



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Análise Espacial de Indicadores Sócio-Econômicos na Compreensão do Manejo da Colheita da Cana-de-Açúcar

Maikon R. de Novaes

SER-301 - Análise Espacial de Dados Geográficos

São José dos Campos
Dezembro de 2008

Ministério da
Ciência e Tecnologia





Definição do Problema

**Manejo da colheita da cana-de-açúcar
no estado de São Paulo**



Definição do Problema

Cana-crua



Definição do Problema



\$\$\$\$\$\$

Definição do Problema





Definição do Problema

Cana-queima

Definição do Problema



Danos ao meio ambiente e à saúde



Hipótese Inicial

**Indicadores sócio-econômicos podem
auxiliar na explicação no tipo de
colheita da cana-de-açúcar**



Explorando o problema...

$$I_c = \frac{(\%A) - (\%B)}{100}$$

I_c = Indicador de Colheita ($-1 \leq \mathbf{I}_c \leq 1$)

A = Quantidade de cana-crua

B = Quantidade de cana-queima

Período de análise: **2006** e **2007**

Explorando o problema...

Identificação de agrupamentos espacial para o modo de colheita:

Média Móvel Espacial:

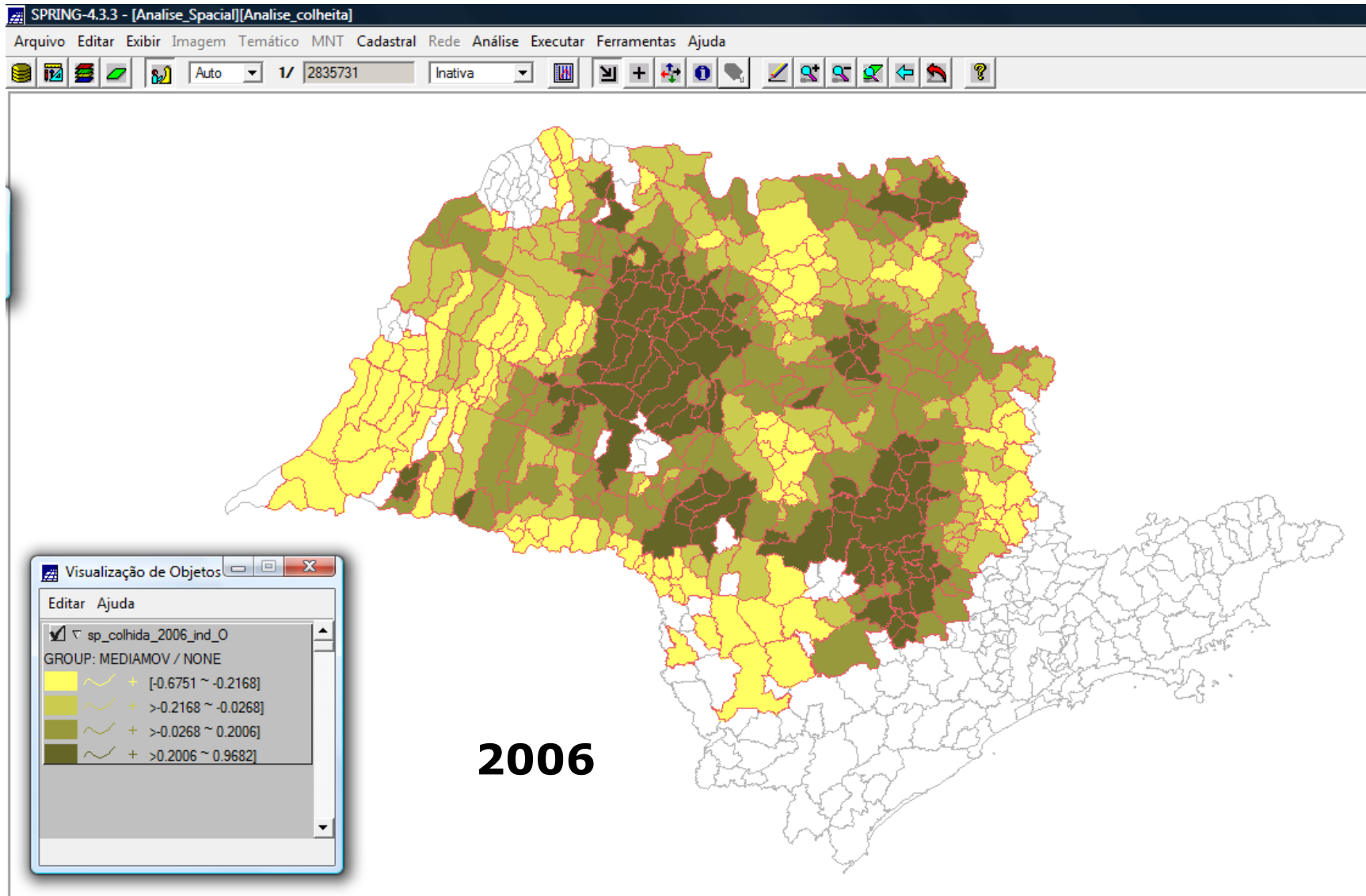
$$\hat{\mu}_i = \frac{\sum_{j=1}^n W_{ij} y_j}{\sum_{j=1}^n W_{ij}} \quad i = 1, 2, \dots, n$$

onde:

- W_{ij} é a matriz de proximidade.
- y_i é o valor do atributo em cada área.
- n é o número de polígonos

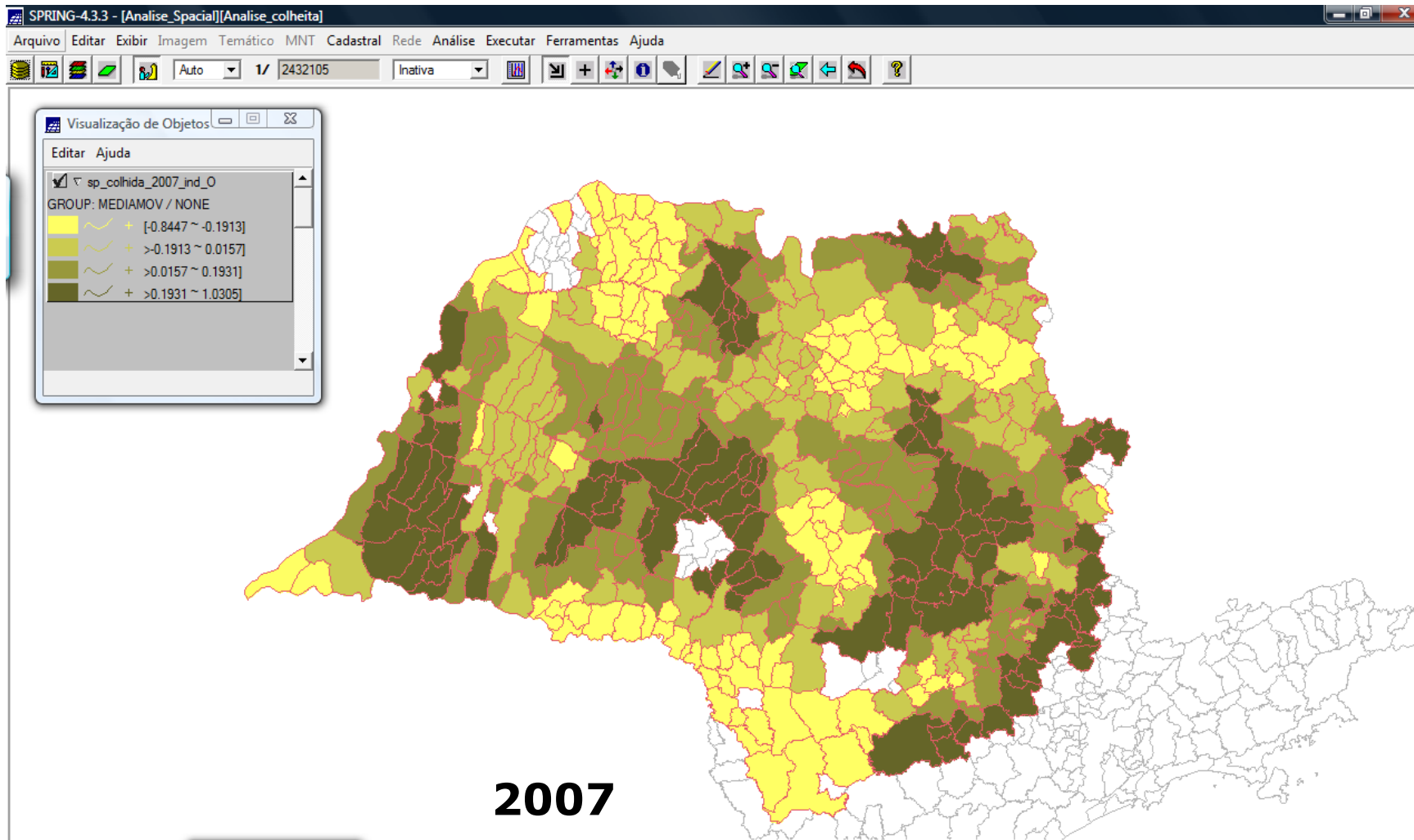


Explorando o problema...





Explorando o problema...



Explorando o problema...

Identificação de agrupamentos espacial para o modo de colheita:

Índice Global de Moran:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

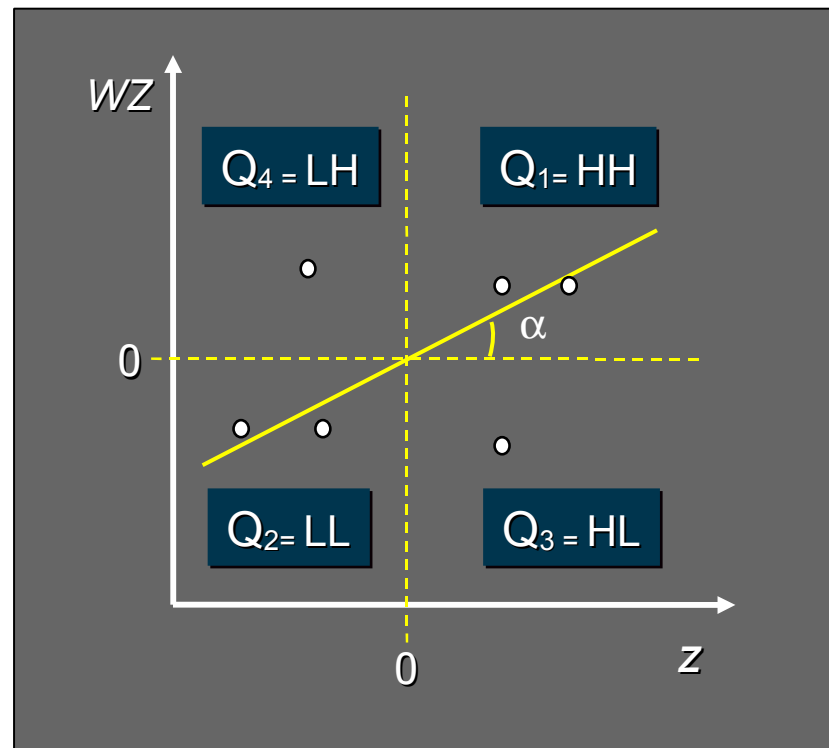
onde:

- n corresponde ao número de áreas,
- y_i é o valor do atributo considerado na área i ,
- \bar{y} representa o valor médio do atributo na região de estudo,
- w_{ij} são os pesos atribuídos conforme a conexão entre as áreas i e j .

Explorando o problema...

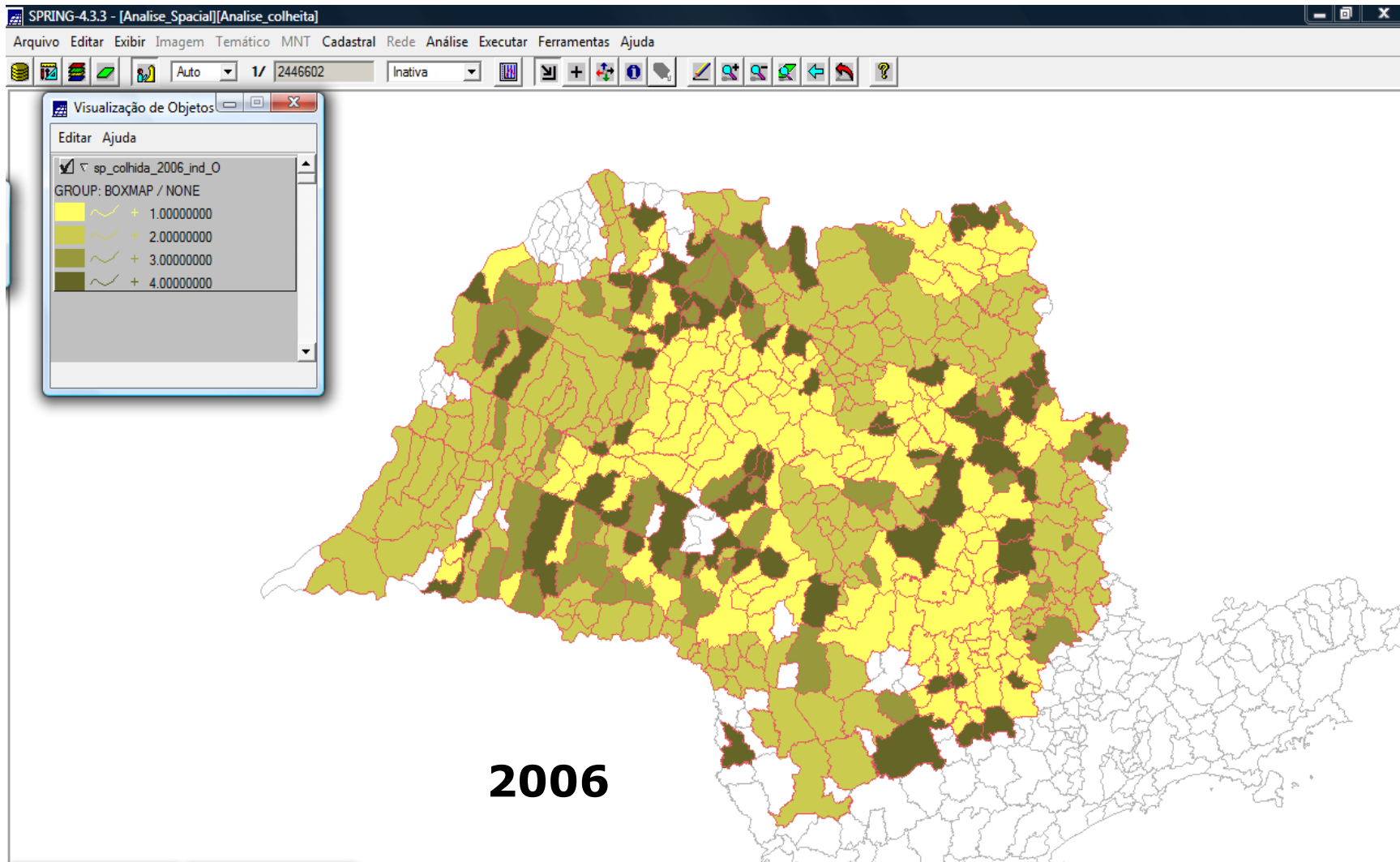
Identificação de agrupamentos espacial para o modo de colheita:

Índice Global de Moran:



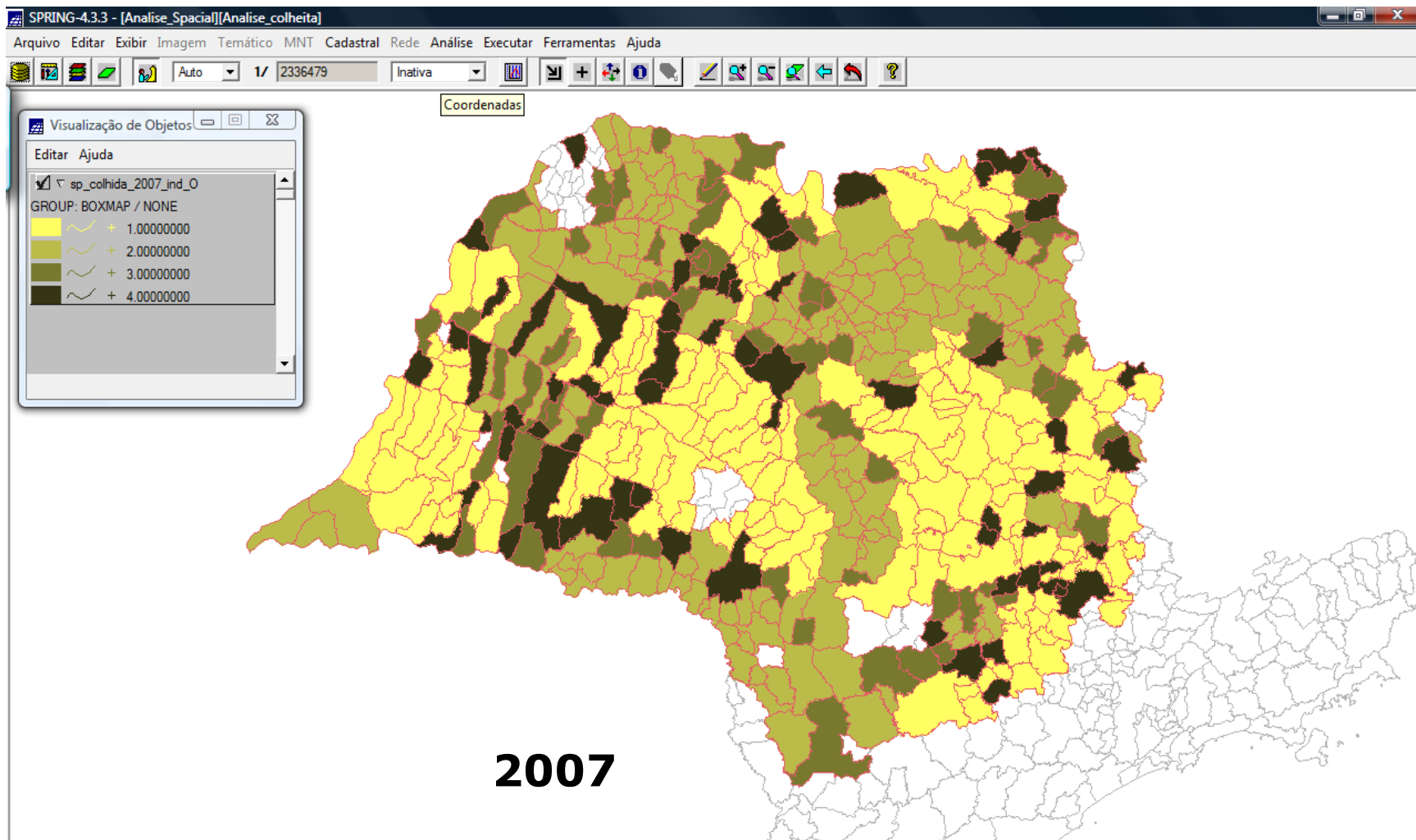


Explorando o problema...



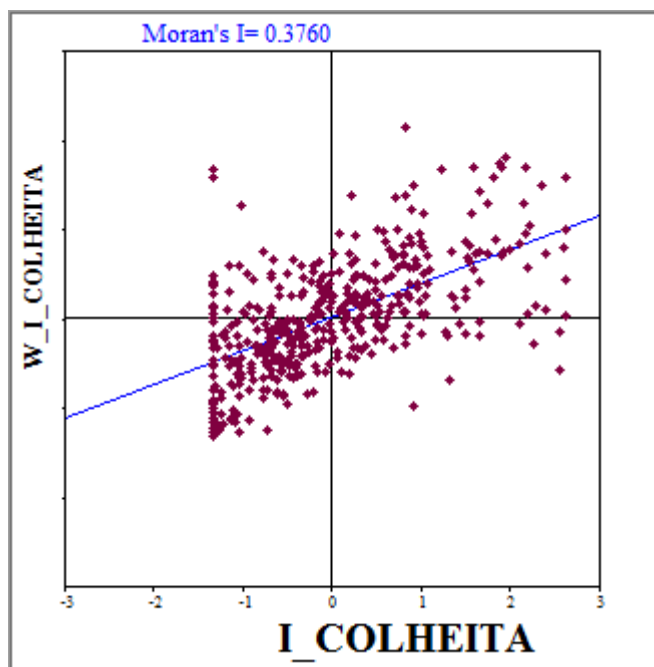


Explorando o problema...



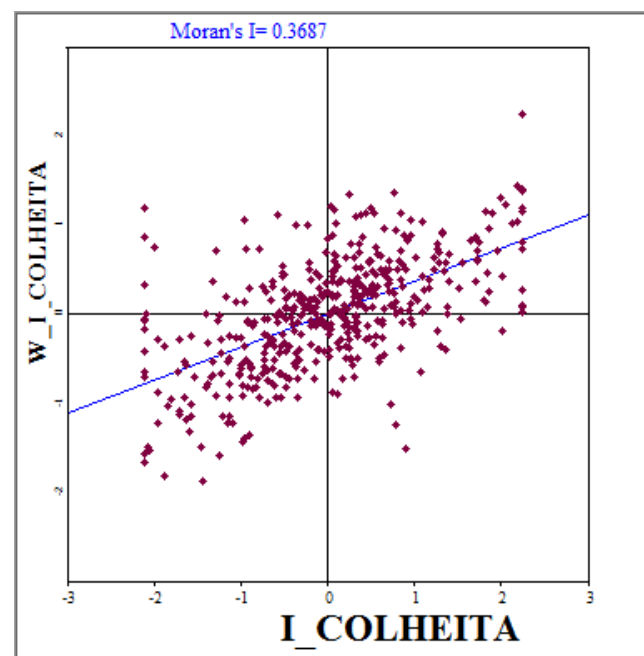
Explorando o problema...

2006



I=0,3760

2007



I=0,3687



Metodologia

Seleção de Variáveis:

- PIB per capita (SEADE, 2005);
- % de empregados na agricultura (SEADE, 2006 e 2007);
- % de empregados na indústria (SEADE 2006 e 2007);
- Rendimento médio na agricultura (SEADE 2006);
- Rendimento médio na indústria (SEADE 2006);
- IDH municipal (IBGE, 2000)
- IPRS (SEADE, 2005)
- Riqueza (SEADE, 2005)
- Longevidade (SEADE, 2005)
- Instrução (SEADE, 2005)



Metodologia

Seleção de Variáveis:

	PIB.PC	AGRO.PERC	AGRO.REND	IDHM	RIQUEZA
PIB.PC	1.000000000	-0.08437037	0.20019694	0.009894945	0.40581817
AGRO.PERC	-0.084370369	1.000000000	0.16751998	-0.072462386	-0.21836453
AGRO.REND	0.200196938	0.16751998	1.000000000	-0.030866130	0.40670329
IDHM	0.009894945	-0.07246239	-0.03086613	1.000000000	0.09788768
RIQUEZA	0.405818173	-0.21836453	0.40670329	0.097887681	1.000000000
LONGEVIDAD	0.135272899	-0.13578961	0.10325213	0.065461361	0.17422423
INSTRUCAO	0.046440872	-0.08200250	0.05340731	0.083990828	-0.06832013
IND.COLHEI	0.024594136	0.10228609	0.05137683	0.060970594	0.04466456

	LONGEVIDAD	INSTRUCAO	IND.COLHEI
PIB.PC	0.13527290	0.046440872	0.024594136
AGRO.PERC	-0.13578961	-0.082002503	0.102286089
AGRO.REND	0.10325213	0.053407309	0.051376831
IDHM	0.06546136	0.083990828	0.060970594
RIQUEZA	0.17422423	-0.068320127	0.044664565
LONGEVIDAD	1.000000000	0.189424239	-0.063218325
INSTRUCAO	0.18942424	1.000000000	-0.009212359
IND.COLHEI	-0.06321833	-0.009212359	1.000000000

2006

	PIB.PC	AGRO.REND	IDHM	RIQUEZA	LONGEVIDAD
PIB.PC	1.000000000	0.193172849	0.004393316	0.39273779	0.11870073
AGRO.REND	0.193172849	1.000000000	-0.037995371	0.40547281	0.10509648
IDHM	0.004393316	-0.037995371	1.000000000	0.08333660	0.05292766
RIQUEZA	0.392737792	0.405472815	0.083336597	1.000000000	0.16760600
LONGEVIDAD	0.118700729	0.105096476	0.052927664	0.16760600	1.000000000
AGRO.PERC	-0.089902513	0.170854259	-0.065025038	-0.21096346	-0.14493323
INSTRUCAO	0.051138486	0.052612600	0.078868638	-0.06349051	0.19791180
IND.COLHEI	0.033034785	-0.004921475	0.013558089	0.05431418	-0.03281345

	AGRO.PERC	INSTRUCAO	IND.COLHEI
PIB.PC	-0.08990251	0.05113849	0.033034785
AGRO.REND	0.17085426	0.05261260	-0.004921475
IDHM	-0.06502504	0.07886864	0.013558089
RIQUEZA	-0.21096346	-0.06349051	0.054314179
LONGEVIDAD	-0.14493323	0.19791180	-0.032813454
AGRO.PERC	1.000000000	-0.09074434	0.010446317
INSTRUCAO	-0.09074434	1.000000000	0.061436516
IND.COLHEI	0.01044632	0.06143652	1.000000000

2007



Metodologia

Variáveis selecionadas:

- PIB per capita (SEADE, 2005);
- % de empregados na agricultura (SEADE, 2006 e 2007);
- % de empregados na indústria (SEADE 2006 e 2007);
- Rendimento médio na agricultura (SEADE 2006);
- Rendimento médio na indústria (SEADE 2006);
- IDH municipal (IBGE, 2000)
- IPRS (SEADE, 2005)
- Riqueza (SEADE, 2005)
- Longevidade (SEADE, 2005)
- Instrução (SEADE, 2005)



Metodologia

Regressão Linear simples:

Modelar a relação linear entre variáveis de modo que uma delas possa ser descrita e o seu valor estimado pelas demais

O modelo de regressão linear simples pode ser escrito como:

$$y = \beta_0 + \beta_1 X + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

β_0 : intercepto

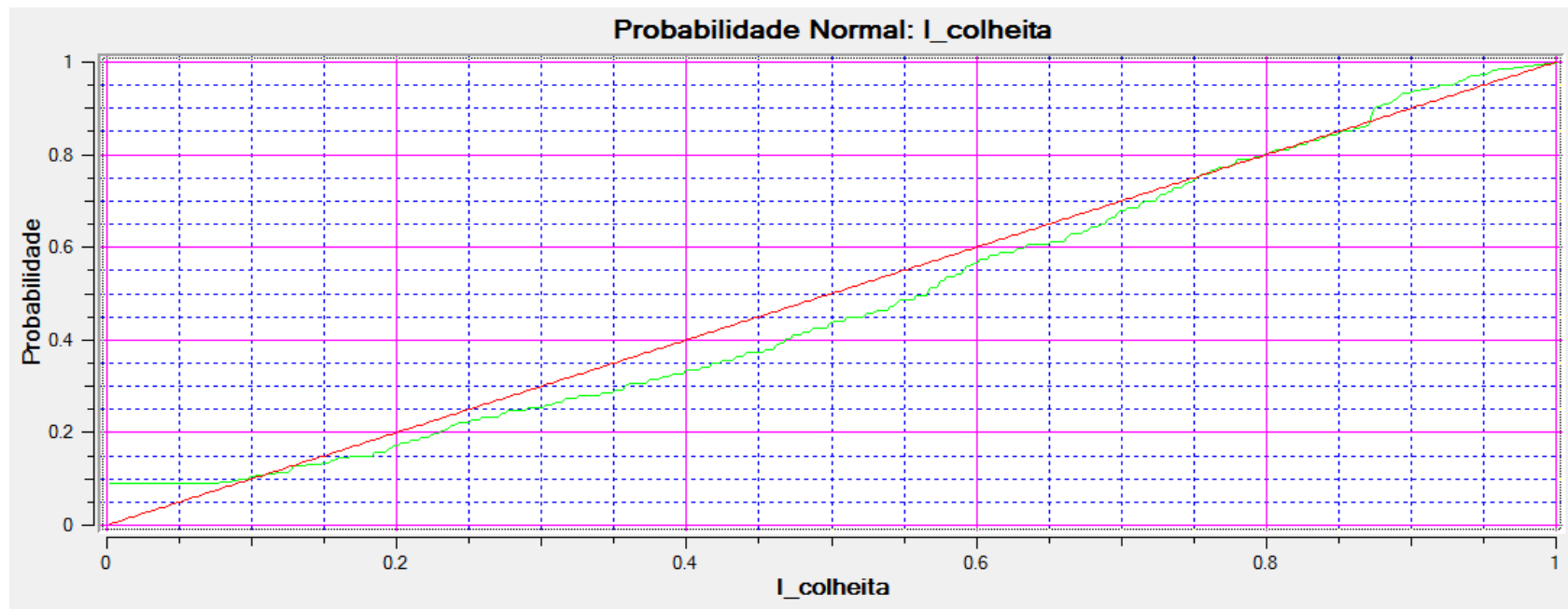
β_1 : inclinação

ε : componente de erro aleatório

Metodologia

Regressão Linear simples:

Teste de Normalidade da variável resposta "Índice de Colheita":



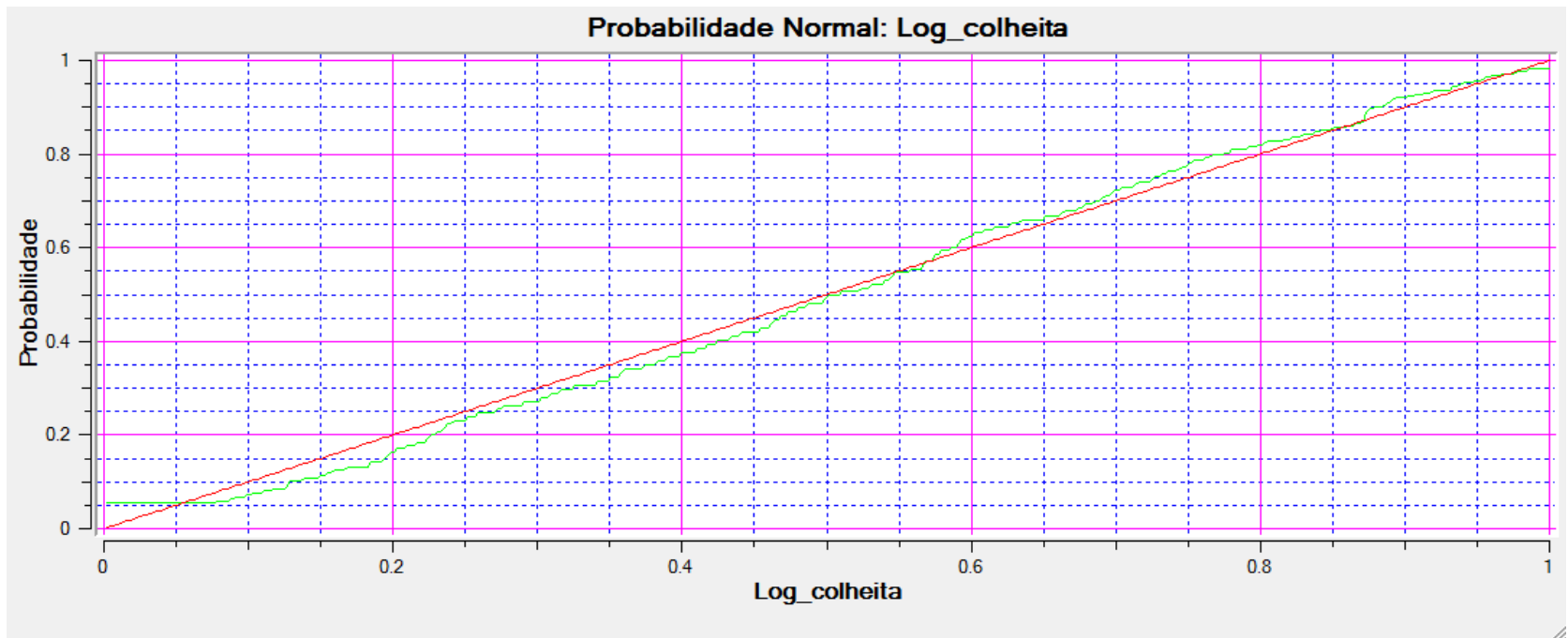
p-valor: 0.0906 (sig. de 0,05)

Média:	1.968260394
Desvio Padrão:	0.457623334
Mediana:	1.985000000
Variância:	0.209419116

Metodologia

Regressão Linear simples:

Transformação em Log. da variável resposta "Índice de Colheita":



p-valor: 0.0029 (sig. de 0,05)

Média:	0.471830971
Desvio Padrão:	0.295538169
Mediana:	0.470003629
Variância:	0.087342809



Metodologia

Regressão Linear simples:

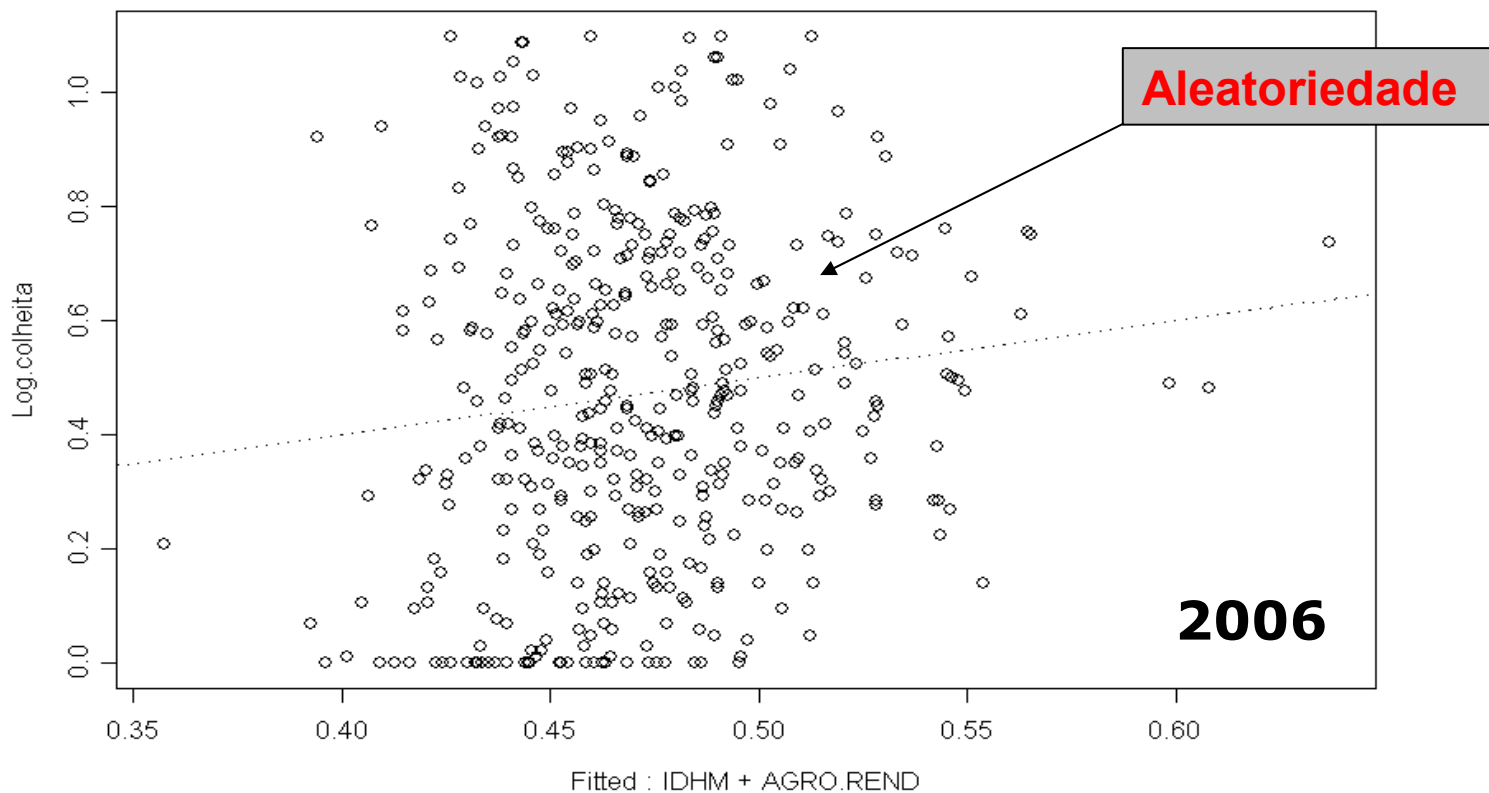
Ranking segundo o β e o p-valor

	2006	
Variável	β	Significância (p-valor)
Agro (%)	0,0017	0,0400
Agro_Renda	0,0002	0,0668
IDHm	0,7000	0,1241
Longevidade	-0,0023	0,2813
PIB per capita	0,0000	0,3199
IPRS	0,0070	0,5841
Instrução	-0,0004	0,8310

	2007	
Variável	β	Significância (p-valor)
IPRS	0,0261	0,0114
Instrução	0,0026	0,1221
PIB per capita	0,0000	0,3902
Agro_Renda	0,0001	0,4461
IDHm	0,2581	0,4867
Longevidade	-0,0006	0,7434
Agro (%)	0,0001	0,8356

Metodologia

Regressão Linear simples:

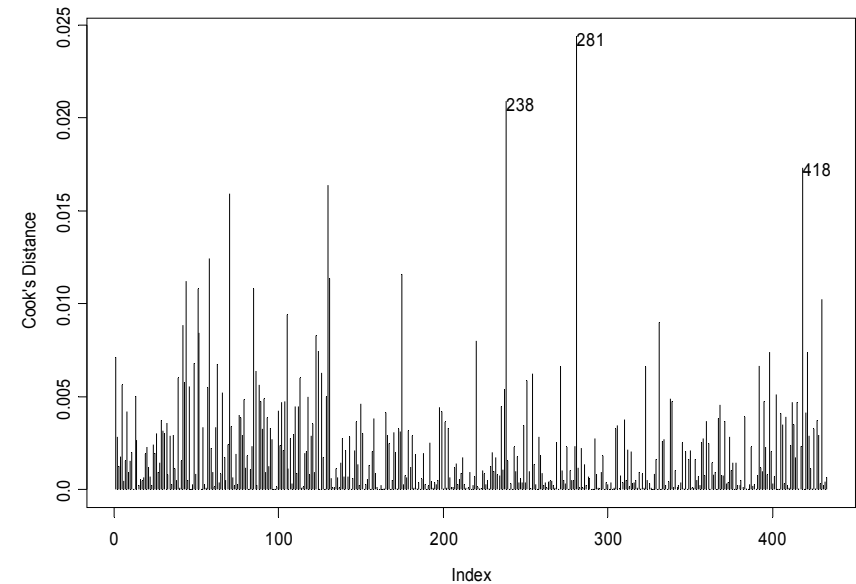
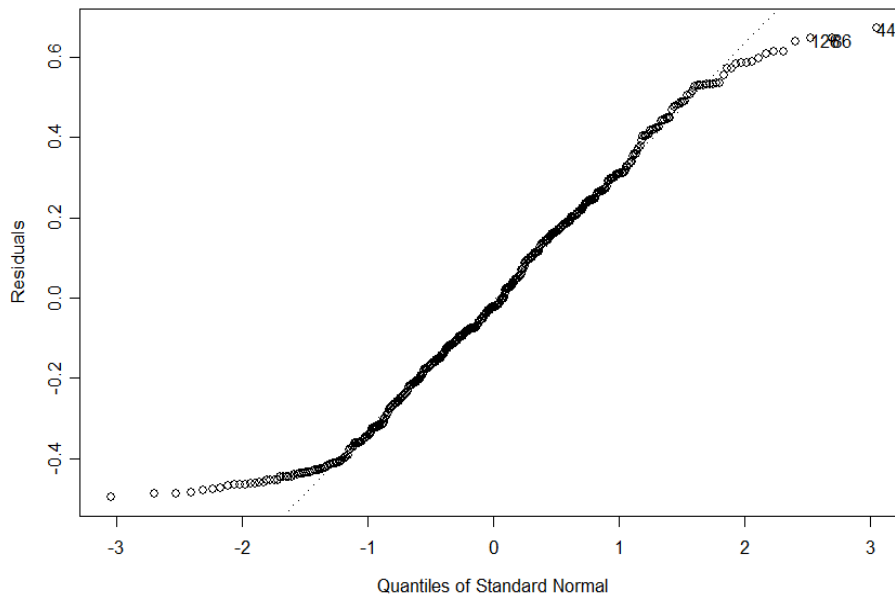


Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-0.1950789	0.359115	-0.5432212	0.5872515
IDHM	0.7265766	0.453234	1.603094	0.1096481
AGRO_REND	0.0001552477	8.215736e-005	1.889639	0.0594793

R^2 0.013669
 Log likelihood -83.612
 Akaike info criterion 173.224

Metodologia

Análise de resíduos: 2006

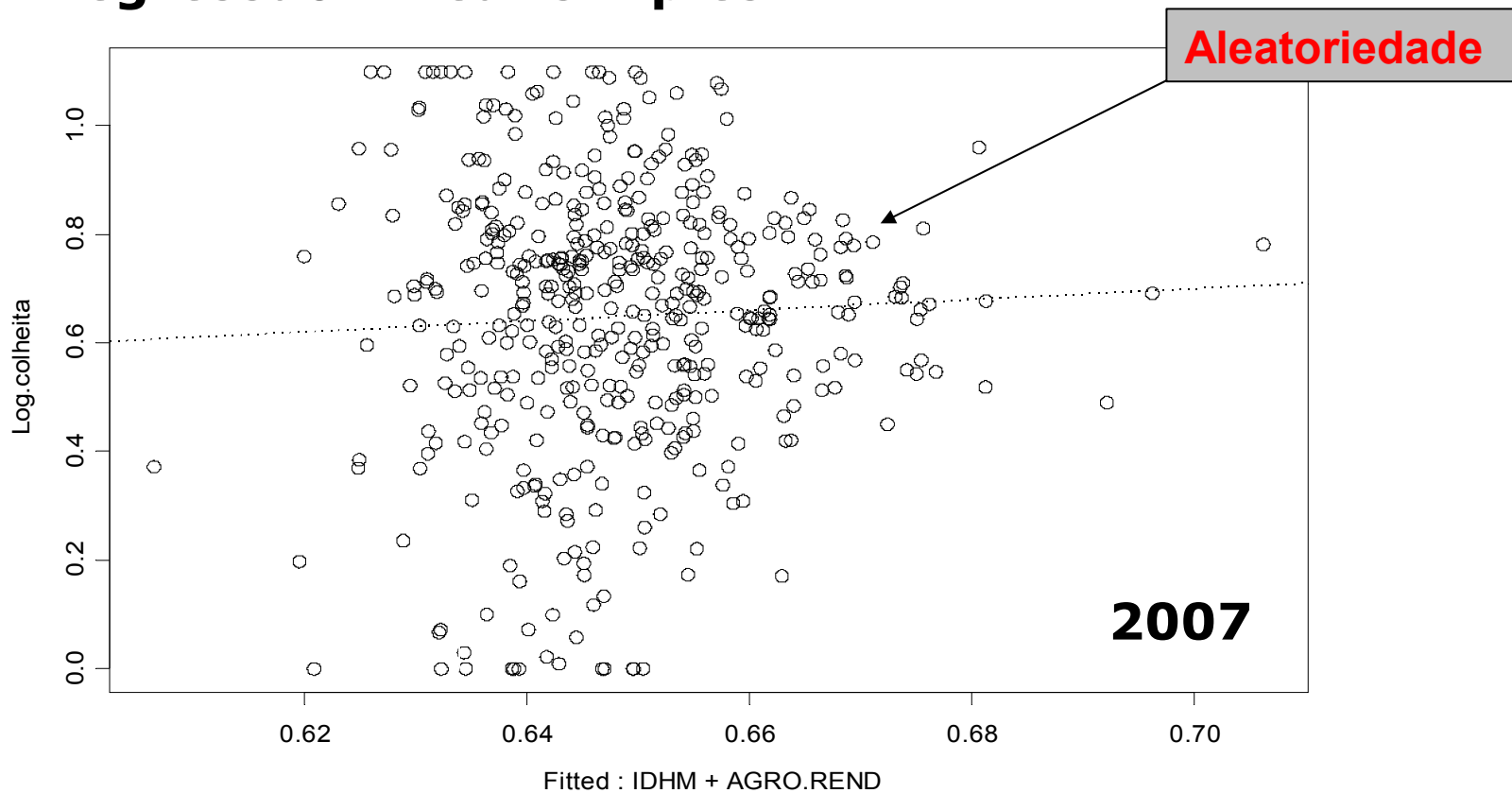


TEST	DF	VALUE	PROB
Jarque-Bera 0.0004458	2	15.43109	
Kolmogorov-Smirnov		0.0489	0.0149
Breusch-Pagan 0.0020784	2	12.35233	
Moran's I (error)		0.379860	



Metodologia

Regressão Linear simples:



Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.4037215	0.2945464	1.370655	0.1711585
IDHM	0.2692617	0.3712067	0.7253687	0.4686031
AGRO_REND	5.3766e-005	6.8121e-005	0.7892286	0.4303860

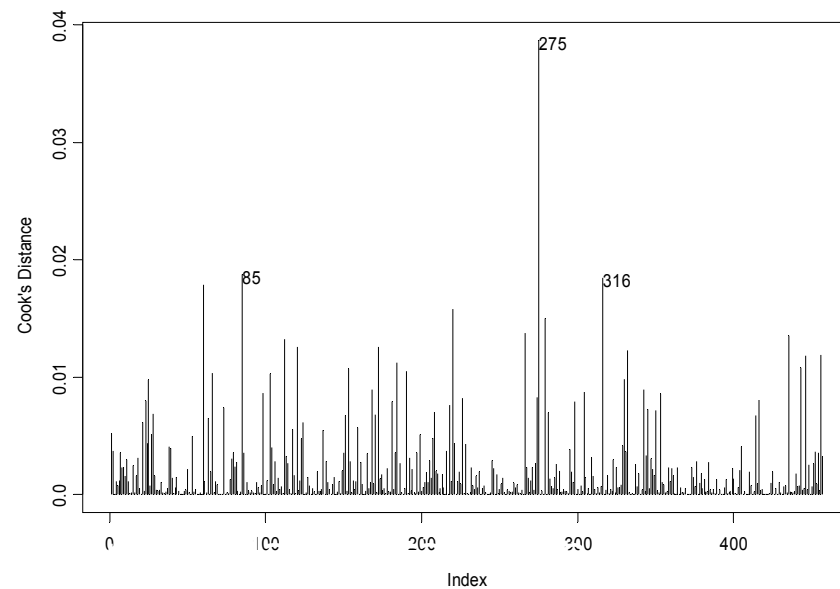
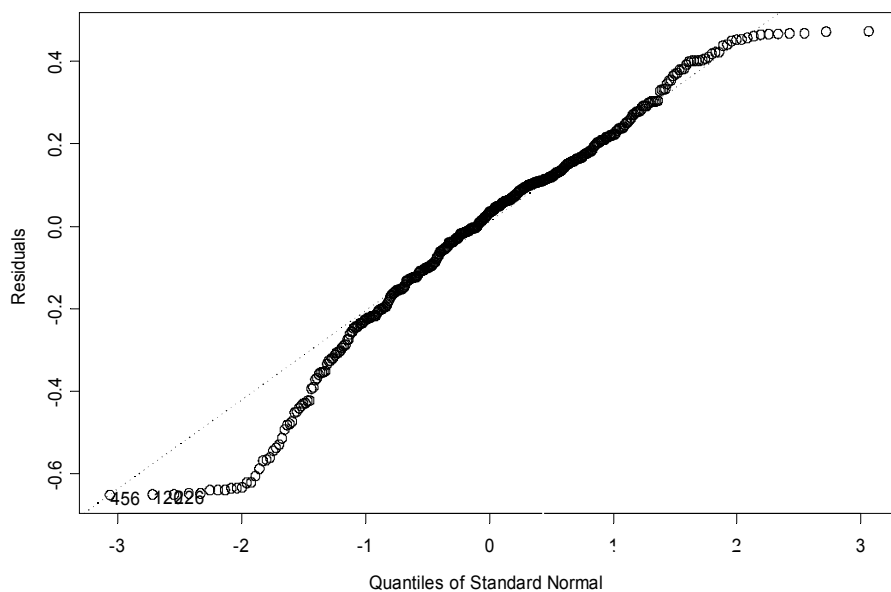
R^2

Log likelihood
Akaike

0.002433
-7.75115
21.5023

Metodologia

Análise de resíduos: 2006



TEST	DF	VALUE	PROB
Jarque-Bera 0.0000123	2	22.60708	
Kolmogorov-Smirnov		0.0714	0
Breusch-Pagan 0.0000007	2	28.25303	
Moran's I (error)		0.360212	



Metodologia

Conclusão preliminar

As variáveis independentes explicam pouco do fenômeno nos 2 anos de análise



Regressão Espacial

Os resíduos apresentaram uma dependência espacial



Metodologia

Regressão Espacial com efeitos Globais (*Spatial Lag*):

Este modelo atribui a autocorrelação espacial à variável dependente Y expresso por:

$$Y = \rho WY + X\beta + \varepsilon$$

Onde:

W = matriz de correlação espacial;

WY = expressa a dependência espacial;

ρ = coeficiente espacial autoregressivo



Resultados

Regressão Espacial com efeitos Globais (*Spatial Lag*):

2006

Regressão Linear
Simples

R² 0.013669
Log likelihood -83.612
Akaike info criterion 173.224

Regressão Espacial
Spatial lag

R² :
0.369569
Log likelihood : 4.84555
Akaike : -1.69111

2007

Regressão Linear
Simples

R² 0.002433
Log likelihood -7.75115
Akaike 21.5023

Regressão Espacial
Spatial lag

R² :
0.308261
Log likelihood : 55.4867
Akaike : -102.973



Resultados

Regressão Espacial com efeitos Globais (*Spatial Lag*):

2006

Variable	Coefficient	Std.Error	z-value	Probability
W_LOG_HEITA	0.9891158	0.004877616	202.7867	0.0000000
CONSTANT	-0.5913959	0.2861261	-2.066907	0.0387428
IDHM	0.6957107	0.3611198	1.926537	0.0540372
AGRO_REND	7.571595e-005	6.54614e-005	1.15665	0.2474154

Regressão Linear Simples - 2006

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-0.1950789	0.359115	-0.5432212	0.5872515
IDHM	0.7265766	0.453234	1.603094	0.1096481
AGRO_REND	0.0001552477	8.215736e-005	1.889639	0.0594793



Resultados

Regressão Espacial com efeitos Globais (*Spatial Lag*):

2007

Variable	Coefficient	Std.Error	z-value	Probability
W_LOG_HEITA	0.6433995	0.04466163	14.40609	0.0000000
CONSTANT	0.04217049	0.2460596	0.1713833	0.8639224
IDHM	0.2116806	0.3080961	0.6870603	0.4920446
AGRO_REND	3.472225e-005	5.654494e-005	0.6140646	0.5391725

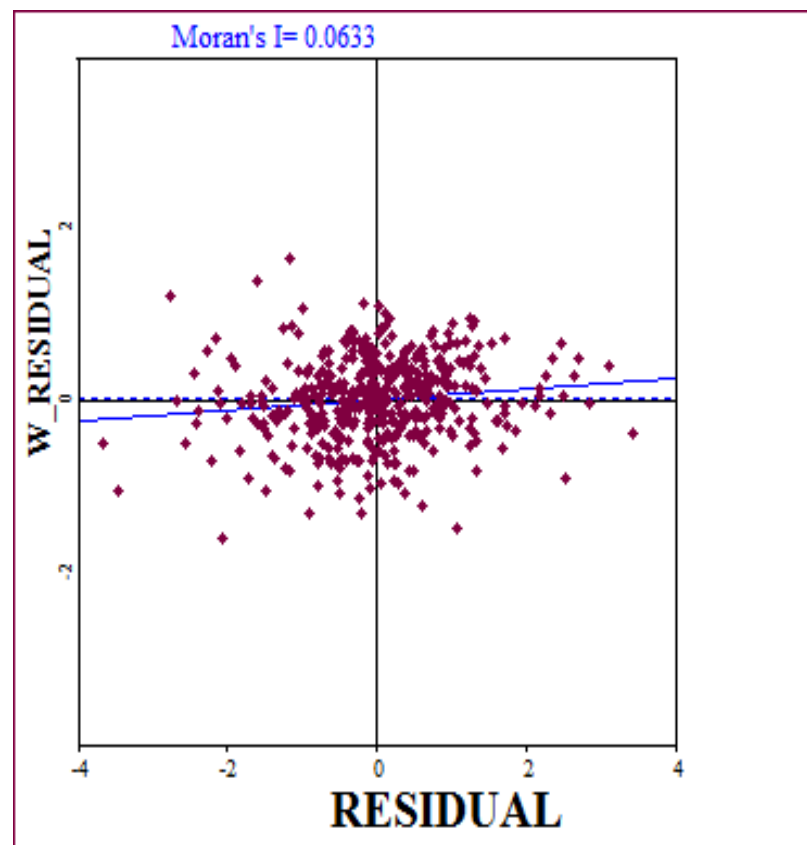
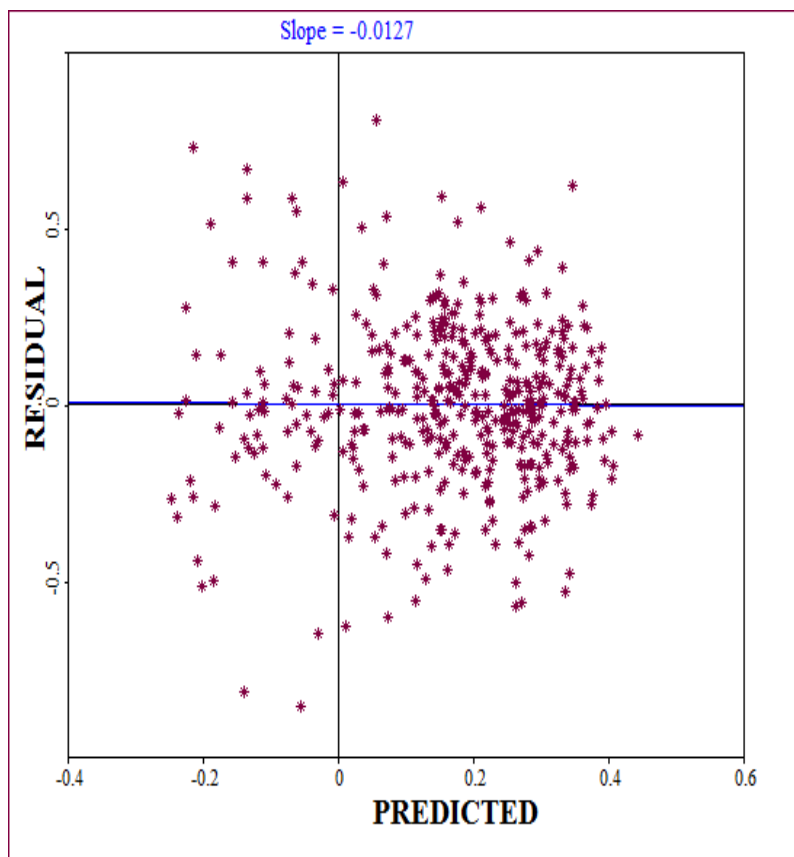
Regressão Linear Simples - 2006

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.4037215	0.2945464	1.370655	0.1711585
IDHM	0.2692617	0.3712067	0.7253687	0.4686031
AGRO_REND	5.3766e-005	6.8121e-005	0.7892286	0.4303860

Resultados

Análise de Resíduos: 2006

$I = 0,0633$

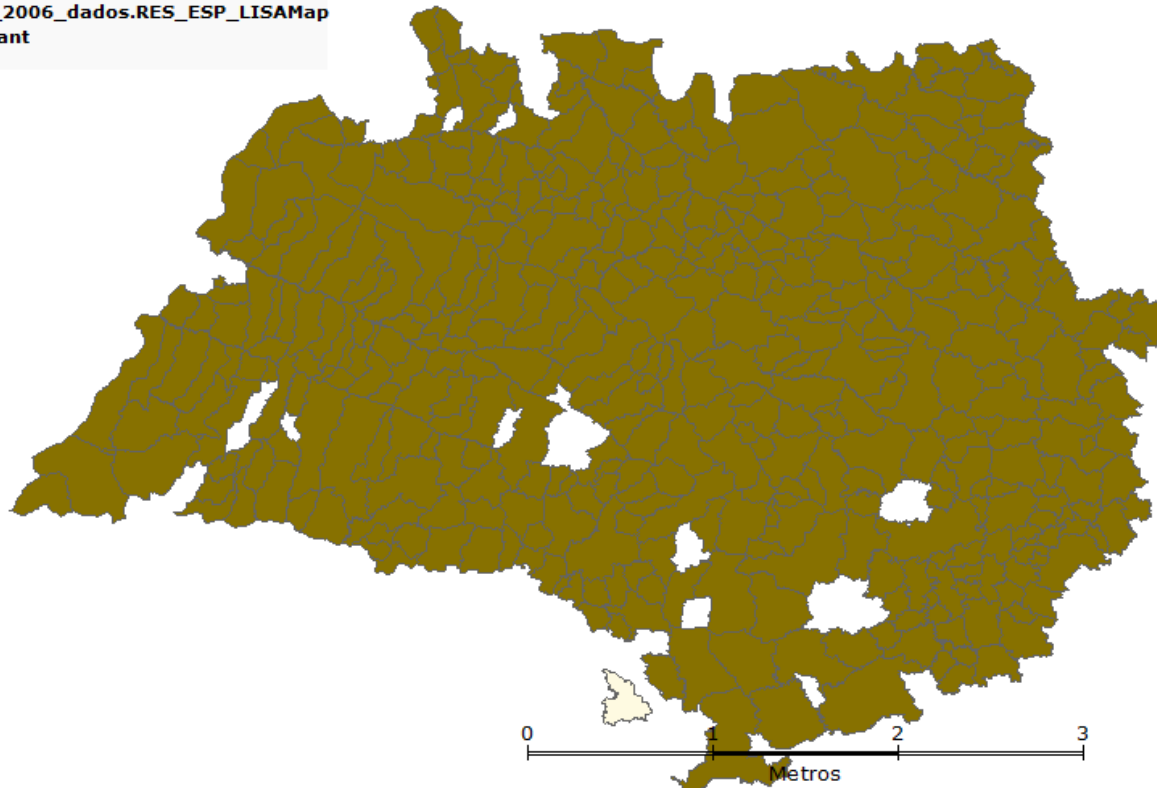




Resultados

Análise de Resíduos: 2006

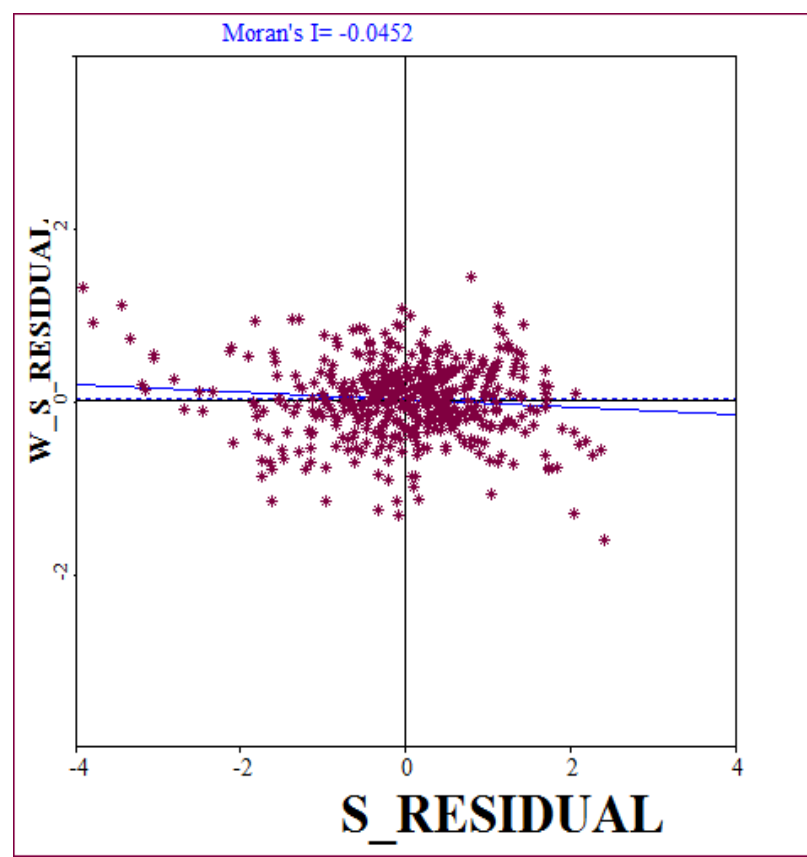
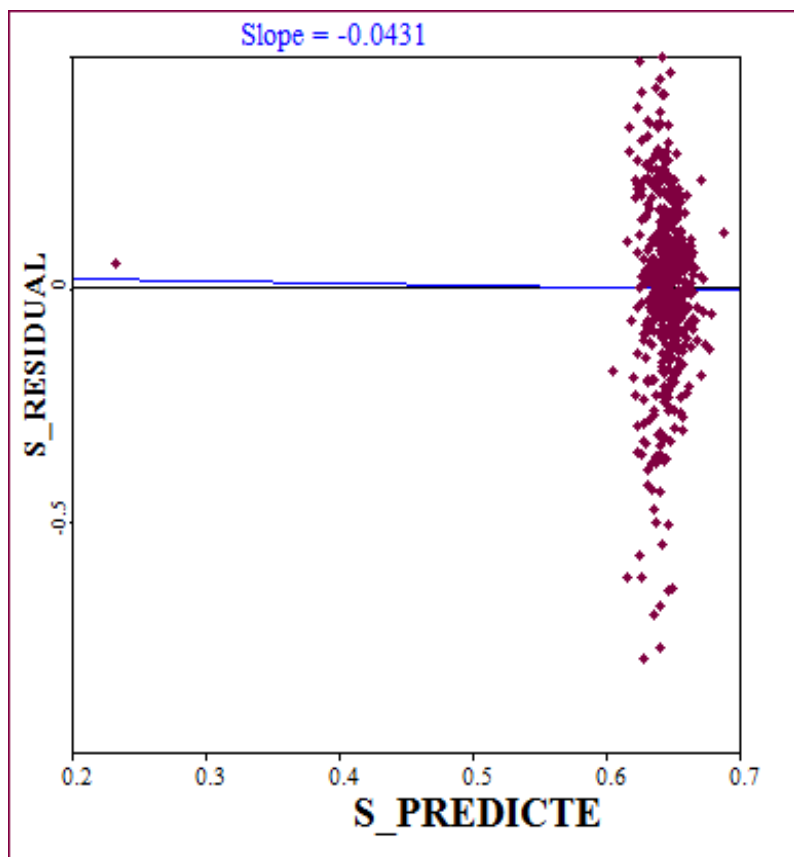
sp_colhida_2006_dados.RES_ESP_LISAMap
not-significant
High-Low



Resultados

Análise de Resíduos: 2007

$I = -0,0452$



Resultados

Análise de Resíduos: 2007

sp_colhida_2007_dados.LISASig

- not-significant
- High-high





Considerações Finais

O modelo por regressão espacial demonstrou que a variável resposta possui forte dependência espacial.

O modelo por regressão espacial permitiu explicar que há evidências que o manejo da colheita da cana sem o uso do fogo ocorre em municípios com níveis sócio-econômico mais elevados em parte retratados através do IDH municipal e da renda média na agricultura.

Apesar de algumas diferenças de significância das variáveis explicativas entre os dois períodos analisados, no ano de 2007 torna-se evidente a tendência de adequação dos fazendeiros aos novos protocolos ambientais.



Softwares Utilizados

TerraView 3.2.0 (Banco de dados, análise exploratória)

S-Plus 4.5 (Regressão linear simples, testes estatísticos)

Spring 4.3.3 (Análise exploratória)

GeoDa 0.9.5i (Regressão espacial - *Spatial lag*)