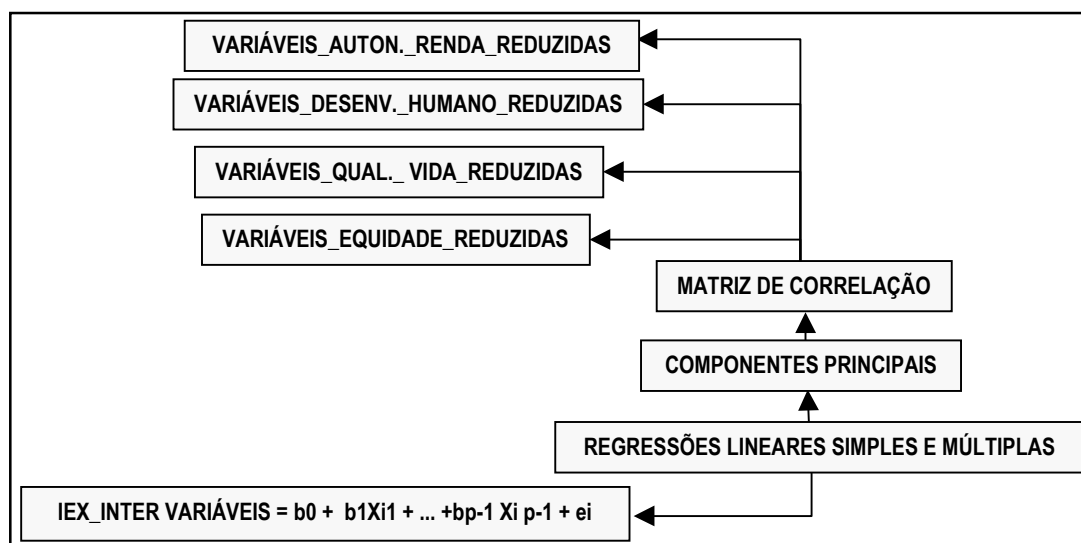


Figura 3.4 – Metodologia específica para o cálculo do Índice de Exclusão/Inclusão Intervariáveis.

3.4- ANÁLISE DOS RESÍDUOS

Como o modelo de regressão contém em si algumas hipóteses sobre o erro, tais como distribuição normal, independência, média igual a zero e variância constante, torna-se necessário uma investigação aprimorada sobre os resíduos provenientes da função de regressão a ser aplicada ao final de todas as análises de regressão realizadas.

A análise dos resíduos será feita a partir de diversos gráficos que mostrem, respectivamente, as seguintes relações:

- I. Resíduos x Observados;

- II. Resíduos x Estimados (\hat{Y});
- III. Observados (Y_i) x Estimados (\hat{Y});
- IV. Distribuição dos Resíduos ao longo da reta de regressão e no histograma;
- V. Resíduos em relação à todas as variáveis que restaram no modelo após as reduções.

CAPÍTULO 4 – RESULTADOS

Este capítulo mostrará os resultados referentes à todas as análises estatísticas e espaciais aplicadas contendo:

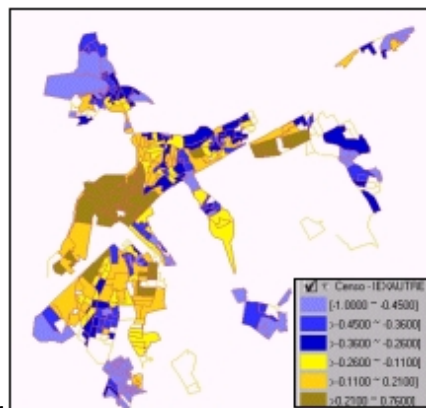
- I. Estatística descritiva das variáveis dependentes (Y) e seu tipo de distribuição;
- II. Variáveis independentes (X) a serem investigadas;
- III. Matriz de Correlação;
- IV. Componentes Principais;
- V. Regressões Lineares Simples e Múltiplas;
- VI. Análise dos Resíduos;
- VII. Apresentação dos “Mapas Mínimos”;
- VIII. Análise Espacial de Áreas para validação dos modelos estatísticos propostos.

Todas as análises acima citadas serão aplicadas e testadas aos três métodos propostos para o cálculo do Índice de Exclusão/Inclusão Social Mínimo, o Intrautópico, o Interutópico e o Intervariáveis.

4.1. – REDUÇÃO DAS VARIÁVEIS INTRA-UTÓPICAS

4.1.1 – REDUÇÃO DAS VARIÁVEIS INTRA-AUTONOMIA DE RENDA

I. Estatística Descritiva:



Descriptive Statistics (trab1.sta)		AUTONOMIA DE RENDA					
	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Minimum	Maximum	Std.Dev.
AUT_REND	350,000	54,470	52,458	56,481	0,000	100,000	19,130

Figura 4.1 – Espacialização da variável dependente Autonomia de Renda em conjunto com sua estatística descritiva.

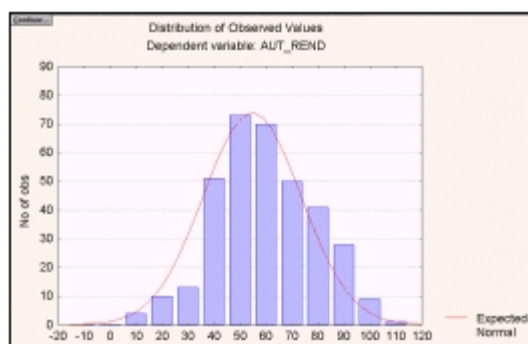


Figura 4.2 - Curva comparativa entre o comportamento do índice de autonomia de renda e a correspondente curva gaussiana de distribuição normal.

II. Matriz de Correlação

A Matriz de Correlação foi aplicada sobre as 11 variáveis iniciais componentes do Índice Autonomia de Renda apresentado no quadro abaixo:

FONTE	INDICADORES	ÍNDICES	CAMPO
CENSO IBGE (1991)	lexi Chefes de família abaixo da linha de Pobreza (sem Rendimento) lexi Chefe de Família na Linha de Pobreza (com ganho até 2 SM	lex Precária Condição de Sobrevivência	lex AUTONOMIA DE RENDA DOS CHEFES DE FAMÍLIA
	lexi sem Rendimento lexi até 0,5 SM lexi de 0,5 até 1 SM lexi de 1 a 2 SM lexi de 2 a 3 SM lexi de 3 a 5 SM lexi de 5 à 10 SM lexi de 10 a 15 SM lexi de 15 a 20 SM lexi mais de 20 SM	lex de Distribuição de Renda dos Chefes de Família	

Analisando-se a Matriz de Correlação, verifica-se a baixa correlação da variável independente de 3 a 5 Salários Mínimos (SM) com a variável dependente Autonomia de Renda. A variável de 10 a 15 SM apresentou-se muito relacionada com as variáveis na Linha da Pobreza (LIN_POBR) e de 15 a 20 SM, as variáveis de 0,5 à 15 SM e de 1 à 2 SM mostraram-se muito relacionadas com LIN_POBR e a variável mais de 20 SM muito relacionada com de 15 à 20 SM.

Portanto, em função da análise da matriz de correlação foram excluídas as variáveis D3_5SM, D10_15SM, D0,5_1SM, D1_2SM e M20_SM, destacadas na Tabela 4.1, por apresentarem baixa correlação com o Índice de Autonomia de Renda ou grande autocorrelação entre si. não sendo boas estimadoras da variável dependente.

Tabela 4.1 - Matriz de Correlação

Correlations (trab1.sta)												
	A_LN_POB	LIN_POBR	DE2_3SM	DE3_5SM	DE5_10SM	D10_15SM	D15_20SM	M20_SM	A_0_5_SM	D0_5_1SM	D1_2SM	AUT_REND
A_LN_POB	1,000	0,270	0,154	0,100	-0,206	-0,352	-0,335	-0,291	0,286	0,201	0,226	-0,528
LIN_POBR	0,270	1,000	0,548	0,061	-0,601	-0,703	-0,638	-0,532	0,627	0,844	0,897	-0,934
DE2_3SM	0,154	0,548	1,000	0,536	-0,290	-0,679	-0,673	-0,578	0,230	0,259	0,668	-0,602
DE3_5SM	0,100	0,061	0,536	1,000	0,302	-0,401	-0,515	-0,531	-0,012	-0,117	0,187	-0,168
DE5_10SM	-0,206	-0,601	-0,290	0,302	1,000	0,522	0,276	0,023	-0,353	-0,497	-0,553	0,602
D10_15SM	-0,352	-0,703	-0,679	-0,401	0,522	1,000	0,891	0,665	-0,416	-0,493	-0,708	0,814
D15_20SM	-0,335	-0,638	-0,673	-0,515	0,276	0,891	1,000	0,815	-0,379	-0,430	-0,655	0,757
M20_SM	-0,291	-0,532	-0,578	-0,531	0,023	0,665	0,815	1,000	-0,324	-0,371	-0,534	0,636
A_0_5_SM	0,286	0,627	0,230	-0,012	-0,353	-0,416	-0,379	-0,324	1,000	0,524	0,381	-0,696
D0_5_1SM	0,201	0,844	0,259	-0,117	-0,497	-0,493	-0,430	-0,371	0,524	1,000	0,557	-0,758
D1_2SM	0,226	0,897	0,668	0,187	-0,553	-0,708	-0,655	-0,534	0,381	0,557	1,000	-0,825
AUT_REND	-0,528	-0,934	-0,602	-0,168	0,602	0,814	0,757	0,636	-0,696	-0,758	-0,825	1,000

 Excluídos pela Matriz de Correlação

 Excluídos pelas Componentes Principais

III. COMPONENTES PRINCIPAIS.

O método de componentes principais (Tabela 4.2) foi utilizado para a identificação de correlação entre as variáveis independentes, com o objetivo de eliminar uma ou mais variáveis. A análise da última componente apresentou o autovalor aproxima-se de zero ($\lambda_6 = 0,199$). Igualando-se a zero a combinação linear das variáveis independentes originais correspondentes a esta componente principal e eliminando-se as variáveis com coeficientes próximos de zero ($< 0,1$ em módulo), chegou-se à seguinte expressão:

$$0,199 \approx 0 = \text{LIN_POBR} * 0,353 - \text{A_0,5_SM} * 0,141 + \text{A_LIN_PO} * 0,044 + \text{DE5_10SM} * 0,151 + \text{D15_20SM} * 0,173 - \text{D2_3SM} * 0,01$$

Sendo assim, das seis variáveis restantes, uma variável passou a ser combinação das outras cinco. E finalmente, analisando as seis variáveis restantes na matriz de correlação, verificou-se que a variável D5_10SM era a mais indicada para ser retirada, devido a sua menor correlação com a variável dependente AUT_RENDA.

$$0,151 * D5_10SM = A_0,5_SM * 141 - D15_20SM * 0,173 - LIN_POBR * 0,353$$

Assim pode-se afirmar que D5_10SM é uma combinação linear das outras variáveis e pode ser excluída.

Tabela 4.2 - Autovetores e Autovalores dos Componentes Principais.

Factor Loadings (Unrotated) (trab1.sta)						
Extraction: Principal components						
(Marked loadings are > .700000)						
	Factor	Factor	Factor	Factor	Factor	Factor
	1	2	3	4	5	6
A_LIN_PO	-0,464	-0,406	0,747	0,232	0,075	0,044
LIN_POB	-0,899	-0,053	-0,215	-0,066	-0,116	0,353
A_0_5_SM	-0,681	-0,422	-0,103	-0,554	0,146	-0,141
DE2_3SM	-0,708	0,599	0,028	0,090	0,361	-0,010
DE5_10SM	0,639	0,325	0,447	-0,514	0,008	0,151
D15_20SM	0,803	-0,382	-0,227	0,078	0,349	0,173
Eigenval	3,042	0,956	0,867	0,643	0,292	0,199
Prp.Totl	0,507	0,159	0,144	0,107	0,049	0,033

IV. REGRESSÃO LINEAR

A realização de regressões lineares a partir de todas as combinações possíveis entre as variáveis evidenciou que o maior R^2 , de 0,9542, foi obtido para a combinação das variáveis A_LN_POB e LIN_POB.

O acréscimo de qualquer das variáveis restantes resultou em um aumento não significativo do R^2 , portanto não foram acrescentadas no modelo. O quadro abaixo apresenta os resultados da regressão.

As superfícies de regressão revelam tendências comportamentais entre as variáveis, ou seja, a existência ou não de interação entre as variáveis resulta em diferentes superfícies de resposta. O modelo genérico de regressão linear não se restringe à respostas de superfície lineares. O termo linear refere-se ao fato de que o parâmetro é linear, mas não necessariamente a superfície de resposta (Neter & Wasserman, 1973).

A superfície de regressão plana simples (Figura 4.3), correspondente ao modelo final da função de regressão, indica a não existência de interação entre as variáveis independentes e uma relação inversamente proporcional entre as mesmas e a variável dependente.

V. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.4**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta, porém a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.5**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.6**.

Na **Figura 4.6**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno.

Já nas **Figuras 4.7 e 4.8** observa-se a distribuição normal dos erros, e a independência das variáveis independentes nas **Figuras 4.9 e 4.10**.

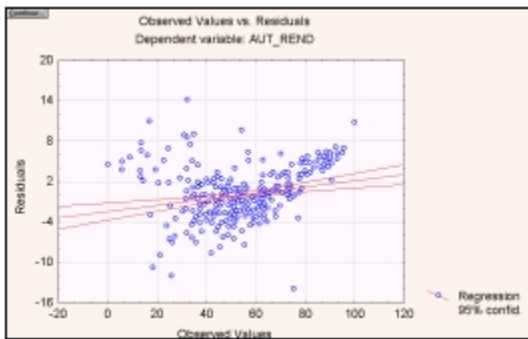


Figura 4.4 - Resíduos x Observados.

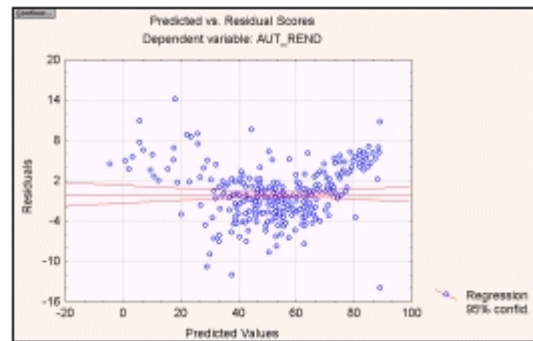


Figura 4.5 - Resíduos x Estimados (\hat{Y}).

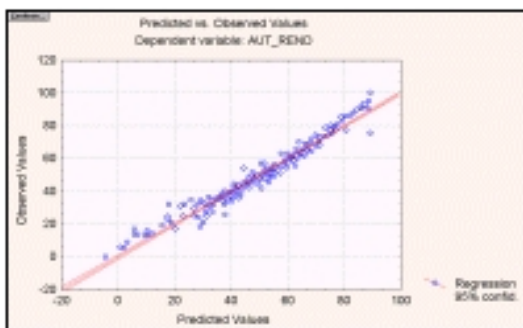


Figura 4.6 - Observados (Y_i) x Estimados (\hat{Y}).

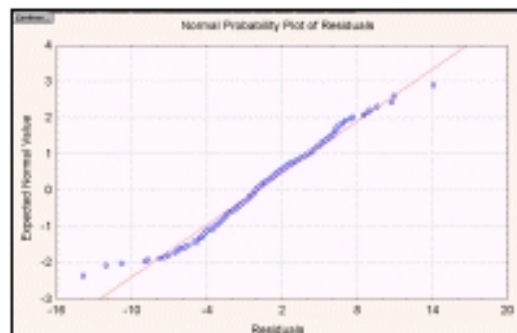


Figura 4.7 - Distribuição normal dos resíduos.

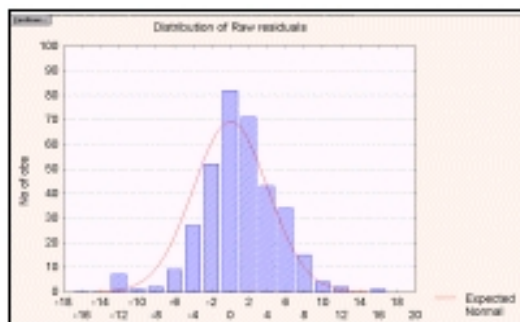


Figura 4.8 - Histograma de distribuição normal dos resíduos.

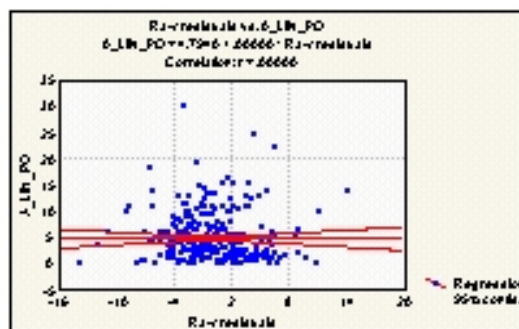


Figura 4.9 Resíduos x A_LIN_PO

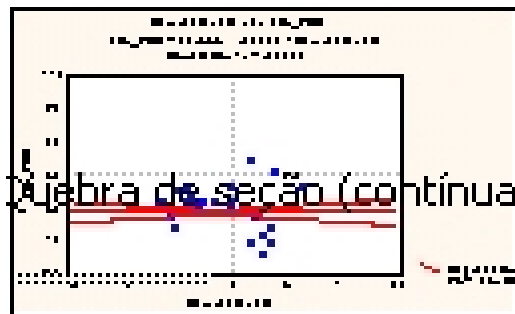
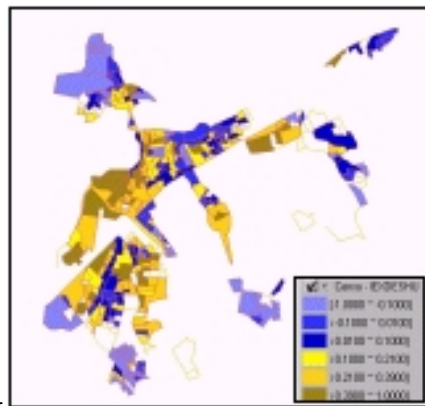


Figura 4.10 -: Resíduos x LIN_POBR

4.1.2 – REDUÇÃO DAS VARIÁVEIS INTRA-DESENVOLVIMENTO HUMANO

I. Estatística Descritiva:



Descriptive Statistics (trab1.sta) DESENVOLVIMENTO HUMANO							
	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Minimum	Maximum	Std.Dev.
DES_HUMA	350,000	60,051	58,292	61,810	0,000	100,000	16,732

Figura 4.11 – Espacialização da variável dependente Desenvolvimento Humano em conjunto com sua estatística descritiva.

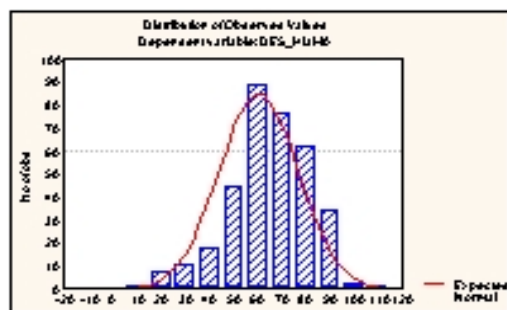


Figura 4.12 - Curva comparativa entre o comportamento do índice de autonomia de renda e a correspondente curva gaussiana de distribuição normal.

II. Matriz de Correlação

A Matriz de Correlação foi aplicada sobre às 9 variáveis iniciais componentes do Índice de Desenvolvimento Humano apresentado abaixo:

FONTE	INDICADORES	ÍNDICES	CAMPO
CENSO IBGE (1991)	lexi Chefes de Família não Alfabetizados lexi Escolaridade Precária (de 1 à 3 anos de estudo) lexi de 4 a 7 anos de estudo lexi de 8 a 10 anos de estudo lexi de 11 a 14 anos de estudo lexi mais de 15 anos de estudo	lex de Desenvolvimento Educativo	lex DESENVOL- VIMENTO HUMANO
	lexi Alfabetização Precoce (com 5 a 9 anos) lexi Alfabetização Tardia (de 10 a 14 anos)	lex Estímulo Educativo	
	lexi não Alfabetizados lexi Escolaridade Precária	lex Escolaridade Precária	
	lexi População acima de 70 anos	lex Longevidade	

Analisando-se a Matriz de Correlação, verifica-se a baixa correlação das variáveis independentes de 8 a 10 anos de Estudo e População acima de 70 anos com a variável dependente Desenvolvimento Humano. As variáveis Alfabetização Precoce (com 5 a 9 anos) e Alfabetização Tardia (de 10 a 14 anos) apresentaram baixa correlação com o Desenvolvimento Humano e alta correlação entre si..

Portanto, em função da análise da matriz de correlação foram excluídas as variáveis de 8 a 10 anos de estudo (D8_10AES), População acima de 70 anos (POP_70), Alfabetização Precoce (A_PREC) e Alfabetização Tardia (A_TARDIA) destacadas na **Tabela 4.5**, por

apresentarem baixa correlação com o índice de Desenvolvimento Humano ou grande autocorrelação entre si, não sendo boas estimadoras da variável dependente.

Tabela 4.5 - Matriz de Correlação

Correlations (trab1.sta)										
	CHF_NALF	ESC_PREC	D4_7AEST	D8_10AES	D11_14AE	M15_AEST	POP_70	A_PRECOC	A_TARDIA	DES_HUMA
CHF_NALF	1,000	0,787	0,447	-0,232	-0,643	-0,532	-0,071	-0,195	0,228	-0,877
ESC_PREC	0,787	1,000	0,654	-0,060	-0,711	-0,679	-0,054	-0,097	0,322	-0,929
D4_7AEST	0,447	0,654	1,000	0,396	-0,575	-0,774	-0,083	0,046	0,273	-0,594
D8_10AES	-0,232	-0,060	0,396	1,000	0,237	-0,382	-0,019	0,291	0,145	0,198
D11_14AE	-0,643	-0,711	-0,575	0,237	1,000	0,526	0,106	0,316	-0,024	0,751
M15_AEST	-0,532	-0,679	-0,774	-0,382	0,526	1,000	0,053	0,157	-0,083	0,615
POP_70	-0,071	-0,054	-0,083	-0,019	0,106	0,053	1,000	-0,147	-0,140	0,064
A_PRECOC	-0,195	-0,097	0,046	0,291	0,316	0,157	-0,147	1,000	0,605	0,254
A_TARDIA	0,228	0,322	0,273	0,145	-0,024	-0,083	-0,140	0,605	1,000	-0,350
DES_HUMA	-0,877	-0,929	-0,594	0,198	0,751	0,615	0,064	0,254	-0,350	1,000



Excluídos pela Matriz de Correlação

Excluídos pela componentes Principais

III. COMPONENTES PRINCIPAIS.

O método de componentes principais (**Tabela 4.6**), aplicado ao Índice de Desenvolvimento Humano (DES_HUM) analisou a última componente na qual o autovalor aproxima-se de zero ($\lambda_9 = 0,138$) objetivando a eliminação de variáveis.

Igualando-se a zero a combinação linear das variáveis independentes originais correspondentes a esta componente principal e eliminando-se as variáveis com coeficientes próximos de zero ($< 0,1$ em módulo), chegou-se à seguinte expressão:

$$0,138 \approx 0 = \text{ESC_PREC} * 0,236 + \text{D4_7AEST} * 0,105 + \text{D11_14AEST} * 0,106 + \text{M15_AEST} * 0,22$$

Sendo assim, das nove variáveis restantes, uma variável passou a ser combinação das outras oito. Analisando as nove variáveis restantes na matriz de correlação, verificou-se que a

variável D4_7AEST era a mais indicada para ser retirada, devido a sua menor correlação com a variável dependente DES_HUM.

$$0,105D4_7AEST = -0,236ESC_PREC - D11_14AEST*0,106 - 0,22*M15AEST$$

Assim pode-se afirmar que D4_7AEST é uma combinação linear das outras variáveis e pode ser excluída.

Tabela 4.6 - Autovetores e Autovalores dos Componentes Principais.

Factor Loadings (Unrotated) (trab1.sta)									
Extraction: Principal components									
(Marked loadings are > .700000)									
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
CHF_NALF	-0,808	-0,216	0,306	0,108	0,333	-0,141	-0,211	-0,136	-0,027
ESC_PREC	-0,917	-0,036	0,166	0,130	0,077	-0,040	0,121	0,190	0,236
D4_7AEST	-0,826	0,294	-0,286	-0,026	-0,248	0,053	0,048	-0,267	0,105
D8_10AES	-0,080	0,674	-0,670	-0,088	0,122	0,058	-0,223	0,110	0,052
D11_14AE	0,807	0,341	-0,123	0,079	0,386	-0,038	0,176	-0,136	0,106
M15_AEST	0,833	-0,108	0,413	0,065	-0,140	0,090	-0,205	-0,047	0,220
POP_70	0,122	-0,221	-0,335	0,905	-0,065	-0,018	-0,018	0,002	-0,021
A_PRECOC	0,146	0,840	0,320	0,121	-0,169	-0,356	-0,008	0,021	-0,026
A_TARDIA	-0,278	0,692	0,508	0,249	0,066	0,334	0,022	0,008	-0,084
Eigenval	3,640	1,950	1,309	0,943	0,399	0,276	0,185	0,160	0,138
Prp.Totl	0,404	0,217	0,145	0,105	0,044	0,031	0,021	0,018	0,015

IV. REGRESSÃO LINEAR

A realização de regressões lineares a partir de todas as combinações possíveis entre as variáveis diagnosticou que o maior R^2 , de 0,918, foi obtido para a combinação das variáveis CHF_NALF e ESC_PREC.

O acréscimo de qualquer das variáveis restantes resultou em um aumento não significativo do R^2 , portanto não foram acrescentadas no modelo. O quadro abaixo apresenta os resultados da regressão.

Multiple Regression Results					
Multiple Regression Results					
Dep. Var. : DES_HUMA	Multiple R :	.95829241	F =	1950.756	
	R ² :	.91832435	df =	2, 347	
No. of cases: 350	adjusted R ² :	.91785360	p =	0.000000	
	Standard error of estimate:	4.795449010			
Intercept: 85.924812475	Std.Error:	.5065038	t(347) =	169.64	p < 0.0000
CHF_NALF beta=-.38 ESC_PREC beta=-.63					

Tabela 4.7 - Resultados da regressão múltipla para as variáveis CHF_NALF e ESC_PREC.

Regression Summary for Dependent Variable: DES_HUMA						
R= .95829241 R ² = .91832435 Adjusted R ² = .91785360						
F(2,347)=1950.8 p<0.0000 Std.Error of estimate: 4.7954						
	BETA	St. Err. of BETA	B	St. Err. of B	t(347)	p-level
Intercept			85,925	0,507	169,643	0,000
CHF_NALF	-0,384	0,025	-0,950	0,062	-15,443	0,000
ESC_PREC	-0,626	0,025	-1,260	0,050	-25,192	0,000

Tabela 4.8 - ANOVA da regressão múltipla para as variáveis CHF_NALF e ESC_PREC.

Analysis of Variance (trab1.sta)					
	Sums of Squares	df	Mean Square	F	p-level
Regress.	89720,47	2	44860,234	1950,756	0
Residual	7979,727	347	22,996		
Total	97700,2				

A seguir observa-se a equação de regressão para as variáveis selecionadas que melhor explicam o Índice de Autonomia de Renda.

$$DES_HUM = 85,925 - 0,95* CH_NALF - 1,26* ESC_PREC$$

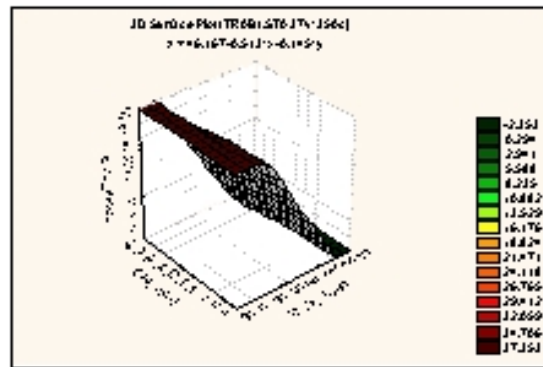


Figura 4.13 - Superfície de regressão plana simples entre Y, Ch_Nalf e Esc_Prec.

A superfície de regressão plana simples (**Figura 4.13**), correspondente ao modelo final da função de regressão, indica a não existência de interação entre as variáveis independentes e uma relação inversamente proporcional entre as mesmas e a variável dependente.

V. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.14**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta, porém a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.15**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.16**.

Na **Figura 4.16**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno.

Já nas **Figuras 4.17 e 4.18** observa-se a distribuição normal dos erros, e a independência das variáveis independentes nas **Figuras 4.19 e 4.20**.

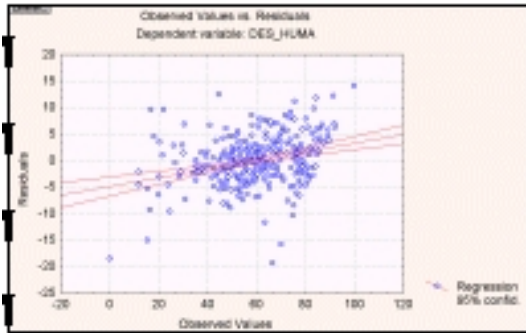


Figura 4.14 - Resíduos x Observados.

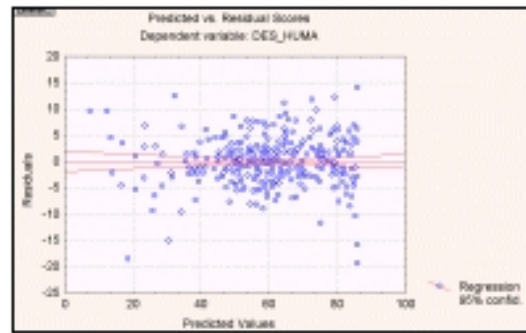


Figura 4.15 - Resíduos x Estimados (\hat{Y}).

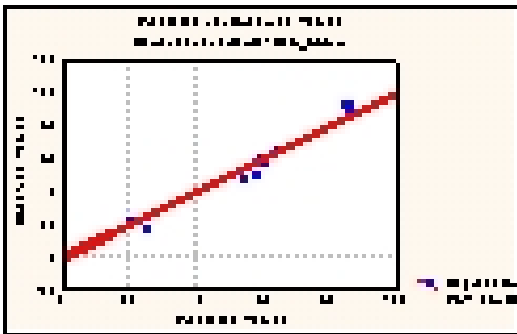


Figura 4.16 - Observados (Y_i) x Estimados (\hat{Y}).

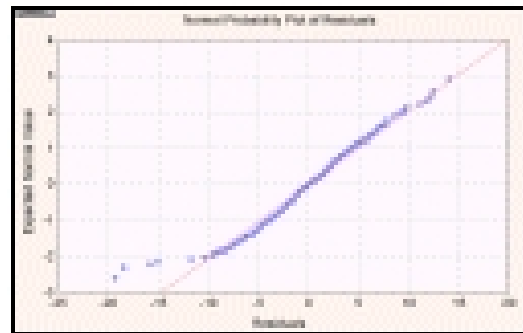


Figura 4.17 - Distribuição normal dos resíduos.

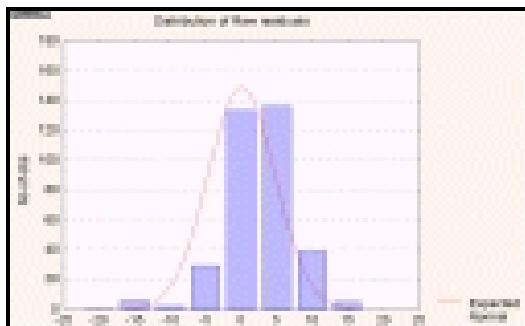


Figura 4.18 - Histograma de distribuição normal dos resíduos.

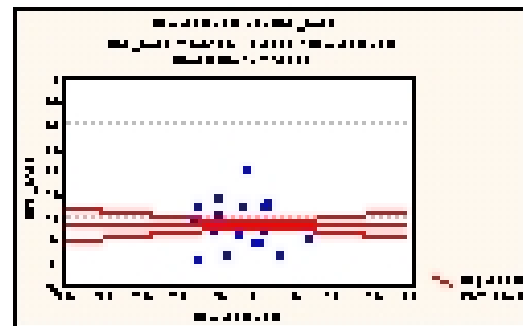


Figura 4.19 - Resíduos x CHF_NALF.

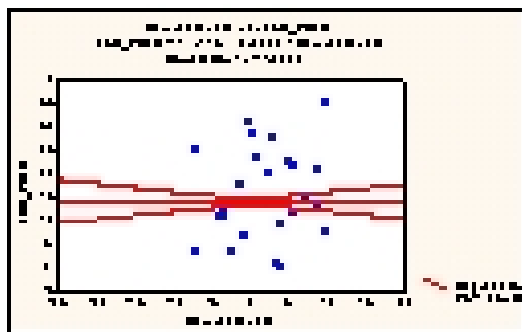
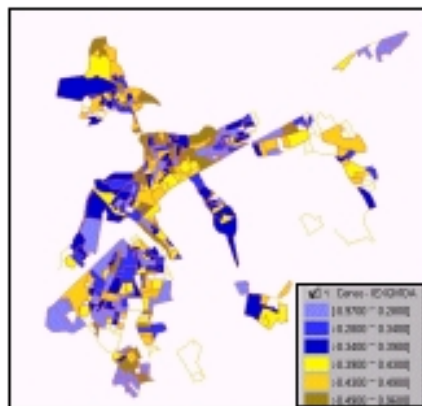


Figura 4.20 - Resíduos x ESC_PREC.

4.1.3 – REDUÇÃO DAS VARIÁVEIS INTRA-QUALIDADE DE VIDA

I. Estatística Descritiva:



Descriptive Statistics (trab1.sta)				QUALIDADE DE VIDA			
	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Minimum	Maximum	Std.Dev.
QUAL_VID	350,000	82,529	80,861	84,197	0,000	100,000	15,865

Figura 4.21 – Espacialização da variável dependente Autonomia de Renda em conjunto com sua estatística descritiva.

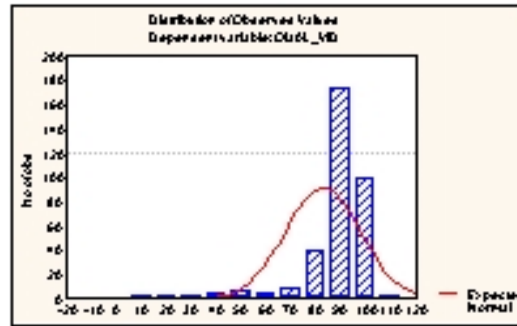


Figura 4.22 - Curva comparativa entre o comportamento do índice de autonomia de renda e a correspondente curva gaussiana de distribuição normal.

II. MATRIZ DE CORRELAÇÃO

A Matriz de Correlação foi aplicada sobre as 10 variáveis iniciais componentes do Índice Autonomia de Renda esquematizado abaixo, sendo que para a variável Condições de Privacidade considerou-se as três variáveis brutas que a constituem, a saber cômodos*, dormitórios* e banheiros* por pessoa por domicílio:

CENSO IBGE (1991)	lexi Precário Abastecimento de Água lexi Precário Instalação sanitária (Esgoto) lexi Precário Tratamento do Lixo	lex Qualidade Ambiental	lex Qualidade Domiciliar	lex QUALIDADE DE VIDA
	lexi Propriedade Domiciliar			
	lexi Densidade Habitacional	Conforto Domiciliar		
	lexi Condições de Privacidade*			
	lexi Conforto Sanitário			
	lexi Habitação Precária			

Analisando-se a Matriz de Correlação, verifica-se a baixa correlação das variáveis independentes Propriedade Domiciliar, Densidade Habitacional e Habitação Precária com a variável dependente Qualidade de Vida. As variáveis Dormitórios por Pessoa por Domicílio, Cômodos por Pessoa por Domicílio e Banheiros por Pessoa por Domicílio apresentaram alta correlação entre si e com a variável dependente Qualidade de Vida.

Portanto, em função da análise da matriz de correlação foram excluídas as variáveis PROP_DOM, DENS_HAB, DORM_P_D, BANH_P_D E e HAB_PREC, destacadas na **Tabela 4.9**, por apresentarem baixa correlação com o *Índice de Qualidade de Vida* ou grande autocorrelação entre si. não sendo boas estimadoras da variável dependente.

Tabela 4.9 - Matriz de Correlação

Correlations (trab1.sta)											
	PREC_A_A	PREC_I_S	PREC_T_L	DENS_HAB	COM_P_DM	DORM_P_D	BAN_P_DO	PROP_DOM	HAB_PREC	CONF_SAN	QUAL_VID
PREC_A_A	1,000	0,519	0,645	-0,107	-0,160	-0,144	-0,086	-0,072	0,016	-0,501	-0,789
PREC_I_S	0,519	1,000	0,588	-0,181	-0,305	-0,205	-0,253	0,104	0,045	-0,860	-0,833
PREC_T_L	0,645	0,588	1,000	-0,091	-0,181	-0,144	-0,167	-0,190	0,054	-0,552	-0,828
DENS_HAB	-0,107	-0,181	-0,091	1,000	0,122	-0,191	-0,015	-0,557	-0,040	-0,228	0,074
COM_P_DM	-0,160	-0,305	-0,181	0,122	1,000	0,687	0,861	-0,029	0,120	0,344	0,373
DORM_P_D	-0,144	-0,205	-0,144	-0,191	0,687	1,000	0,611	0,181	0,115	0,347	0,332
BAN_P_DO	-0,086	-0,253	-0,167	-0,015	0,861	0,611	1,000	0,052	0,101	0,327	0,334
PROP_DOM	-0,072	0,104	-0,190	-0,557	-0,029	0,181	0,052	1,000	-0,122	0,091	0,227
HAB_PREC	0,016	0,045	0,054	-0,040	0,120	0,115	0,101	-0,122	1,000	0,000	-0,116
CONF_SAN	-0,501	-0,860	-0,552	-0,228	0,344	0,347	0,327	0,091	0,000	1,000	0,793
QUAL_VID	-0,789	-0,833	-0,828	0,074	0,373	0,332	0,334	0,227	-0,116	0,793	1,000



Excluídos pela Matriz de Correlação
Excluídos pelas Componentes Principais

III. COMPONENTES PRINCIPAIS.

O método de componentes principais (**Tabela 4.10**), aplicado ao Índice de Qualidade de Vida (Q_VIDA) analisou a última componente na qual o autovalor aproxima-se de zero ($\lambda_5= 0,138$) objetivando a eliminação de variáveis.

Igualando-se a zero a combinação linear das variáveis independentes originais correspondentes a esta componente principal e eliminando-se as variáveis com coeficientes próximos de zero ($< 0,1$ em módulo), chegou-se à seguinte expressão:

$$0,138 \approx 0 = 0,266* \text{PREC_I_S} + 0,259*\text{CONF_SAN}$$

Sendo assim, das cinco variáveis restantes, uma variável passou a ser combinação das outras quatro. Analisando as cinco variáveis restantes na matriz de correlação, verificou-se que a variável CONF_SAN era a mais indicada para ser retirada, devido a sua menor correlação com a variável dependente Q_VIDA.

$$\text{CONF_SAN} = -0,266/0,259 * \text{PREC_I_S}$$

Assim pode-se afirmar que CONF_SAN é uma combinação linear das outras variáveis e pode ser excluída.

Tabela 4.10 - Autovetores e Autovalores dos Componentes Principais.

Factor Loadings (Unrotated) (trab1.sta)					
Extraction: Principal components					
(Marked loadings are > .700000)					
	Factor	Factor	Factor	Factor	Factor
	1	2	3	4	5
PREC_A_A	-0,757	0,347	-0,404	-0,379	0,004
PREC_I_S	-0,891	-0,050	0,365	-0,023	0,266
PREC_T_L	-0,801	0,300	-0,259	0,448	-0,021
COM_P_DM	0,427	0,841	0,330	-0,014	-0,013
CONF_SAN	0,881	0,112	-0,373	0,065	0,259
Expl.Var	2,968	0,934	0,611	0,349	0,138
Prp.Totl	0,594	0,187	0,122	0,070	0,028

IV. REGRESSÃO LINEAR

A realização de regressões lineares a partir de todas as combinações possíveis entre as variáveis diagnosticou que o maior R^2 , de 0,928, foi obtido para a combinação das variáveis PREC_A_A, PREC_I_S e PREC_T_L, retirada de uma variável reduz o R^2 para 0,86.

O acréscimo de qualquer das variáveis restantes resultou em um aumento não significativo do R^2 , portanto não foram acrescentadas no modelo. O quadro abaixo apresenta os resultados da regressão.

Multiple Regression Results						
Multiple Regression Results						
Dep. Var. :	QUAL_VID	Multiple R :	.96316140	F =	1479.428	
		R ² :	.92767988	df =	3, 346	
No. of cases:	350	adjusted R ² :	.92705283	p =	0.000000	
		Standard error of estimate:	4.284823308			
Intercept:	89.279237901	Std.Error:	.2557055	t(346) =	349.15	p < 0.0000
PREC_A_A beta=-.33 PREC_I_S beta=-.46 PREC_T_L beta=-.35						

Tabela 4.11 - Resultados da regressão múltipla para as variáveis PREC_A_A, PREC_I_S e PREC_T_L.

Regression Summary for Dependent Variable: QUAL_VID						
R= .96316140 R ² = .92767988 Adjusted R ² = .92705283						
F(3,346)=1479.4 p<0.0000 Std.Error of estimate: 4.2848						
	BETA	St. Err. of BETA	B	St. Err. of B	t(346)	p-level
Intercept			89,279	0,256	349,149	0,000
PREC_A_A	-0,327	0,019	-0,218	0,013	-16,836	0,000
PREC_I_S	-0,460	0,018	-0,194	0,008	-25,065	0,000
PREC_T_L	-0,346	0,021	-0,265	0,016	-16,849	0,000

Tabela 4.12 - ANOVA da regressão múltipla para as variáveis CHF_NALF e ESC_PREC.

Analysis of Variance (trab1.sta)					
	Sums of Squares	df	Mean Square	F	p-level
Regress.	81485,617	3,000	27161,873	1479,428	0,000
Residual	6352,460	346,000	18,360		
Total	87838,078				

A seguir observa-se a equação de regressão para as variáveis selecionadas que melhor explicam o Índice de Qualidade de Vida.

$$Q_VIDA = 89,279 - 0,218* PREC_A_A - 0,194*PREC_I_S - 0,265* PREC_T_L$$

V. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.23**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta, porém a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.24**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.25**.

Na **Figura 4.25**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno.

Já nas **Figuras 4.26 e 4.27** observa-se a distribuição normal dos erros, e a independência das variáveis independentes nas **Figuras 4.28, 4.29 e 4.30**.

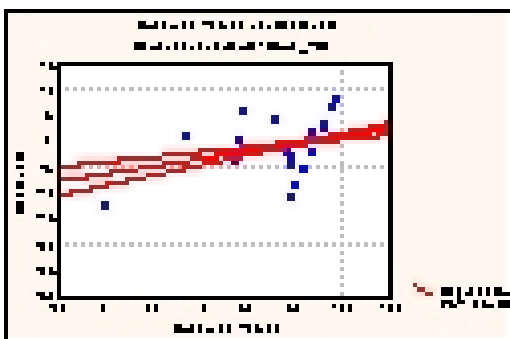


Figura 4.23 - Resíduos x Observados.

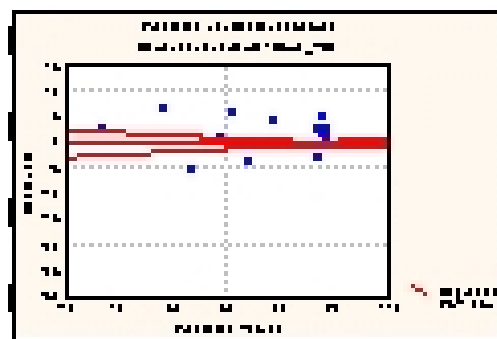


Figura 4.24 - Resíduos x Estimados (\hat{Y}).

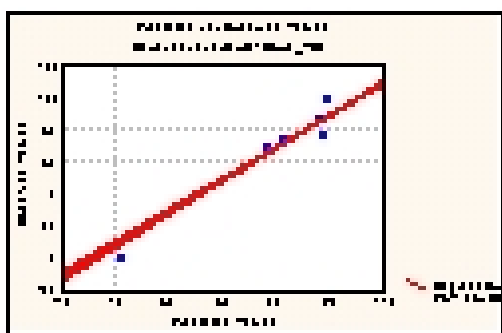


Figura 4.25 - Observados (Y_i) x Estimados (\hat{Y}).

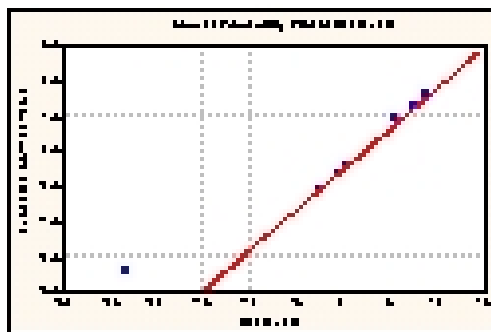


Figura 4.26 - Distribuição normal dos resíduos.

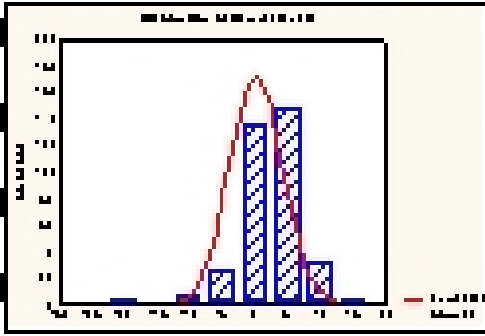


Figura 4.27 - Histograma de distribuição normal dos resíduos.

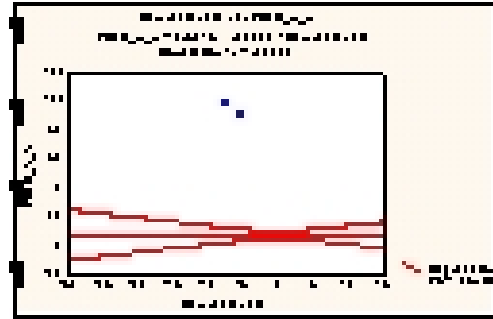


Figura 4.28 - Resíduos x PREC-A_A.

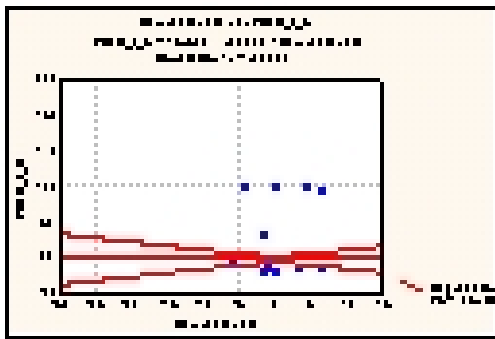


Figura 4.29 - Resíduos x PREC-I_S.

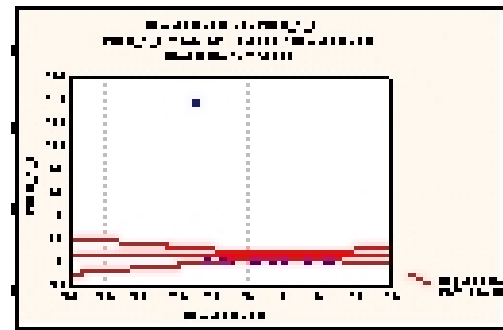
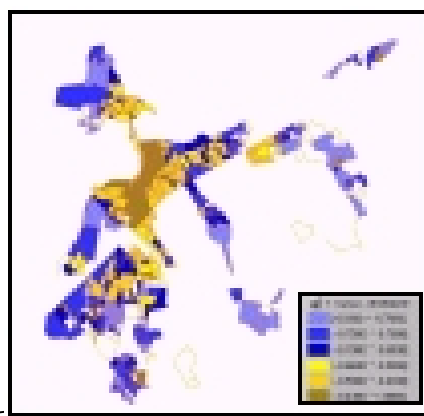


Figura 4.30 - Resíduos x PREC-T_L.

4.1.4 - REDUÇÃO DAS VARIÁVEIS INTRA-EQUIDADE

I. Estatística Descritiva:



Descriptive Statistics (trab1.sta)			EQUIDADE				
	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Minimum	Maximum	Std.Dev.
EQUIDADE	350,000	41,645	40,549	42,740	0,000	100,000	10,423

Figura 4.31 – Espacialização da variável dependente Equidade em conjunto com sua estatística descritiva.

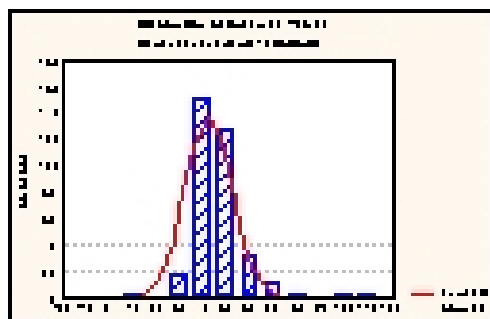


Figura 4.32 - Curva comparativa entre o comportamento do Índice de Equidade a correspondente curva gaussiana de distribuição normal.

II. MATRIZ DE CORRELAÇÃO

A Matriz de Correlação foi aplicada sobre as 2 variáveis iniciais componentes do Índice de Equidade esquematizado abaixo:

CENSO IBGE (1991)	lexi Mulheres não Alfabetizadas lexi Concentração de Mulheres Chefes de Família	lex EQUIDADE
------------------------------	--	---------------------

Analisando-se a Matriz de Correlação, verifica-se a baixa correlação da variável independente Mulheres não Alfabetizadas (M_NALF) com a variável dependente Equidade, portanto, em função desta análise a mesma foi excluída como em destaque na **Tabela 4.13**.

Tabela 4.13 - Matriz de Correlação

Correlations (trab1.sta)			
	M_NALF	M_CHF	EQUIDADE
M_NALF	1,000	0,205	0,100
M_CHF	0,205	1,000	0,904
EQUIDADE	0,100	0,904	1,000
	Excluído pela Matriz de Correlação		

III. REGRESSÃO LINEAR

Para regressão realizada verificou-se que o maior R^2 foi de 0,825 obtido para a combinação das variáveis M_CHF e M_NALF.

No entanto a regressão com apenas a variável M_CH apresentou pouca diferença em relação a anterior, tendo um R^2 de 0,817, o que resultou na exclusão da variável M_NALF. O quadro abaixo apresenta os resultados da regressão.

Multiple Regression Results						
Multiple Regression Results						
Dep. Var. : EQUIDADE	Multiple R :	.90391493	F =	1554.286		
	R ² :	.81706221	df =	1, 348		
No. of cases: 350	adjusted R ² :	.81653652	p =	0.000000		
	Standard error of estimate:	4.464276097				
Intercept:	31.292867284	Std.Error:	.3548062	t(348) =	88.197	p < 0.0000
M_CHF beta=.904						

Tabela 4.14 - Resultados da regressão múltipla para a variável M_CH

Regression Summary for Dependent Variable: EQUIDADE						
R= .90391493 R ² = .81706221 Adjusted R ² = .81653652						
F(1,348)=1554.3 p<0.0000 Std.Error of estimate: 4.4643						
	BETA	St. Err. of BETA	B	St. Err. of B	t(348)	p-level
Intercpt			31,293	0,355	88,197	0,000
M_CHF	0,904	0,023	0,457	0,012	39,424	0,000

Tabela 4.15 - ANOVA da regressão múltipla para as variável M_CH

Analysis of Variance (trab1.sta)					
	Sums of Squares	df	Mean Square	F	p-level
Regress.	30976,549	1,000	30976,549	1554,286	0,000
Residual	6935,557	348,000	19,930		
Total	37912,105				

A seguir observa-se a equação de regressão para as variáveis seleccionadas que melhor explicam o Índice de Equidade.

$$\text{EQUIDADE} = 31,293 + 0,457 * \text{M_CHF}$$

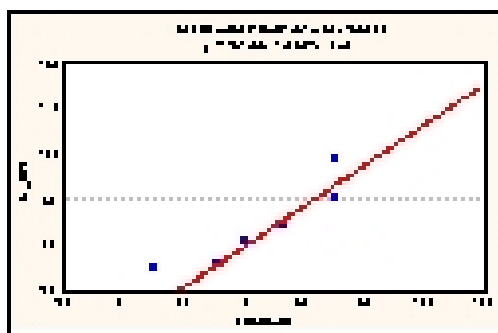


Figura 4.33 - Reta de Regressão com os valores estimados da variável mulheres chefes de família em relação a equidade.

IV. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.34**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta, porém a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.35**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.36**.

Na **Figura 4.36**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno.

Já nas **Figuras 4.37 e 4.38** observa-se a distribuição normal dos erros, e a independência das variáveis independentes nas **Figuras 4.39**.

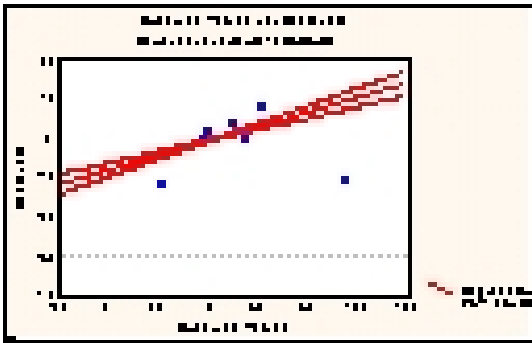


Figura 4.34 - Resíduos x Observados.

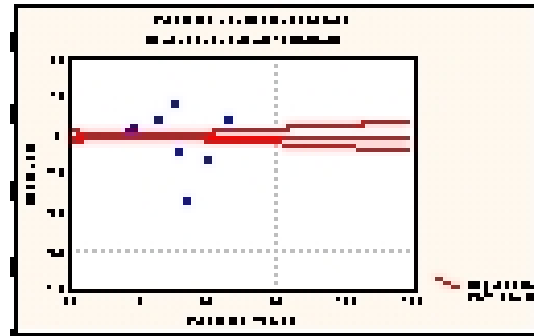


Figura 4.35 - Resíduos x Estimados (\hat{Y}).

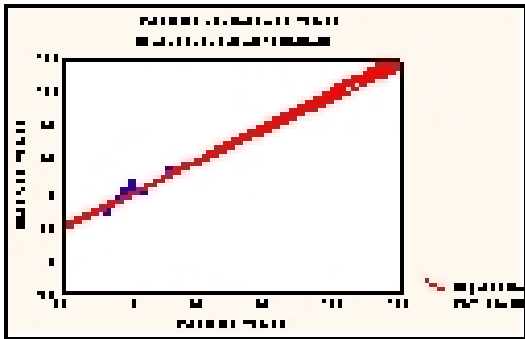


Figura 4.36 - Observados (Y_i) x Estimados (\hat{Y}).

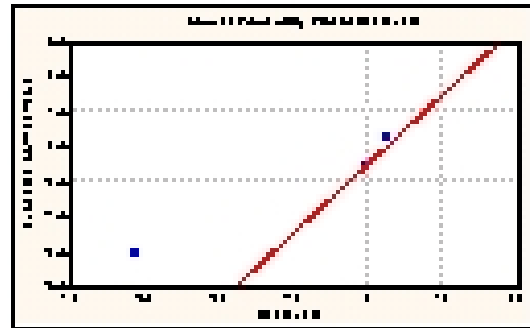


Figura 4.37 - Distribuição normal dos resíduos.

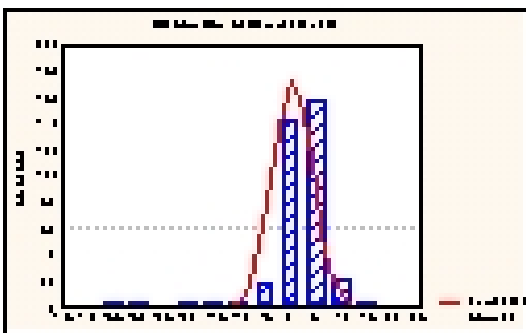


Figura 4.38 - Histograma de distribuição normal dos resíduos.

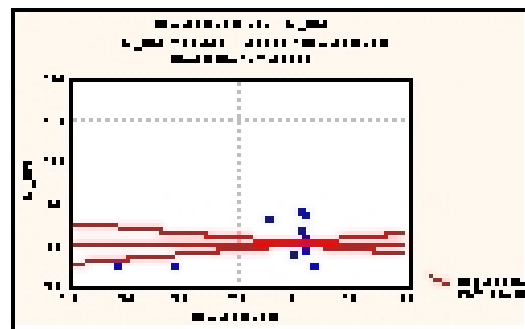


Figura 4.39 - Resíduos x M_CHF .

4.1.5 – ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTRA-UTÓPICO

A partir das Regressões Lineares realizadas para as quatro Utopias (Índices Intermediários) componentes do Índice de Exclusão Social Original inicialmente calculado pela metodologia de Sposati (1996), chegamos a quatro equações de regressão que possibilitaram o cálculo de novos valores para os mesmos a partir da substituição dos bs (variáveis independentes) pelos valores originais. As quatro equações obtidas pelas regressões apresentam-se relacionadas abaixo:

- $AUT_REN_INTRAUTOP = 89,229 - 1,3142 * A_LIN_PO - 1,033 * LIN_POBR$
- $DES_HUM_INTRAUTOP = 85,925 - 0,95 * CH_NALF - 1,26 * ESC_PREC$
- $Q_VIDA_INTRAUTOP = 89,279 - 0,218 * PREC_A_A - 0,194 * PREC_I_S - 0,265 * PREC_T_L$
- $EQUIDADE_INTRAUTOP = 31,293 + 0,457 * M_CHF$

O novo Índice de Exclusão/Inclusão Intrautópico foi obtido através da soma linear dos novos valores das Utopias (N_AUT_RE, N_DES_HU, N_Q_VIDA e N_EQUID) e posteriormente escalonado de -1 à 1

4.2 - REDUÇÃO INTER-UTÓPICA

4.2.1 - ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTER-UTÓPICO

O Índice de Exclusão/Inclusão Social Interutópico foi construído semelhantemente ao Índice de Exclusão/Inclusão Intrautópico, diferenciando-se por uma continuação da análise de regressão a partir das Novas Utopias reduzidas, visando caracterizar em ordem crescente

de importância a contribuição estatística das variáveis para a composição do Índice de Exclusão/Inclusão Social Original. Deste modo as análises de matriz de correlação, componentes principais e regressões lineares foram aplicadas tendo por variável dependente o Índice Original e por variáveis independentes as quatro novas Utopias resultantes da redução Intrautópica.

Os resultados das análises geradas para este novo Índice de Exclusão/Inclusão Social denominado Interutópico serão apresentadas abaixo seguindo o padrão dos itens acima.

I. Estatística Descritiva:

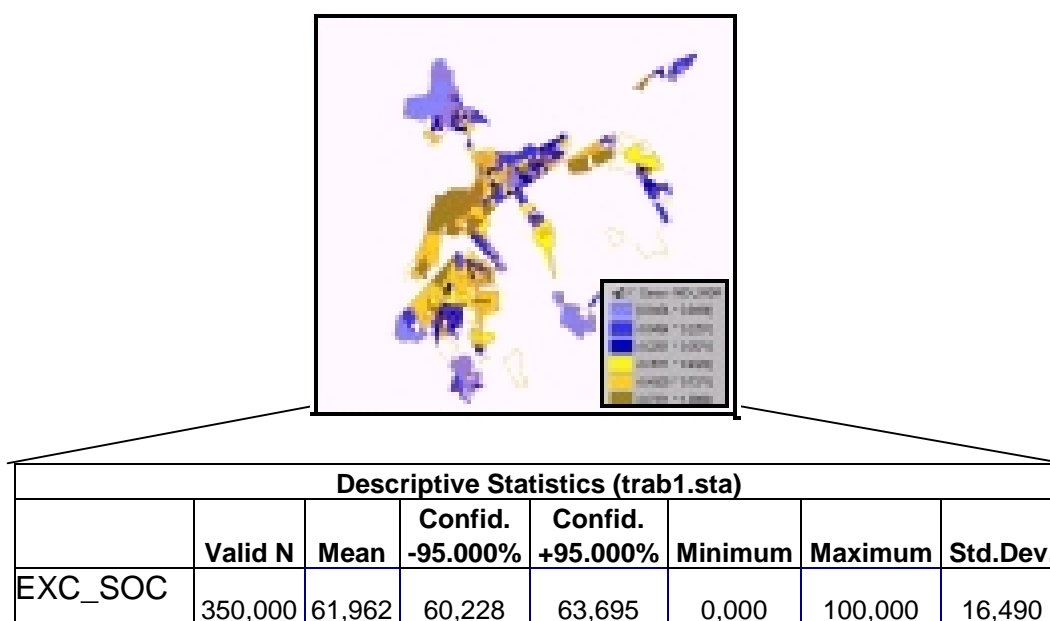


Figura 4.40 – Espacialização da variável dependente Exclusão/Inclusão Social Indireto em conjunto com sua estatística descritiva.

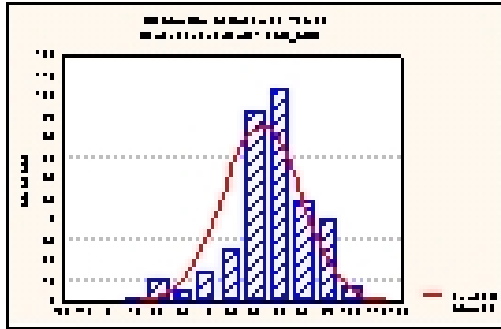


Figura 5.41 - Curva comparativa entre o comportamento do índice de Exclusão/Inclusão Social Indireto e a correspondente curva gaussiana de distribuição normal.

II. MATRIZ DE CORRELAÇÃO

A Matriz de Correlação foi aplicada sobre os 4 Novos Índices Intermediários componentes do Índice Exclusão/Inclusão Social citados abaixo:

- I. Índice de Autonomia de Renda Intrautópico;
- II. Índice de Desenvolvimento Humano Intrautópico;
- III. Índice de Qualidade de Vida Intrautópico;
- IV. Índice de Equidade Intrautópico.

Analisando-se a Matriz de Correlação, verifica-se a baixa correlação da variável independente N_EQUIDADE com a variável dependente Exclusão Social/Inclusão Independente. Portanto, em função da análise da matriz de correlação a mesma foi excluída tornando-se um mal estimador para variável dependente.

Tabela 4.15 - Matriz de Correlação

	N_AUT_RE	N_DES_HU	N_QU_VID	N_EQUIDA	EXC_SOC
N_AUT_RE	1,000	0,855	0,438	0,264	0,827
N_DES_HU	0,855	1,000	0,589	0,412	0,928
N_QU_VID	0,438	0,589	1,000	0,290	0,722
N_EQUIDA	0,264	0,412	0,290	1,000	0,560
EXC_SOC	0,827	0,928	0,722	0,560	1,000



Excluídos pela Matriz de Correlação
Excluídos pela Componentes Principais

III. COMPONENTES PRINCIPAIS.

O método de componentes principais (**Tabela 4.16**), aplicado às quatro novas utopias em relação ao índice original, analisou a última componente na qual o autovalor aproxima-se de zero ($\lambda_6 = 0,125$) objetivando a eliminação de variáveis.

A análise parte da última componente, a qual o autovalor aproxima-se de zero ($\lambda_6 = 0,125$). Igualando-se a zero a combinação linear das variáveis independentes originais correspondentes a esta componente principal e eliminando-se as variáveis com coeficientes próximos de zero ($< 0,1$ em módulo), chegou-se à seguinte expressão:

$$0 \approx -0.226.N_AUT_RE + 0.265.N_DES_HU$$

Sendo assim, das 4 variáveis restantes, uma variável passou a ser combinação das outras três. Analisando as 4 variáveis restantes na matriz de correlação, verificou-se que a variável N_AUT_RE era a mais indicada para ser retirada, devido a sua menor correlação com a variável dependente Exclusão Social Indireta.

$$0.226.N_AUT_RE = 0.265.N_DES_HU$$

Assim pode-se afirmar que N_AUT_RE é função de N_DES_HU e pode ser excluída do modelo.

Tabela 4.16 - Autovetores e Autovalores dos Componentes Principais.

Factor Loadings (Unrotated) (trab1.sta) Extraction: Principal components (Marked loadings are > .700000)			
	Factor 1	Factor 2	Factor 3
N_AUT_RE	-0,897	0,379	-0,226
N_DES_HU	-0,951	0,162	0,265
N_QU_VID	-0,750	-0,658	-0,065
Expl.Var	2,272	0,603	0,125
Prp.Totl	0,757	0,201	0,042

IV. REGRESSÃO LINEAR

A realização de regressões lineares a partir das combinações possíveis com as duas variáveis restantes, observa-se que com apenas a variável N_DES_HU no modelo, o R^2 já era de 0.8608. Com a inclusão da variável N_QU_VID no modelo, o R^2 aumenta para 0.908. Ou seja, um acréscimo de apenas 0,0472 no R^2 . O acréscimo de qualquer das variáveis restantes resultou em um aumento não significativo do R^2 , tornando desnecessária acrescentá-las no modelo. O quadro abaixo apresenta os resultados da regressão evidenciando o pequeno acréscimo no R^2 .

Tabela 4.17 - Resultados da regressão múltipla para a variável N_DES_HU

Regression Summary for Dependent Variable: EXC_SOC R= .92777938 R²= .86077457 Adjusted R²= .86037450 F(1,348)=2151.5 p<0.0000 Std.Error of estimate: 6.1617						
	BETA	St. Err. of BETA	B	St. Err. of B	t(348)	p-level
N_DES_HU	0,928	0,020	0,954	0,021	46,385	0,000
Intercpt			4,666	1,278	3,650	0,000

Tabela 4.18 - ANOVA da regressão múltipla para a variável N_DES_HU

Analysis of Variance (trab1.sta)					
	Sums of Squares	df	Mean Square	F	p-level
Regress.	81686,383	1,000	81686,383	2151,543	0,000
Residual	13212,312	348,000	37,966		
Total	94898,695				

A seguir observa-se a equação de regressão para a variável selecionada que melhor explicou o Índice de Exclusão/Inclusão Social Indiretamente.

$$\text{EXCL_INTERUTOP} = 4.666 + 0.954.N_DES_HUM$$

V. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.42**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta e a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.43**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.44**.

Na **Figura 4.44**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno. Nas **Figuras 4.45 e 4.46** observa-se a distribuição normal dos erros, e a independência das variáveis independentes na **Figura 4.47**.

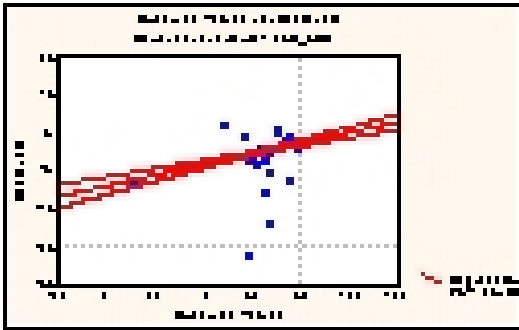


Figura 4.42 - Resíduos x Observados.

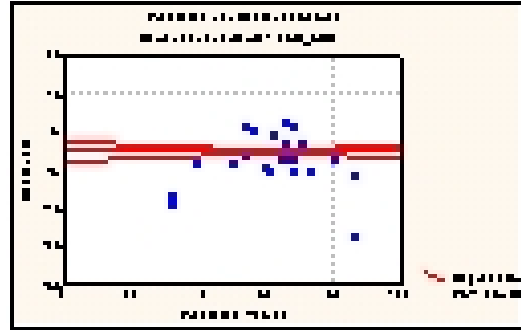


Figura 4.43 - Resíduos x Estimados (\hat{Y}).

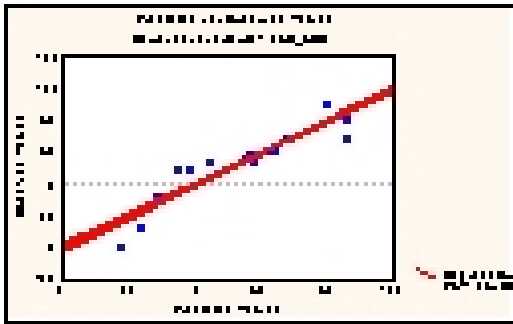


Figura 4.44 - Observados (Y_i) x Estimados (\hat{Y}).

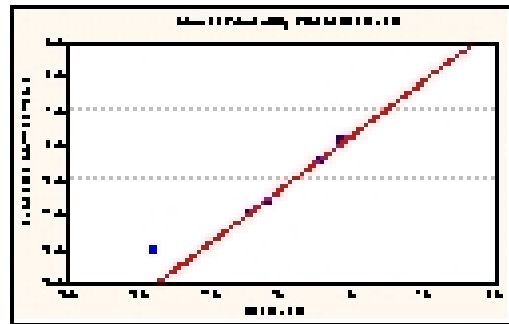


Figura 4.45 - Distribuição normal dos resíduos.

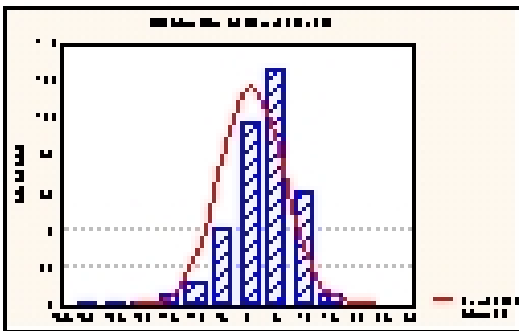


Figura 4.46 - Histograma de distribuição normal dos resíduos.

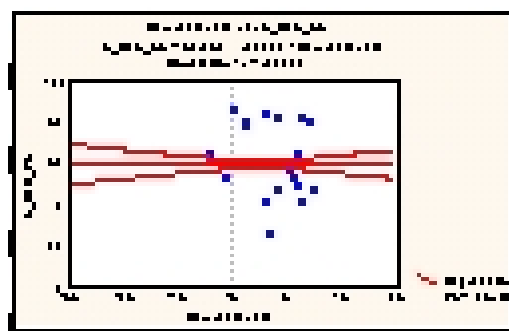


Figura4.47 - Resíduos x N_DES_HU

4.3 - REDUÇÃO INTER-VARIÁVEIS

4.3.1 - ÍNDICE DE EXCLUSÃO/INCLUSÃO SOCIAL MÍNIMO INTER-VARIÁVEIS

O último método desenvolvido busca avaliar os efeitos de uma desvinculação total do universo da quatro Utopias em. relação à Exclusão/Inclusão Social. Para tal os métodos estatísticos utilizados foram aplicados tendo por variável dependente o Índice de Exclusão/Inclusão Social Original e por variáveis independentes as 8 variáveis que restaram após a primeira redução Intrautópica.

Este novo índice de Exclusão/Inclusão Social foi chamado de Intervariáveis porque pretende entender qual(is) das variável(is) bruta(s) contribui(em), estatisticamente, com maior relevância para o fenômeno. A seguir são apresentadas as análises realizadas seguindo o padrão anteriormente proposto.

I. ESTATÍSTICA DESCRITIVA.

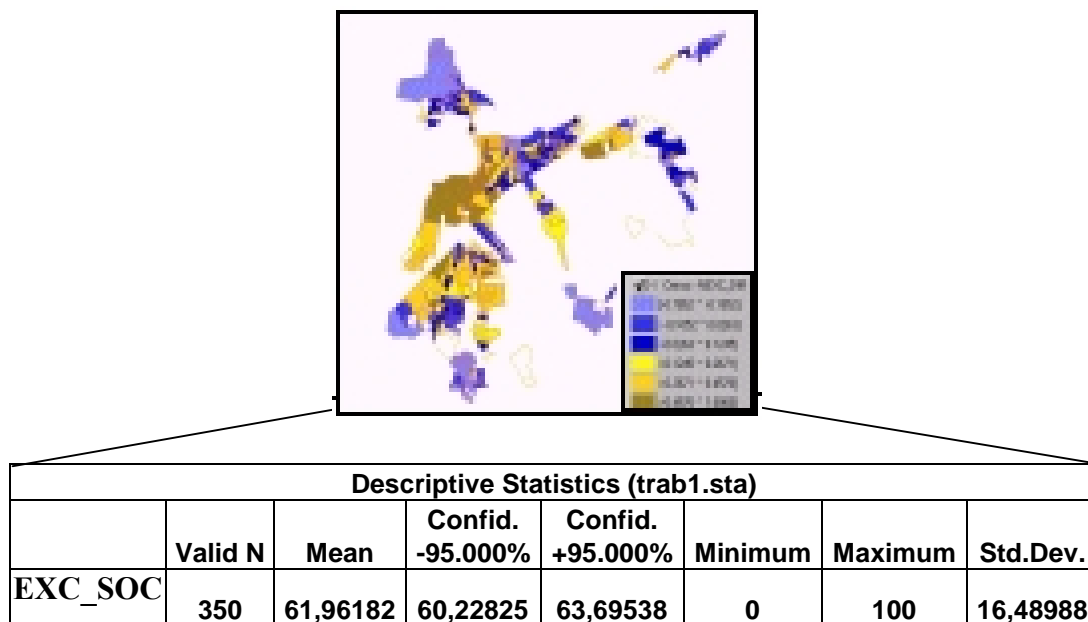


Figura 4.48 – Espacialização da variável dependente Exclusão/Inclusão Social Direta em conjunto com sua estatística descritiva.

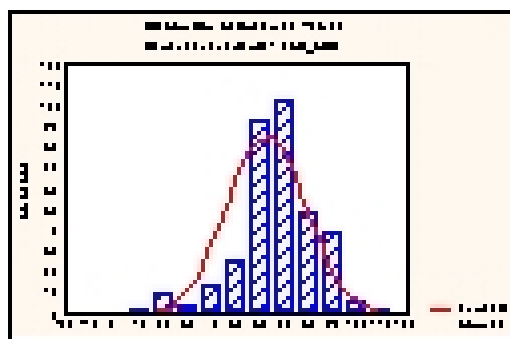


Figura 4.49 - Curva comparativa entre o comportamento do índice Exclusão/Inclusão Social Direto e a correspondente curva gaussiana de distribuição normal.

II. MATRIZ DE CORRELAÇÃO NUMÉRICA

A Matriz de Correlação foi aplicada sobre as 9 variáveis selecionadas para os Índices Intermediários (item 3.2) esquematizadas abaixo:

Tabela 4.19 - Matriz de Correlação

Correlations (trab1.sta)									
	A_LIN_PO	LIN_POB	CHF_NALF	ESC_PREC	M_CHF	PREC_A_A	PREC_I_S	PREC_T_L	EXC_SOC
A_LIN_PO	1,000	0,270	0,290	0,354	-0,182	-0,038	0,190	0,032	-0,375
LIN_POB	0,270	1,000	0,821	0,850	-0,289	0,311	0,449	0,443	-0,853
CHF_NALF	0,290	0,821	1,000	0,787	-0,365	0,449	0,580	0,613	-0,876
ESC_PREC	0,354	0,850	0,787	1,000	-0,407	0,338	0,493	0,426	-0,882
M_CHF	-0,182	-0,289	-0,365	-0,407	1,000	-0,224	-0,285	-0,217	0,560
PREC_A_A	-0,038	0,311	0,449	0,338	-0,224	1,000	0,519	0,645	-0,532
PREC_I_S	0,190	0,449	0,580	0,493	-0,285	0,519	1,000	0,588	-0,673
PREC_T_L	0,032	0,443	0,613	0,426	-0,217	0,645	0,588	1,000	-0,614
EXC_SOC	-0,375	-0,853	-0,876	-0,882	0,560	-0,532	-0,673	-0,614	1,000



Excluídos pela Matriz de Correlação

Excluídos pela componentes Principais

Pela análise da matriz de correlação pode-se excluir a variável A_LIN_PO, a qual apresenta baixa correlação com a exclusão social.

III. COMPONENTES PRINCIPAIS.

O método de componentes principais (**Tabela 4.20**), aplicado às variáveis reduzidas em relação ao índice original, analisou a última componente na qual o autovalor aproxima-se de zero ($\lambda_5= 0,126$) objetivando a eliminação de variáveis.

Igualando-se a zero a combinação linear das variáveis independentes originais correspondentes a esta componente principal e eliminando-se as variáveis com coeficientes próximos de zero ($< 0,1$ em módulo), chegou-se à seguinte expressão:

$$0 \approx -0.268.LIN_POB + 0.144.CHF_NALF + 0.176.ESC_PREC$$

Sendo assim, das 8 variáveis restantes, uma variável passou a ser combinação das outras 7. Analisando as 7 variáveis restantes na matriz de correlação, verificou-se que LIN_POB era a mais indicada para ser retirada, devido a sua menor correlação com a variável dependente N_EX_DIR.

$$0.268.LIN_POB \approx 0.144.CHF_NALF + 0.176.ESC_PREC$$

Pode-se afirmar que LIN_POB é uma combinação linear de CHF_NALF e de ESC_PREC e que portanto pode ser retirada do modelo.

Tabela 4.21 - Autovetores e Autovalores dos Componentes Principais.

Factor Loadings (Unrotated) (trab1.sta) Extraction: Principal components (Marked loadings are > .700000)							
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
LIN_POB	0,827	-0,402	-0,270	-0,085	0,045	0,007	-0,268
CHF_NALF	0,906	-0,176	-0,147	-0,037	-0,106	0,307	0,144
ESC_PREC	0,843	-0,407	-0,107	-0,032	0,132	-0,250	0,176
M_CHF	-0,480	0,264	-0,832	0,039	0,063	-0,011	0,038
PREC_A_A	0,646	0,598	0,097	-0,308	0,345	0,032	-0,007
PREC_I_S	0,746	0,299	0,040	0,585	0,100	-0,002	-0,030
PREC_T_L	0,751	0,469	-0,061	-0,116	-0,426	-0,129	-0,016
Expl.Var	3,984	1,095	0,814	0,461	0,345	0,175	0,126
Prp.Totl	0,569	0,156	0,116	0,066	0,049	0,025	0,018

IV. REGRESSÃO LINEAR

A realização de regressões lineares a partir de todas as combinações possíveis entre as variáveis, resultou em uma equação de regressão com três termos. O acréscimo de qualquer das variáveis restantes resultou em um aumento não significativo do R^2 , portanto não foram acrescentadas no modelo. O quadro abaixo apresenta os resultados da regressão.

Multiple Regression Results		
Dep. Var. : EXC_SOC	Multiple R : .95097977	F = 1090.610
	R ² : .90436253	df = 3, 346
No. of cases: 350	adjusted R ² : .90353330	p = 0.000000
	Standard error of estimate: 5.121605522	
Intercept: 79.679838729	Std.Error: .7614317	t(346) = 104.64 p < 0.0000
CHF_NALF beta=-.45 ESC_PREC beta=-.44 M_CHF beta=.217		

Tabela 4.22 - Resultados da regressão múltipla para as variáveis M_CHF, ESC_PREC e CHF_NALF.

Regression Summary for Dependent Variable: EXC_SOC R= .95097977 R ² = .90436253 Adjusted R ² = .90353330 F(3,346)=1090.6 p<0.0000 Std.Error of estimate: 5.1216						
	BETA	St. Err. of BETA	B	St. Err. of B	t(346)	p-level
Intercpt			79,680	0,761	104,645	0,000
CHF_NALF	-0,451	0,027	-1,101	0,066	-16,702	0,000
ESC_PREC	-0,439	0,028	-0,870	0,055	-15,930	0,000
M_CHF	0,217	0,018	0,173	0,015	11,889	0,000

Tabela 4.23 - ANOVA da regressão múltipla para as variáveis M_CHF, ESC_PREC e CHF_NALF.

Analysis of Variance (trab1.sta)					
	Sums of Squares	df	Mean Square	F	p-level
Regress.	85822,83	3	28607,61	1090,609	0
Residual	9075,872	346	26,23084		
Total	94898,7				

A seguir observa-se a equação de regressão para as variáveis selecionadas que melhor explicam o Índice de Autonomia de Renda.

$$\text{IEXCL_INTERV} = 79.68 - 1.101.\text{CHF_NALF} - 0.87.\text{ESC_PREC} + 0.173.\text{M_CH}$$

V. ANÁLISE DOS RESÍDUOS

Como pode ser observado na **Figura 4.50**, os resíduos possuem uma relação diretamente proporcional com os valores observados relativos a reta e a distribuição dos pontos ao longo da reta mostra que não há relação entre ambos. Na **Figura 4.51**, após a regressão constata-se uma ausência de relação entre os valores estimados e os resíduos. A variância constante dos erros, bem como a média igual a zero podem ser verificados na **Figura 4.52**.

Na **Figura 4.52**, constata-se que a distribuição dos valores observados é proporcional à reta de regressão, isso dá subsídio para a afirmação de que a regressão linear modela bem o fenômeno. Nas **Figuras 4.53 e 4.54** observa-se a distribuição normal dos erros, e a independência das variáveis independentes nas **Figuras 4.55, 4.56 e 4.57**.

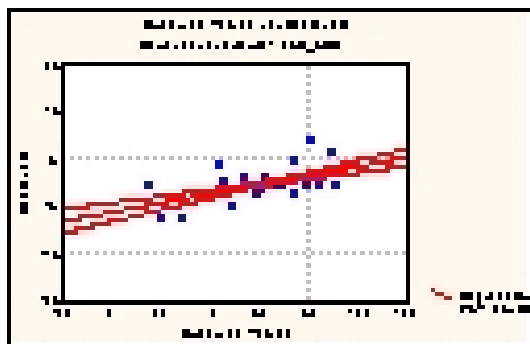


Figura 4.50 - Resíduos x Observados (\hat{Y}).

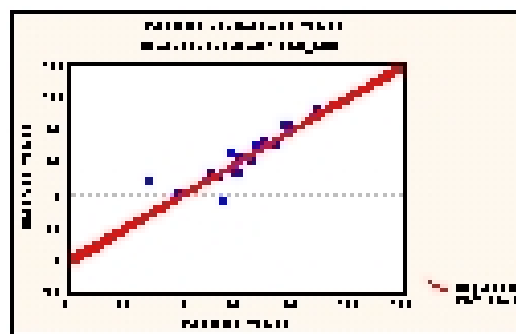


Figura 4.51 - Resíduos x Estimados.

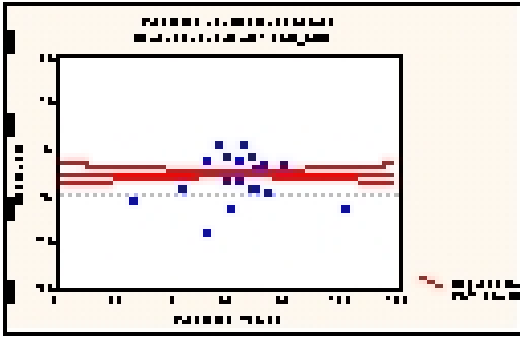


Figura 4.52 - Observados (Y_i) x Estimados (\hat{Y}).

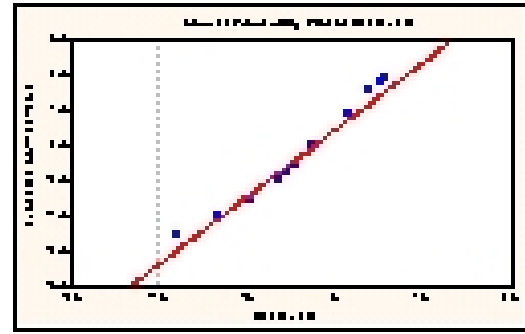


Figura 4.53 - Distribuição normal dos resíduos.

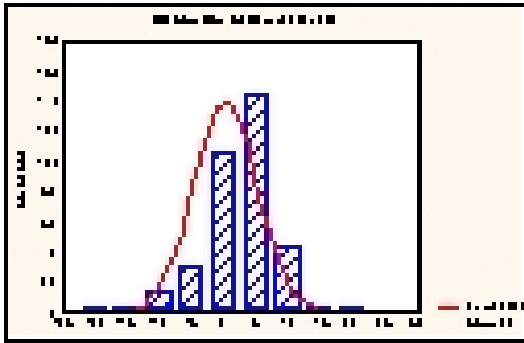


Figura 4.54 - Histograma de distribuição normal dos resíduos.

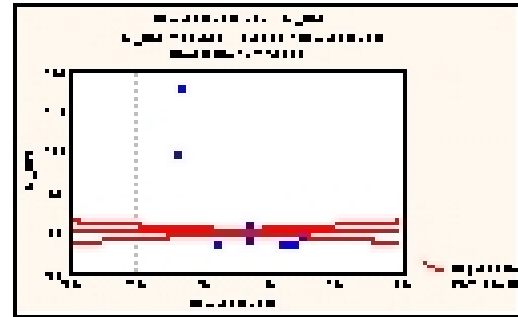


Figura 4.55 - Resíduos x M_CHF

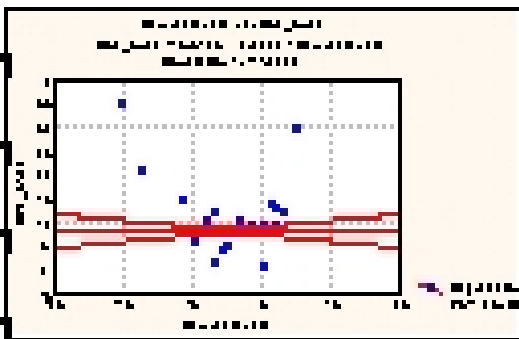


Figura 4.56 - Resíduos CHF_NALF .

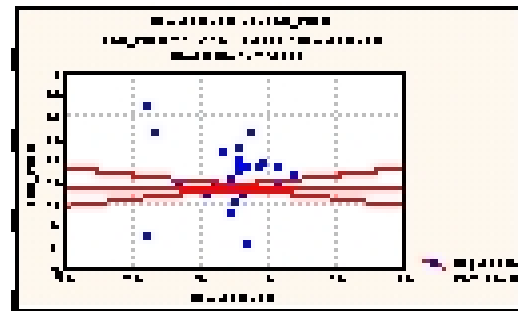
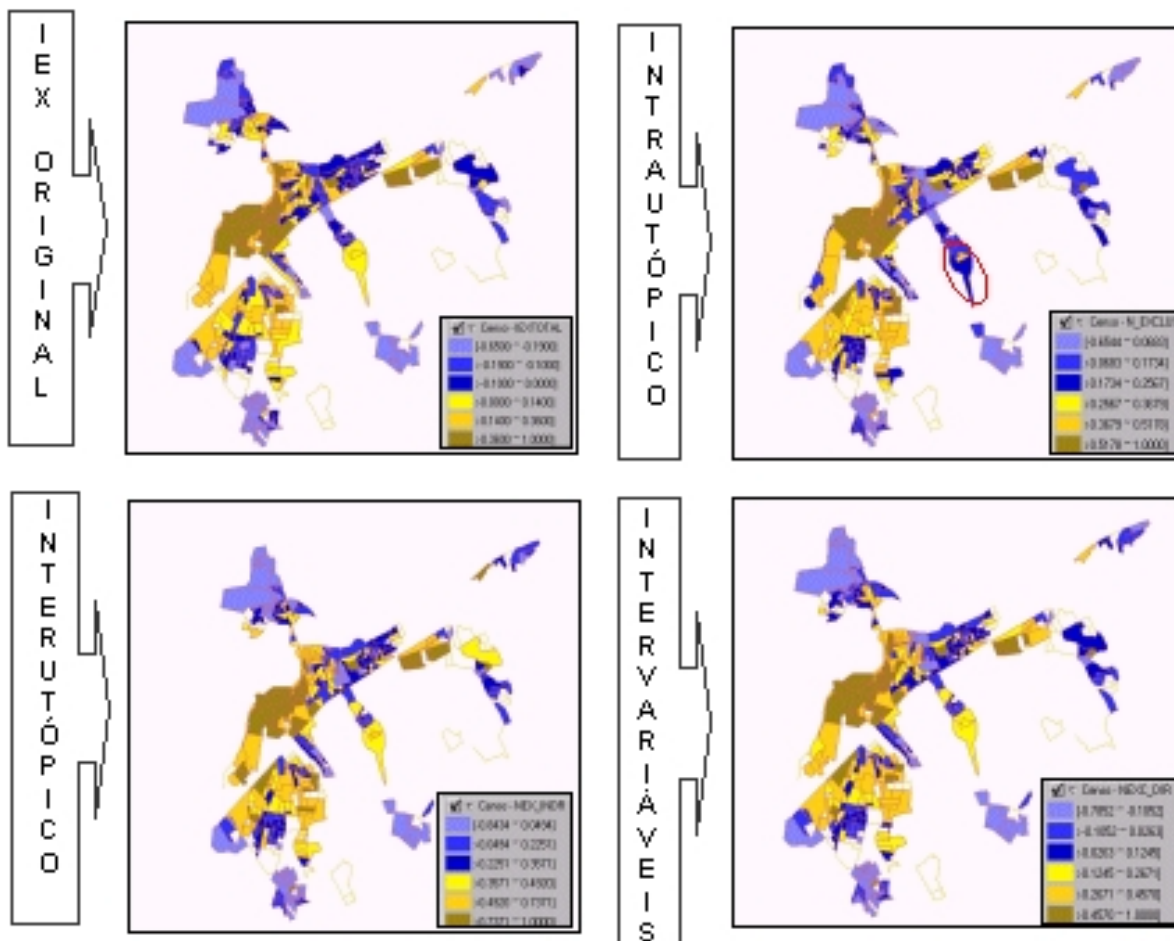


Figura 4.57 - Resíduos x ESC_PREC

4.4 – APRESENTAÇÃO DOS MAPAS MÍNIMOS



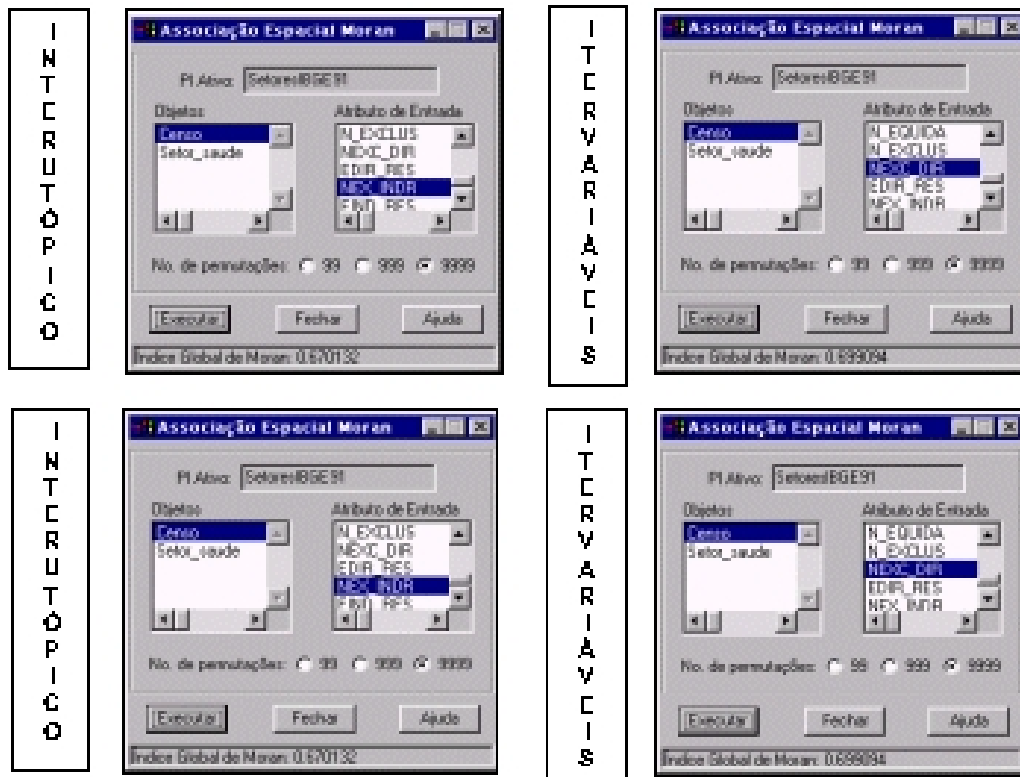
4.5 – ANÁLISE ESPACIAL DE ÁREAS

I. ÍNDICE GLOBAL DE MORAN

O Índice Global de Moran, para os três novos Índices e para o original, fornece o grau de autocorrelação espacial da variável estudada, ou seja, mede quanto o valor da mesma numa região é independente dos valores dos vizinhos. Este indicador detecta afastamentos de uma distribuição espacial aleatória.

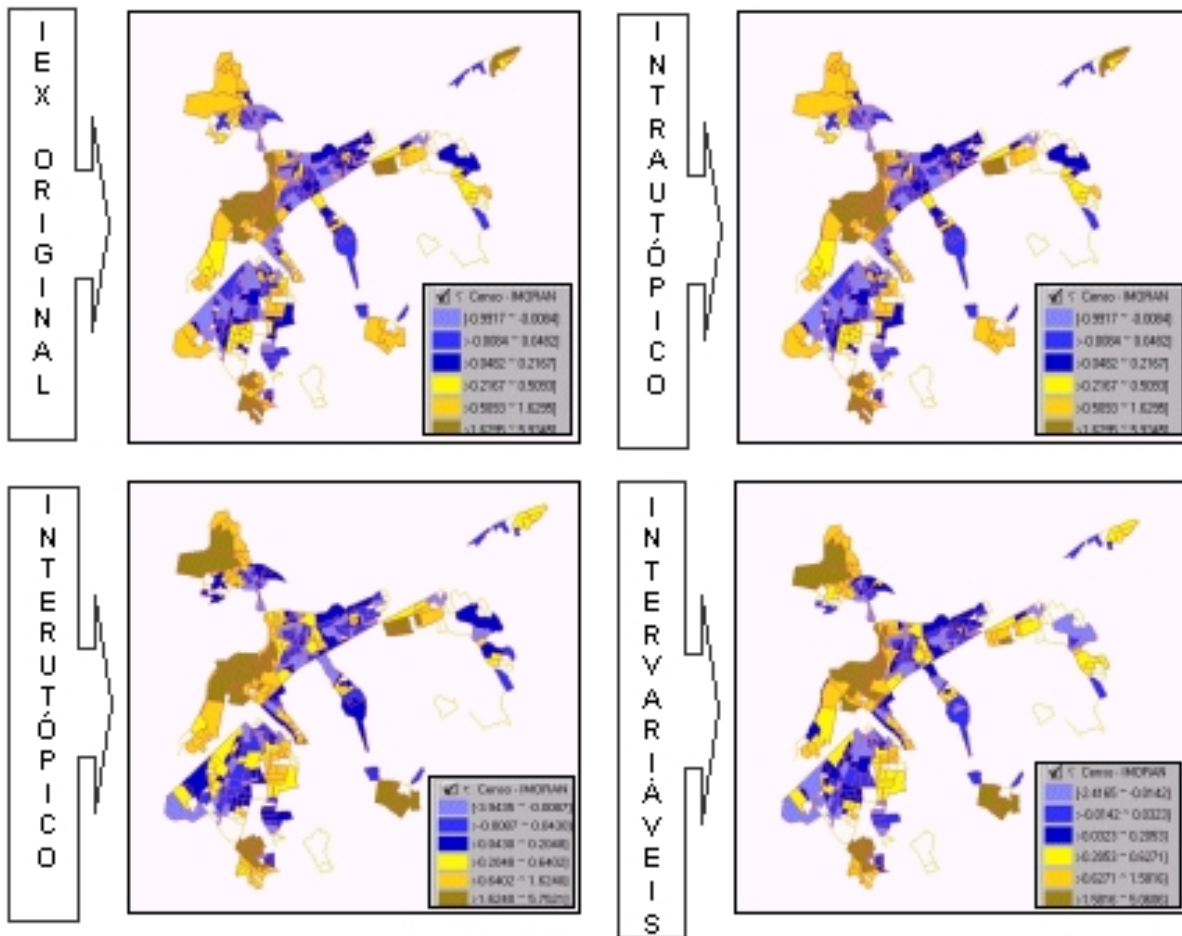
Verifica-se que os índices Globais de Moran obtidos foram 0,6835 (original), 0,6887(recalculado), 0,6990 (direto) e 0,6701 (indireto) indicando uma alta correlação

espacial positiva para a região como um todo. Tal resultado não deve ser analisado de maneira conclusiva para a região, pois pelo fato do índice ser global, este pode mascarar regimes espaciais locais não detectados pelo mesmo. Em função disto, torna-se necessário a continuação da análise exploratória dos dados através dos indicadores locais de associação espacial (LISA), particularmente o Índice Local de Moran.



I. ÍNDICE LOCAL DE MORAN

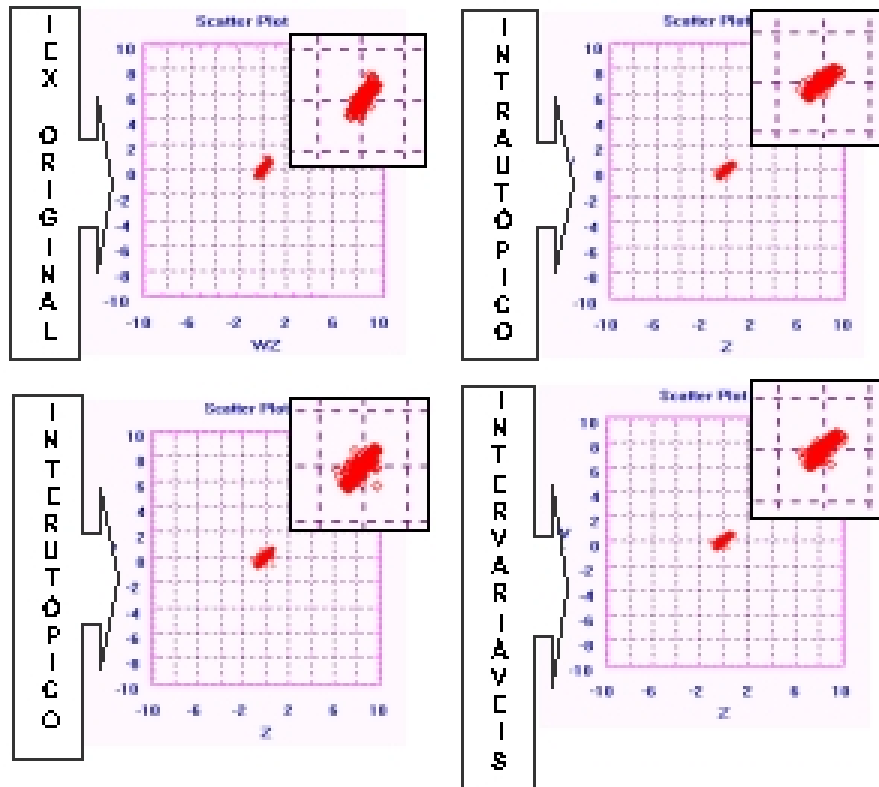
O Índice Local de Moran computa o índice local de cada polígono, pela multiplicação do desvio em relação à média global dos resíduos, pela média dos desvios dos vizinhos. Este índice permite a identificação de agrupamentos de polígonos com valores de atributos semelhantes (*clusters*) que apresentam mesmo regime espacial, assim como polígonos anômalos (*outliers*) e áreas de transição. Um aspecto importante do Índice Local de Moran diz respeito ao nível de significância associado para cada valor calculado.



II. DIAGRAMA DE ESPALHAMENTO DE MORAN

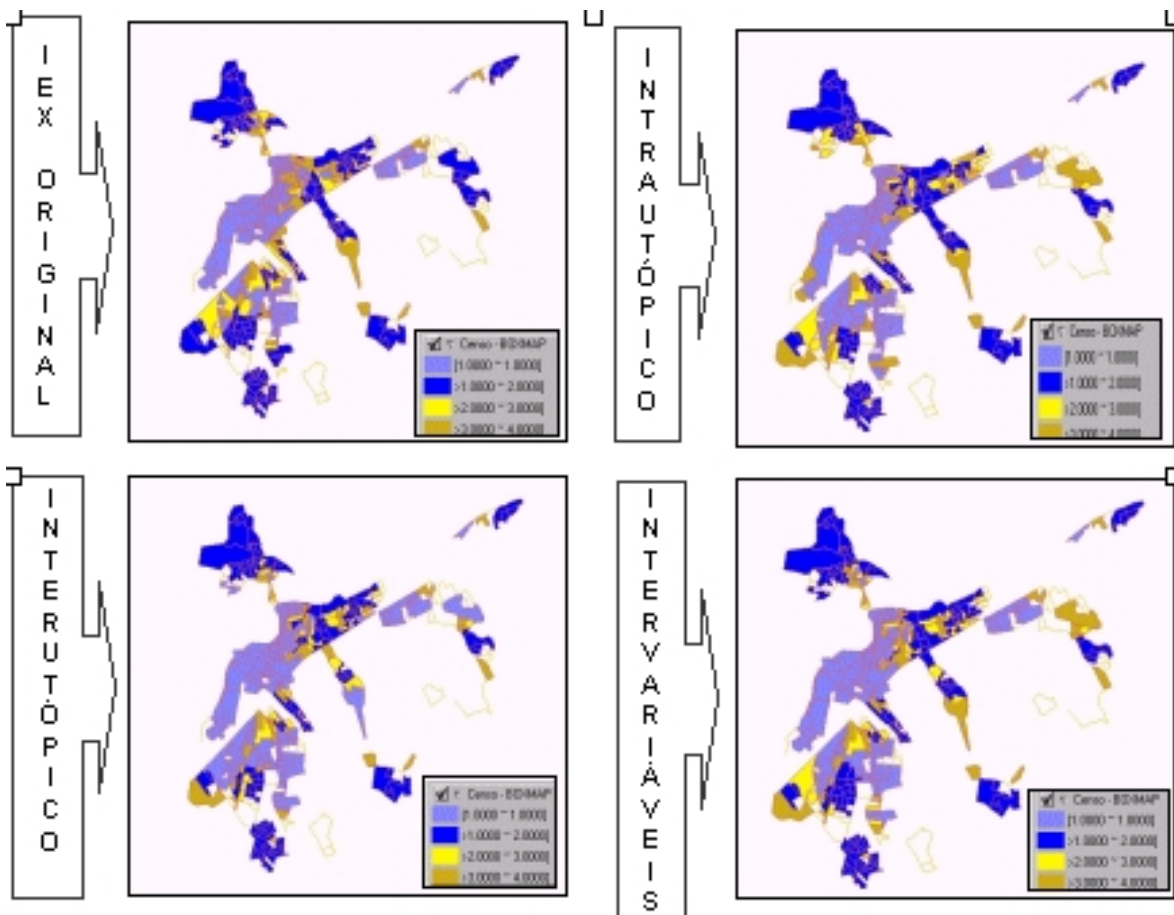
O diagrama de espalhamento de Moran é uma maneira alternativa de visualizar o Índice Global de Moran, onde se pode verificar o relacionamento entre os valores observados dos desvios, de cada polígono, em relação à média global (Z) e os valores das médias locais (WZ). O Índice Global de Moran equivale ao coeficiente de regressão linear que indica a inclinação da reta de regressão de WZ em Z . Os quadrantes do diagrama são divididos em HH (*high, high*), LL (*low, low*), HL (*high, low*) e LH (*low, high*), sendo que os dois primeiros indicam correlação espacial positiva (presença de regime espacial) e os dois últimos, correlação espacial negativa (*presença de região de transição ou outlier*).

Verifica-se abaixo menor quantidade de pontos nos quadrantes LH e HL, comparando-se com a quantidade de pontos nos quadrantes HH e LL. Tal fato explica o alto valor obtido nos Índices Globais de Moran, tendo em vista que o produto $Z \times WZ$ dos pontos pertencentes aos quadrantes HH e LL contribuem positivamente no somatório do cálculo do referido índice.



III. BOX MAP

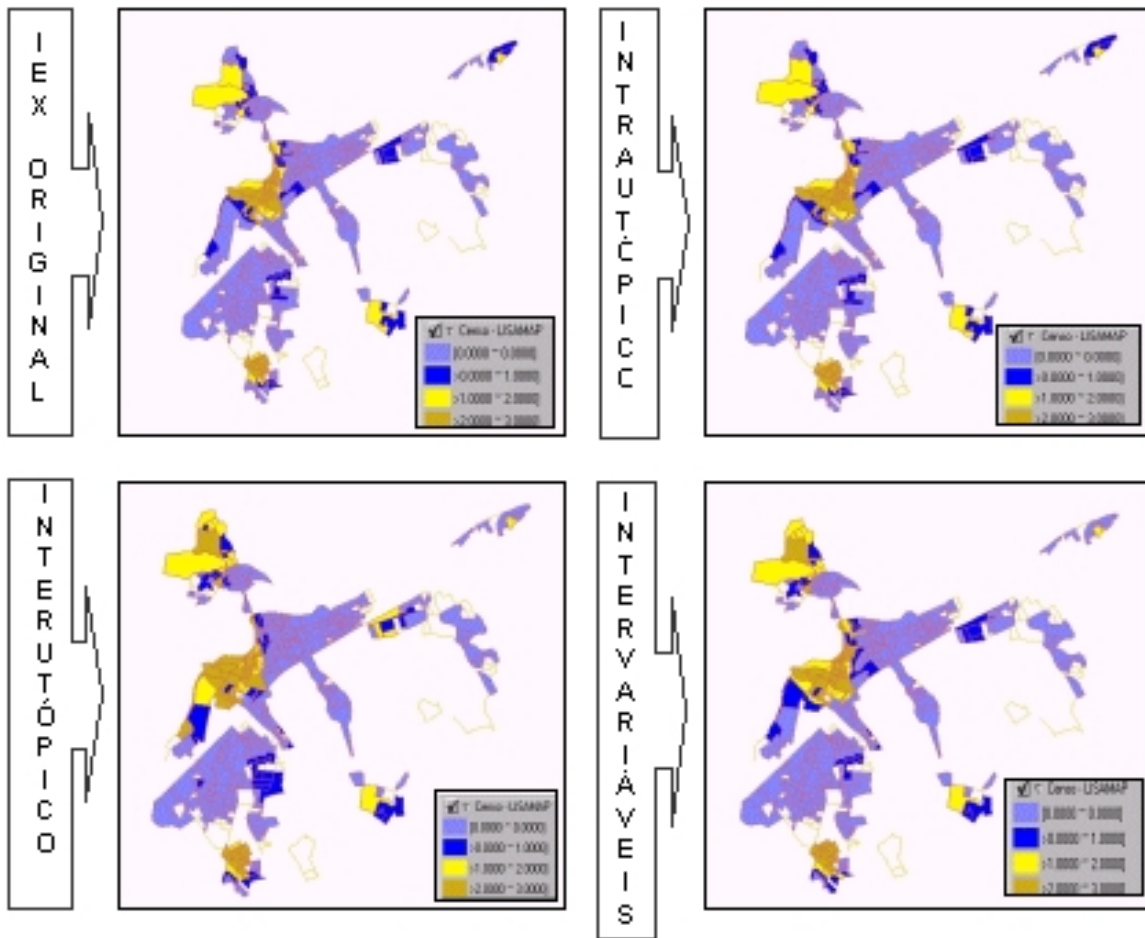
O Box Map classifica os polígonos, presentes no mapa, em 1 ou 2 (dependência espacial positiva (HH) ou negativa (LL)) ou em 3 ou 4 (áreas de transição ou presença de *out liers* (LH, HL)). Os referidos mapas são uma espacialização do Diagrama de Espalhamento de Moran.



Pelos mapas acima verifica-se a distribuição espacial da classificação dos polígonos em relação aos quadrantes do Diagrama de Espalhamento de Moran, cujos resultados apresentam continuidades, indicando a existência de grandes agrupamentos (*clusters*).

IV. LISA MAP

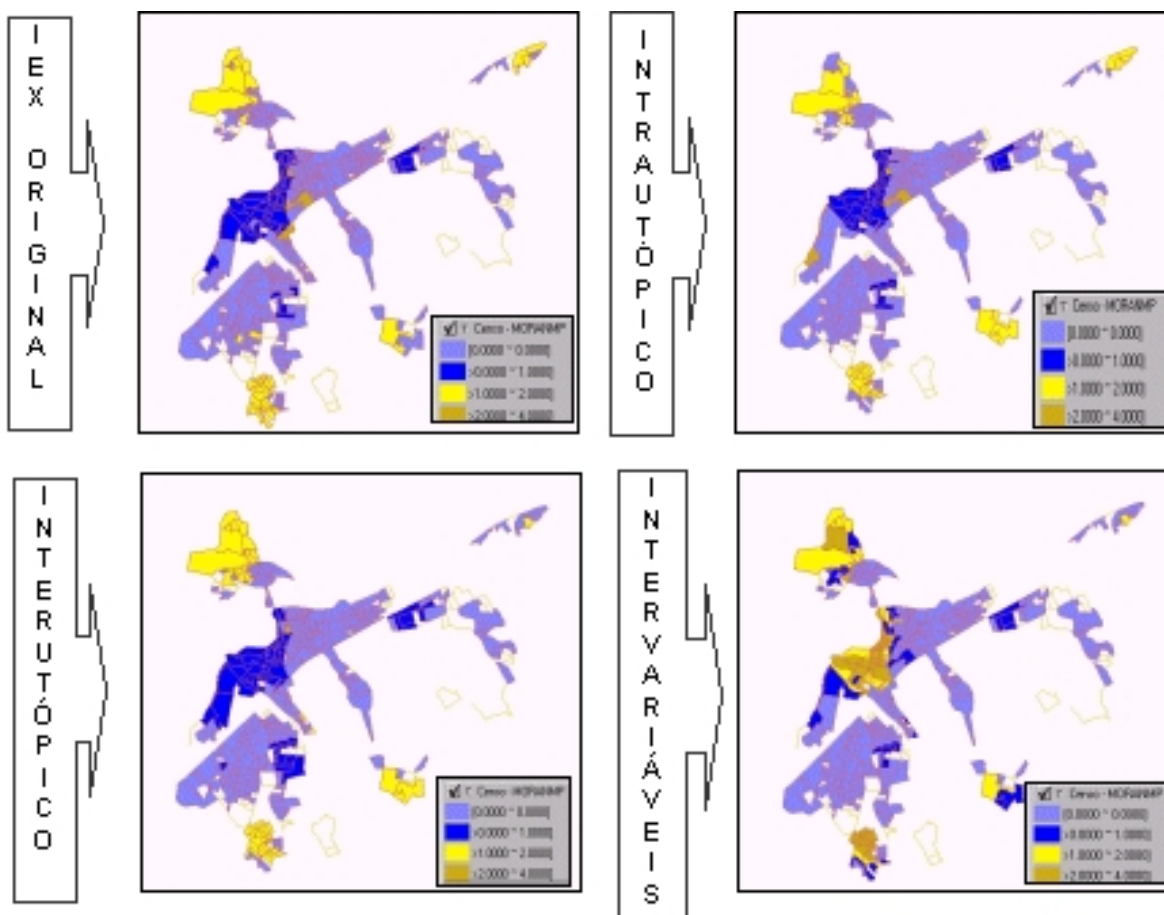
O LISA Map avalia o nível de significância dos índices locais de Moran, obtidos para cada polígono. Tal avaliação da significância é feita comparando os valores de IMoran com uma série de valores obtidos por meio de permutação dos valores dos vizinhos sob a hipótese nula de inexistência de autocorrelação. Os valores de significância, são, então classificados em quatro grupos: não significante, significância de 5%, 1% e 0,1%.



Pelos mapas acima se verifica a distribuição espacial dos valores de significância, cujos resultados apresentam predominância da classificação "não significante" dos valores do Índice Local de Moran obtidos para cada polígono. Porém os mapas acima revelam regiões específicas com agrupamentos significativos (Santana, Campo dos Alemães, Jd. Esplanada, Putim e Jd. Diamante).

V. MORAN MAP

Moran Map é um mapa síntese que combina o LISA MAP e o BOX MAP, destacando os polígonos com significância abaixo de 5% e seu respectivo quadrante.



Pelos mapas acima se verifica a distribuição espacial da classificação, para os polígonos considerados significantes (5%, 1% e 0,1%), em relação aos quadrantes do Diagrama de Espalhamento de Moran, sendo os demais "não significantes", cujos resultados mostram predominância da classificação como "não significante", em concordância com o resultado apresentado no LISA Map. Este comportamento evidencia a existência de agrupamentos significantes em locais específicos (regiões isoladas) para os novos Índices de Exclusão/Inclusão Social no município de São José dos Campos.

CAPÍTULO 5 – CONCLUSÕES

As conclusões estão divididas segundo as análises estatísticas e espaciais realizadas para cada novo índice de Exclusão/Inclusão social calculado.

O Índice de Exclusão/Inclusão Social Intrautópico diferenciou-se levemente do original, em sua espacialização, em função da variável M_CHF, a única considerada no Equidade_Intrautópico, gerando uma confusão nas análises. O número de mulheres chefes de família não é uma característica exclusiva de famílias carentes mas uma realidade da nova ordem social que se instala atualmente.

O Índice de Exclusão/Inclusão Social Intervariáveis, bem como o Intervariáveis também manteve, padrões semelhantes ao Índice Original.

O Índice Global de Moran apresentou valores muito próximos entre os três novos Índices, indicando significativa correlação espacial, ou seja manutenção dos agrupamentos para os novos índices gerados.

A forma elipsoidal do Diagrama de Espalhamento de Moran indica que há tendência de autocorrelação local para todos os Índices.

O Box Map revelou comportamentos semelhantes para todas as distribuições, HH, LL, LH e HL. O LISA Map manteve as regiões diagnosticadas como significantes para todos os índices gerados.

Entretanto a análise qualitativa a partir da espacialização dos novos índices traz apenas uma percepção do comportamento de cada uma, para compararmos quantitativamente os “Mapas Mínimos” utilizaremos como ferramenta o gráfico de barras (**Figura 5.1**) comparando os três novos Índices com relação ao Original Figura .

Comparando-se os mapas de barras abaixo, referentes as diferenças entre os três novos índices de Exclusão/Inclusão social, pode-se constatar que o Índice Interutópico apresenta menor instabilidade, sendo que o resultado menos satisfatório foi obtido pelo novo Exclusão/Inclusão.

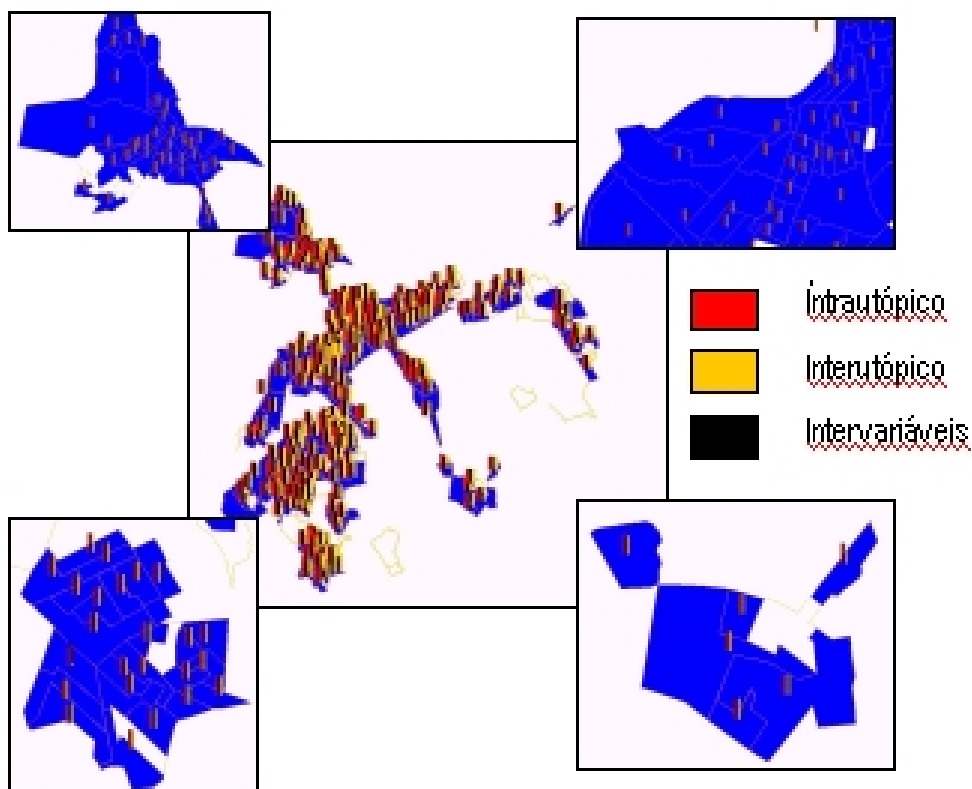
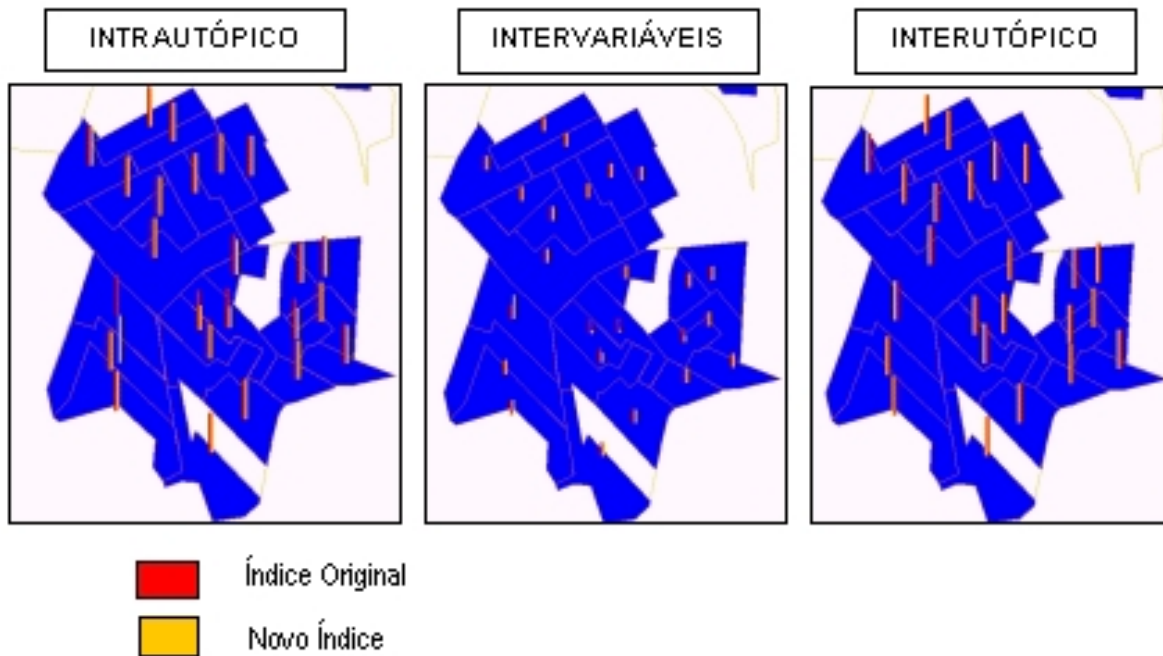
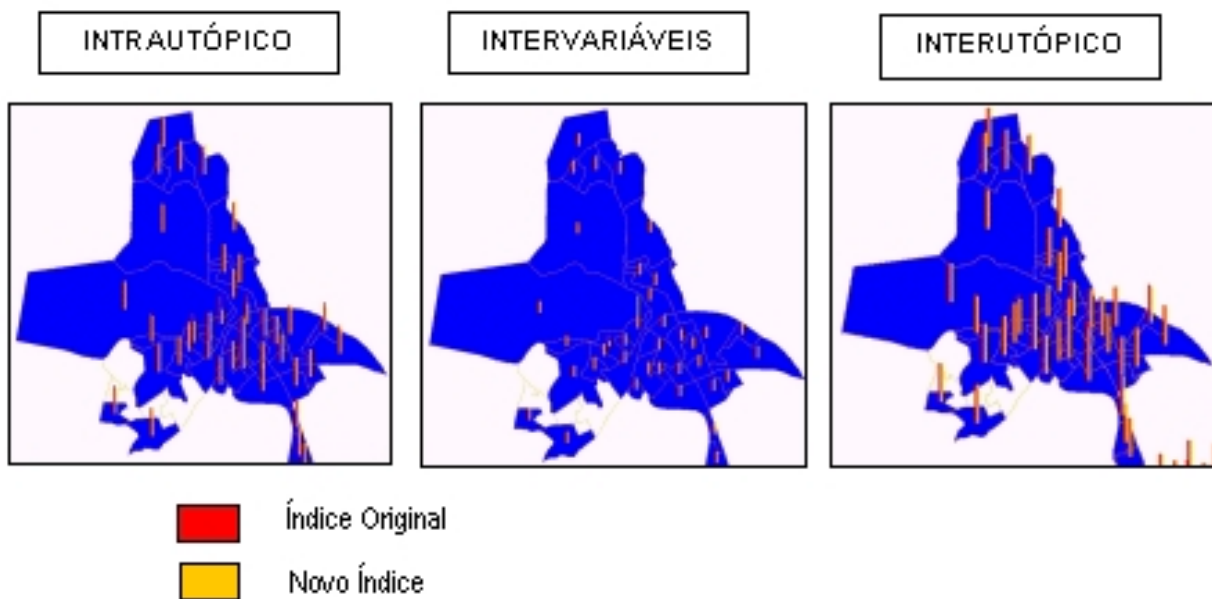


Figura 5.1 – Gráfico de Barras comparando os três Índices gerados.

Os mapas de barras abaixo apresentam uma comparação em detalhe entre os novos índices e o índice original (Sposati, 1996), para a área teste escolhida da região do Campo dos Alemães. Os resultados mostram que o índice indireto apresentou-se mais semelhante ao original.



Os mapas de barras abaixo apresentam uma comparação em detalhe entre os novos índices e o índice original (Sposati, 1996), para a área teste escolhida da região do Bairro Santana. Os resultados mostram que o índice indireto apresentou-se mais semelhante ao original.



Estatisticamente não há diferenças significativas entre os métodos Intervariáveis e Interutópico, o que é comprovado pela semelhança dos R^2 de 0,9043 e 0,8608 respectivamente.

Portanto, para São José dos Campos, de acordo com os dados de 1991 (IBGE) é possível estimar “Mapas Mínimos” para a Exclusão/Inclusão Social com um número reduzido de variáveis. Tornou-se evidente também a busca de um método mais preciso e eficaz para comparar quantitativamente os mapas gerados pelos novos métodos.

Também existem perspectivas relacionadas a abrangência do método proposto de forma que possa ser aplicado em diferentes realidades urbanas através da utilização dos dados Censo disponíveis para todo o País. O delineamento de Políticas Públicas globais e locais através destes métodos seria um experimento interessante, também a ser realizado.

REFERÊNCIAS BIBLIOGRÁFICAS:

Fisher, M. ; Scholten, H. ; Unkin, D. *Spatial Analytica Perspectives on GIS*, Taylor & Francis, Londres, Cap. 8, 1996.

Bailey, T.; Gatrell, A.C., *Interactive Spatial Data Analysis.*, Longman Group Limited, England, 1995.

Câmara, G.; Monteiro, A. M.; Carvalho, M. “*Análise Espacial e Geoprocessamento*”
Curso on-line, INPE, 2000.

Chatterjee, S.; Price, B. *Regression Analysis by Example*. John Wiley e Sons, New York, 1977.

Maxwell, S. The Meaning and Measurement of Poverty, *Poverty Briefing*, Vol 3, Fev. 1999.

Neter, J. e Wasserman W. *Applied Linear Statistical Models*. Irwin-Dorsey Limited, Georgetown, 1974

Sposati, A., *Mapa da Exclusão/Inclusão da Cidade de São Paulo*, Editora da PUC-SP, São Paulo, 1996.