Versão 2.1.15_06_26

Tutorial

Sumário

Introdução 4
O que é o IpeaGEO 4
Quem somos 4
Instalação do IpeaGEO 6
Onde Encontrar 6
Requisitos Tecnológicos 6
Utilizando o IpeaGEO 7
Iniciando o IpeaGEO 7
Abrindo Projeto Salvo 9
Arquivo *Shape* (Malha Digital) 9
Importar Malha 10
Adicionar Camadas (*Layers*) Auxiliares 15
Alterando as Propriedades das Camadas (Layers) 17
Base de Dados 23
Importar Dados 23
Exportar Dados 34
Funcionalidades 36
Mapa Temático 39
Funções Estatísticas do IpeaGEO 44
Análise Exploratória dos Dados 44
Análise Gráfica 44
Calculadora 51
Correlações 58
Estatísticas Descritivas 60
Geração de Variáveis *Dummy* 63
Matriz de Vizinhança 66
Tabelas de Frequência 74
Tabulações Cruzadas 76
Analise Multivariada 80
Analise de Clusters 80
Analise de Componentes Principais 84
Analise Fatorial 87
Cálculo de Taxas 93
Econometria Básica 97
Propensity Score Matching 97
Regressão com Dados Binários 103
Regressão Linear 107

Modelos Lineares Generalizados 111
Estatística Básica 115
Análise e testes para médias 115
Descrição dos Dados 121
Distribuições Contínuas 124
Distribuições Discretas 129
Métodos Multicritérios de Apoio as Decisões 133
Funções Espaciais do IpeaGEO 140
Áreas Mínimas Comparáveis 140
Compatibilização de Variáveis 140
Conglomerados Espaciais 147
Hierárquico 147
Dependência Espacial 156
Índices Globais 157
Índices Locais 161
Econometria Espacial 165
Cross-Section 165
Indicadores de Segregação Espacial 173
Referências Bibliográficas 181

Introdução

O que é o IpeaGEO

O IpeaGEO é um software desenvolvido pela Assessoria de Métodos Quantitativos da Diretoria de Estudos e Políticas Regionais, Urbanos e Ambientais do Ipea - Dirur.

Ele funciona como uma ferramenta gratuita de análises estatísticas, voltada especialmente para análises espaciais. A inclusão de técnicas espaciais consolidadas e foco no território nacional são alguns dos diferenciais do programa.

Quem somos

Instituto de Pesquisa Econômica Aplicada - Ipea

O Instituto de Pesquisa Econômica Aplicada (Ipea) é uma fundação pública federal vinculada à Secretaria de Assuntos Estratégicos da Presidência da República. Suas atividades de pesquisa fornecem suporte técnico e institucional às ações governamentais para a formulação e reformulação de políticas públicas e programas de desenvolvimento brasileiros. Os trabalhos do Ipea são disponibilizados para a sociedade por meio de inúmeras e regulares publicações e seminários e, mais recentemente, via programa semanal de TV em canal fechado.

Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais - Dirur

A atuação da Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) tem como foco a promoção e realização de estudos e pesquisas diretamente relacionados com a definição de estratégias de apoio à formulação e execução das políticas regional e urbana e seus respectivos instrumentos, mais especificamente sobre as questões de desenvolvimento nas esferas regional e intraurbana, as relações entre estabilidade macroeconômica e o crescimento econômico de longo prazo, assim como sobre finanças públicas e política fiscal. Incumbe-se a Dirur de subsidiar as políticas setoriais voltadas para atenuação das desigualdades regionais, superação dos problemas urbanos, e o fomento do crescimento e desenvolvimento regional e urbano.

Assessoria de Métodos Quantitativos da Dirur

Responsáveis técnicos:

Alexandre Xavier Ywata de Carvalho, Ph.D.

Pedro Henrique Melo Albuquerque, D.Sc.

Waldery Rodrigues Júnior, D.Sc.

Equipe de desenvolvimento:

Alex Rodrigues do Nascimento

Camilo Rey Laureto

Cauê de Castro Dobbin

Demerson André Polli

Fabio Augusto Scalet Medina

Fernanda Rocha Gomes da Silva

Gabriela Drummond Marques da Silva

Gilberto Rezende de Almeida Júnior

Guilherme Costa Chadud Moreira

Gustavo Gomes Basso

Igor Ferreira do Nascimento

Luiz Felipe Dantas Guimarães

Marcius Correia Lima Filho

Marina Garcia Pena

Rafael Dantas Guimarães

Instalação do IpeaGEO

Onde Encontrar

O download do software IpeaGEO é gratuito e está disponível no seguinte endereço: http://www.ipea.gov.br/IpeaGEO

Requisitos Tecnológicos

Requisitos Mínimos de Hardware/Software:

HARDWARE : Mínimo de 2 GB de disco rígido, 512 MB de memória RAM e processador de 1 GHz.

SISTEMA OPERACIONAL : Windows XP com Service Pack 2, Windows 2003, Windows Vista ou superior.

Microsoft .NET Framework 4 Client Profile (caso o usuário não possua, o IpeaGEO o encaminha para o site de download).

Conexão com a Internet é necessária para carregamento de dados e mapas hospedados no site do IPEA.

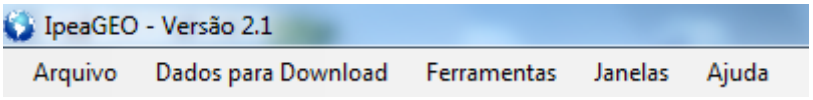
Utilizando o IpeaGEO

Iniciando o IpeaGEO

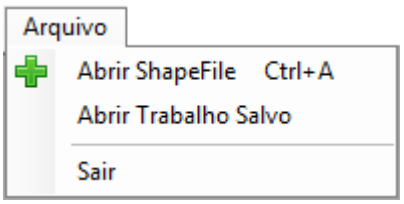
Visão Geral

Tela inicial do IpeaGEO

Ao abrir o IpeaGEO as seguintes opções estão disponíveis:



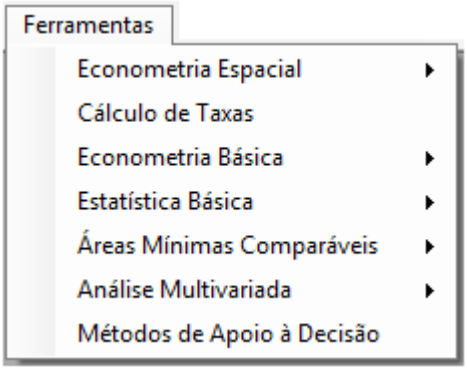
Suas funcionalidades são:



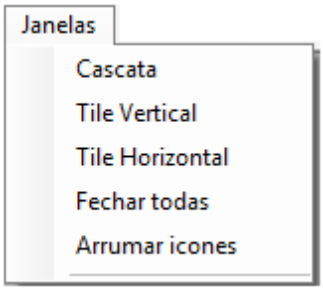
- Abrir ShapeFile (Ctrl + A) - Permite abrir um arquivo do tipo malha digital (*shape*) do seu computador.

- Abrir Trabalho Salvo – permite abrir projeto salvo no IpeaGEO.
- Sair – Encerra o IpeaGEO.

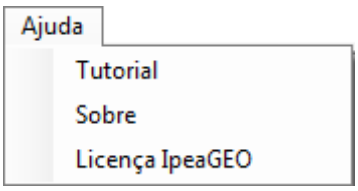
- Dados para Download – Permite ao usuário abrir um arquivo do tipo malha digital (shape) e dados já existentes no site do Ipea.



- Ferramentas – Nesse menu estão contidas as ferramentas estatísticas e econométricas que podem ser aplicadas a uma tabela de dados própria do usuário sem concatenar a uma malha digital (shape).



- Janelas – Caso haja mais de uma janela aberta no programa, essa opção permite ao usuário organiza-las de acordo com sua preferência.

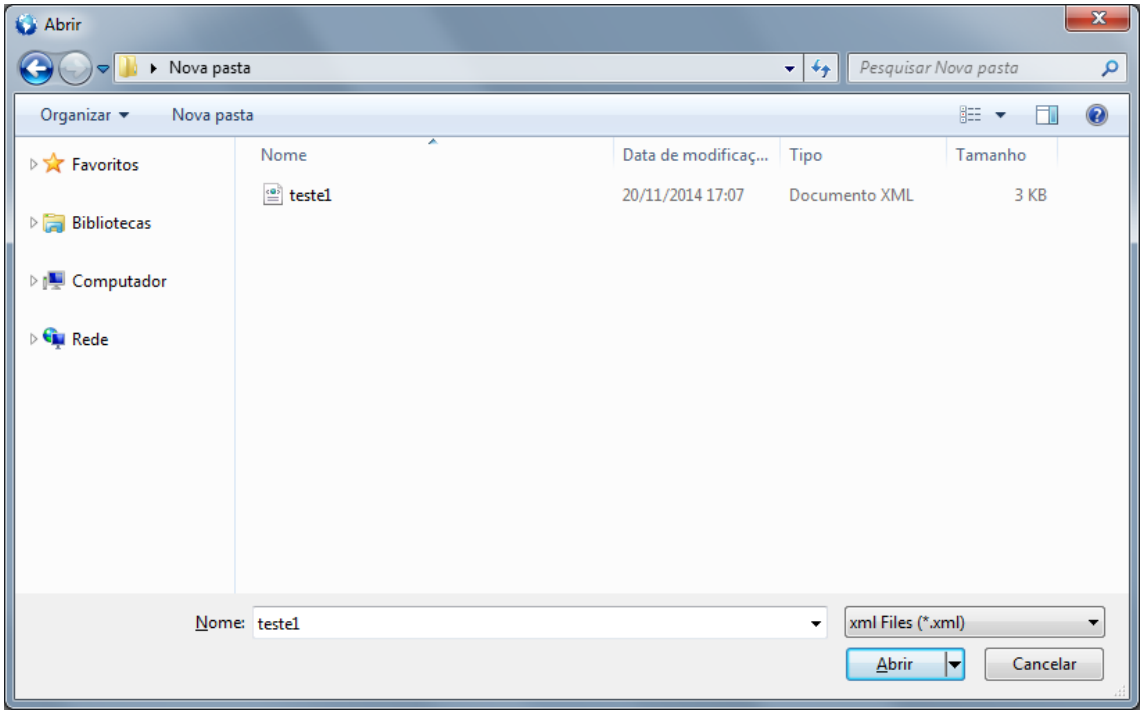


- Ajuda – Permite o acesso ao tutorial do software (este documento), e também informações sobre a equipe de desenvolvimento.

Abrindo Projeto Salvo

O botão “Abrir Trabalho Salvo” permite ao usuário abrir arquivos que tenham sido trabalhados e salvos pelo IpeaGEO. Ao salvar o projeto, dois arquivos (um .txt e outro .xml), que são necessários no processo de abertura, são criados.

Localize o arquivo .xml desejado na janela que se abrirá, como a de baixo:



O programa salva no XML a localização do arquivo shape utilizado assim como do .txt, que devem ser mantidos no mesmo local em que estavam no momento do salvamento para que o projeto possa ser recuperado.

Arquivo *Shape* (Malha Digital)

Visão Geral

O *shape* (*malha digital)* é um conjunto de arquivos que descrevem as coordenadas de pontos e de polígonos, bem como as respectivas projeções no espaço e identificações de cada objeto. Um polígono é um conjunto de pares ordenados (x, y) de coordenadas espaciais, de tal forma que o último ponto seja idêntico ao primeiro, formando uma região fechada no plano. A malha digital é composta por no mínimo três arquivos: “.shp”, “.dbf” e “.shx“. Tais arquivos devem estar obrigatoriamente armazenados na mesma pasta para que a malha digital possa ser aberta. Essas malhas podem ser encontradas no IpeaGEO por meio da opção "Dados para Download" presente no software ou no site do IBGE.

Importar Malha

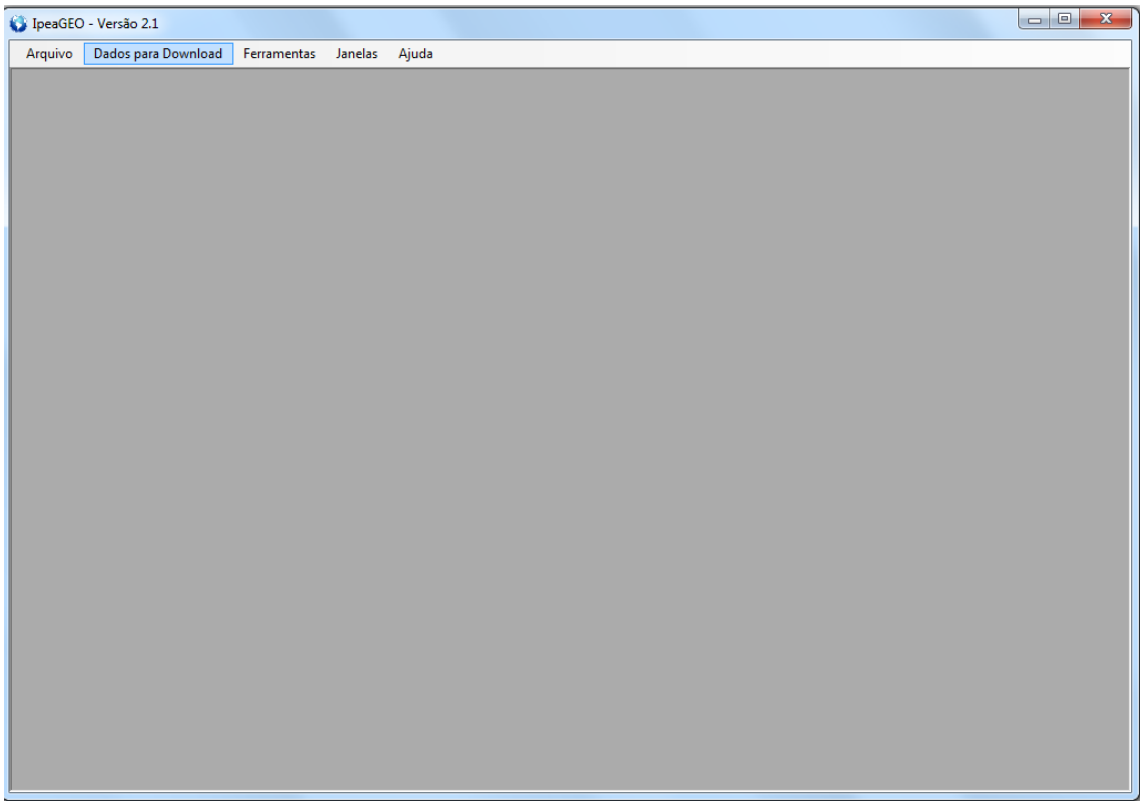
Visão Geral

No IpeaGEO, é possível importar a malha digital de duas maneiras: diretamente de um arquivo armazenado no computador do usuário ou realizar o download no próprio programa.

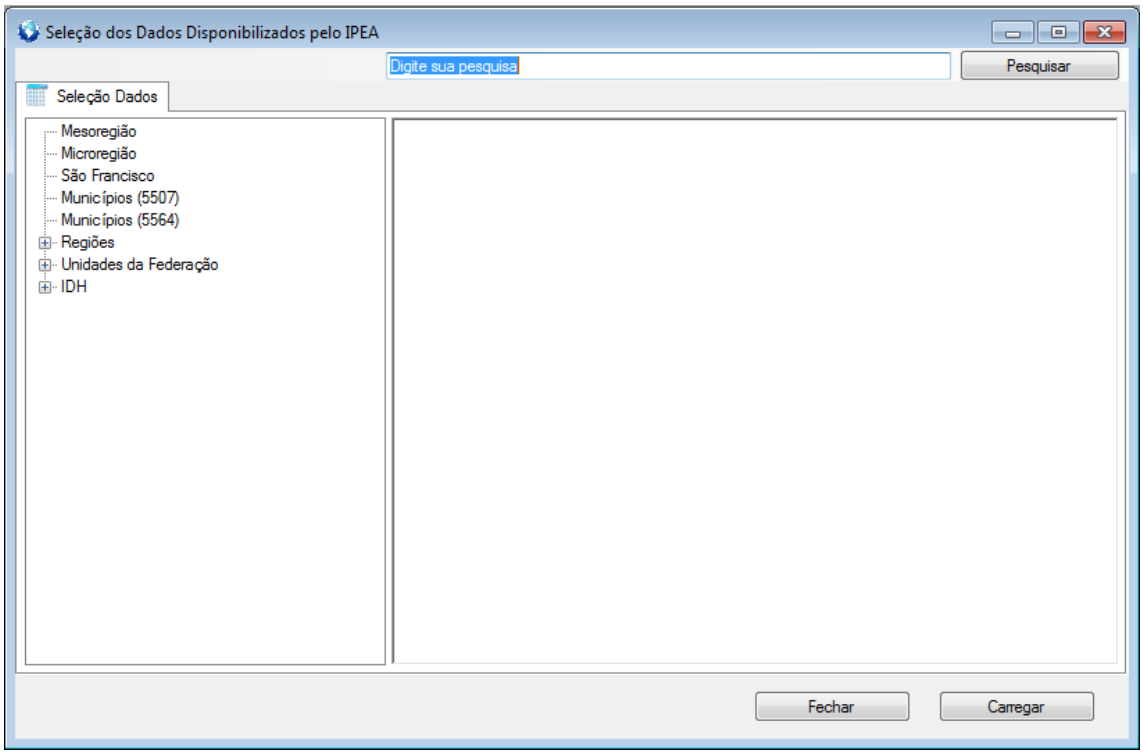
Exemplo

Importando o arquivo malha do Ipea

Para importar um arquivo de malha digital do Ipea clique em *Dados para Download*



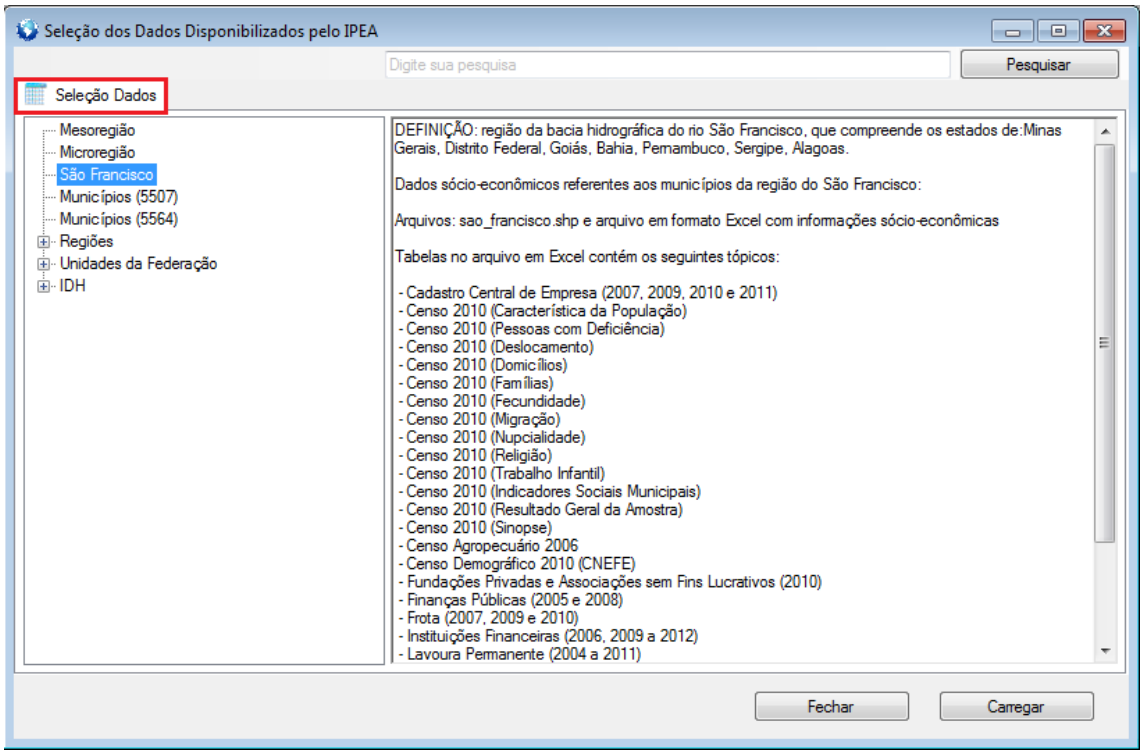
O seguinte formulário irá se abrir:



Existem duas formas de se realizar a pesquisa para importar o arquivo de malha digital no IpeaGEO: por meio do componente “Seleção Dados” ou pela caixa de pesquisa.

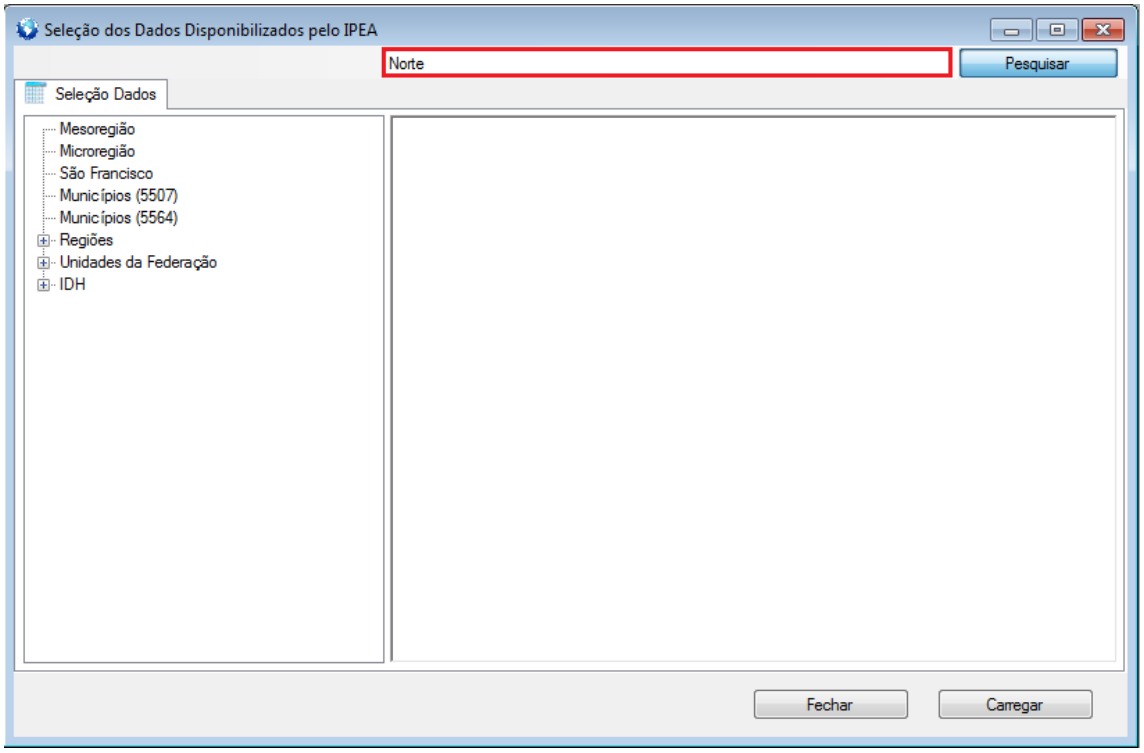
Opção 1: Componente “Seleção Dados”

Na aba “Seleção Dados” estão contidas as opções de malhas digitais disponíveis no site do Ipea. Para baixá-las, selecione a opção desejada (é possível somente uma de cada vez) e em seguida deve-se clicar no botão “Carregar”.



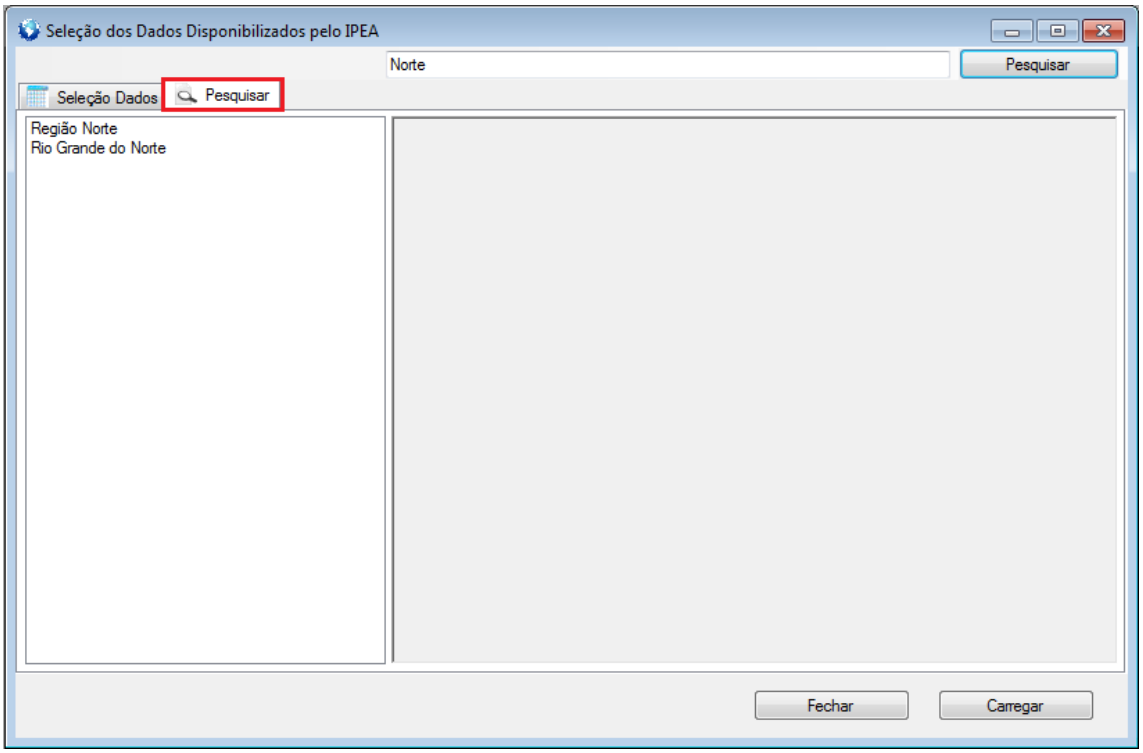
Opção 2: Caixa de pesquisa

Para baixar uma malha digital utilizando a caixa de pesquisa deve-se digitar a palavra de interesse e, em seguida, clicar no botão “Pesquisar”.

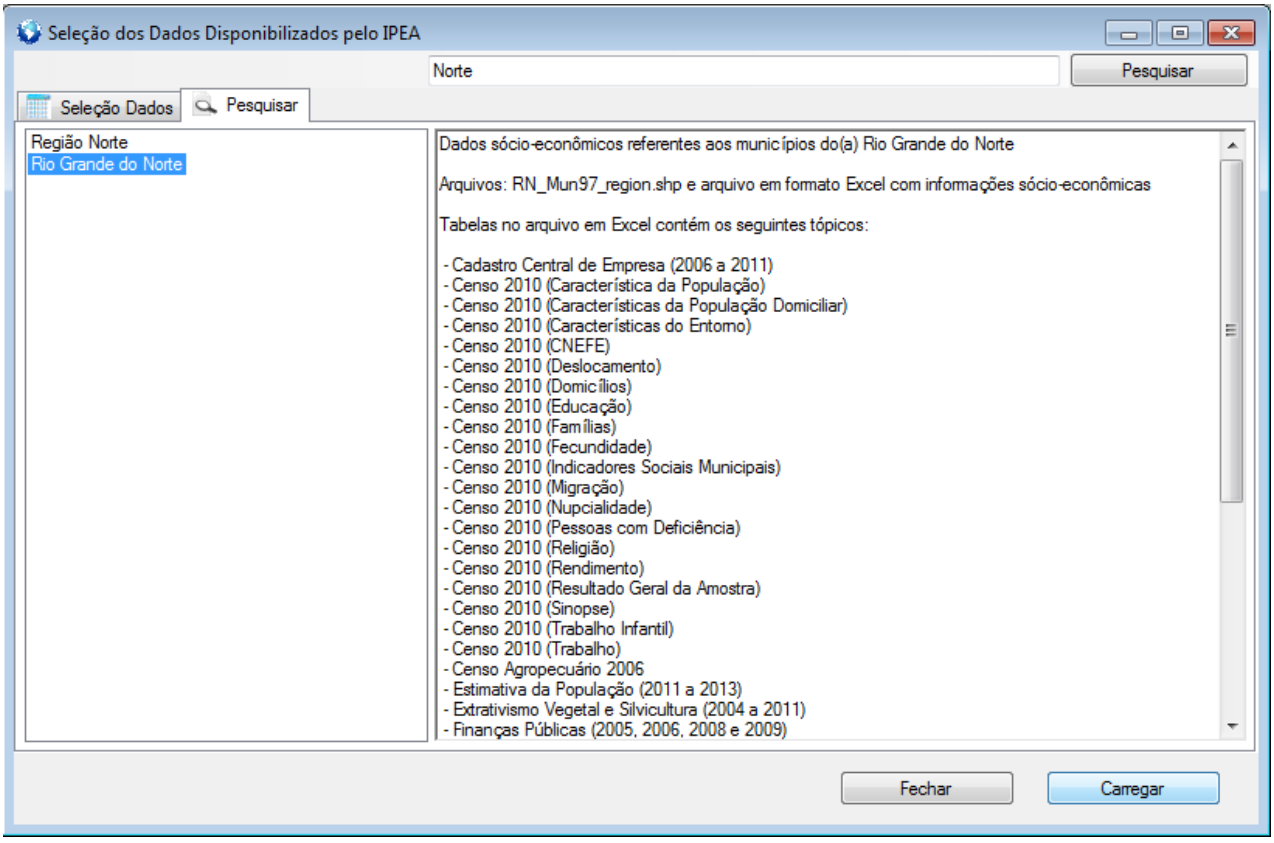


Após clicar no botão pesquisar uma nova aba ficará disponível com os dados

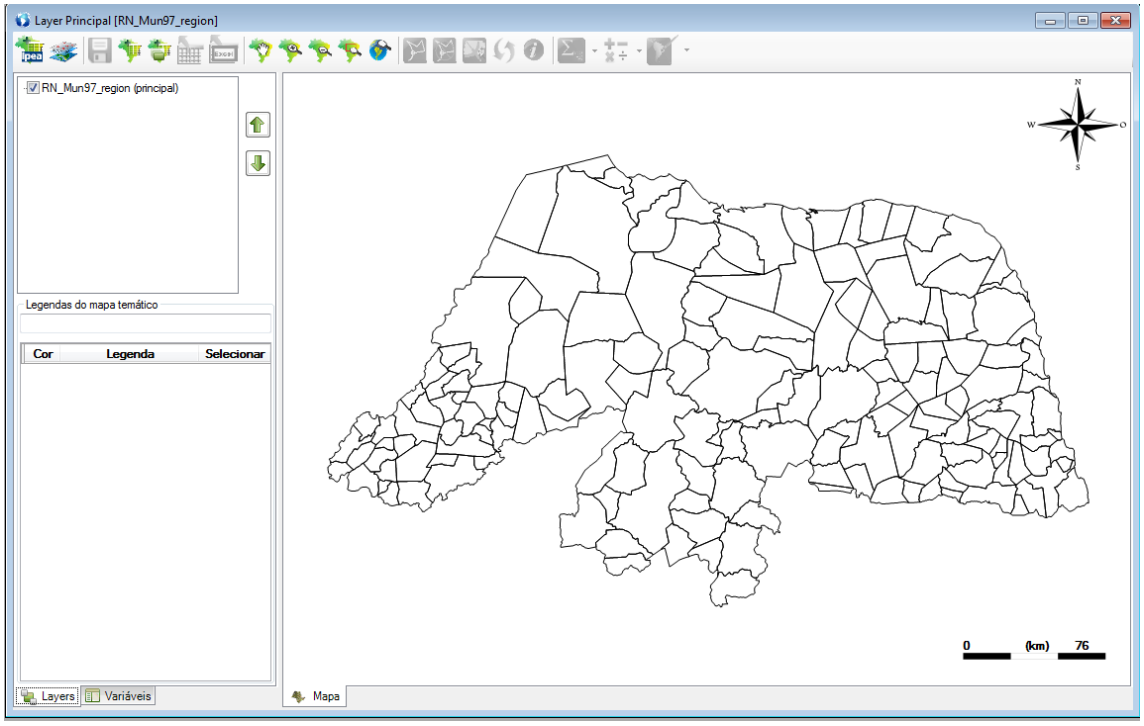
encontrados na pesquisa.



Deve-se selecionar o arquivo desejado e em seguida clicar no botão “Carregar” para importar o arquivo malha digital de interesse.

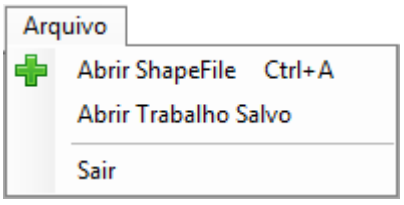


Em qualquer um dos procedimentos acima uma nova janela irá se abrir, contendo o arquivo malha digital selecionado.

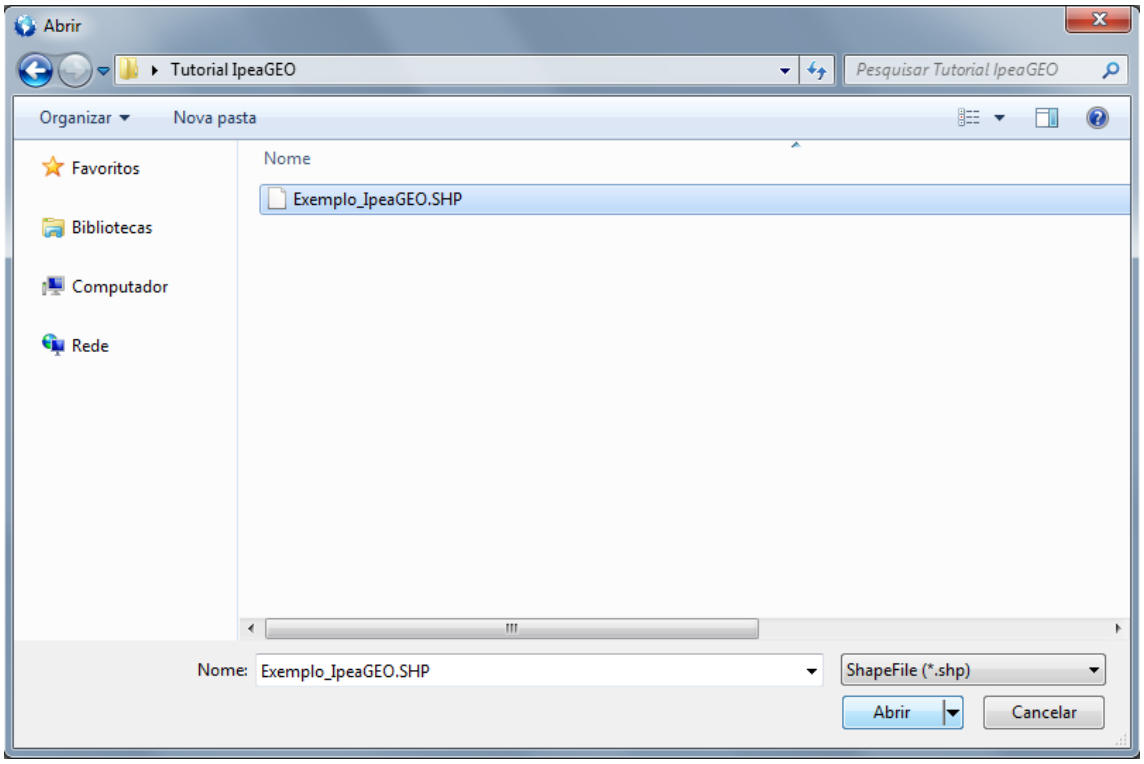


Importando o arquivo malha do computador

Para abrir um arquivo de malha digital próprio, localizado em seu computador, clique em *Arquivo* e em seguida em *Abrir ShapeFile*:



Localize e selecione o arquivo de interesse com a extensão “.shp” e clique em abrir:



Para que a malha seja importada corretamente, os arquivos com extensões “.dbf” e “.shx” também precisam estar na pasta onde está localizado o arquivo “.shp”.

Adicionar Camadas (*Layers*) Auxiliares

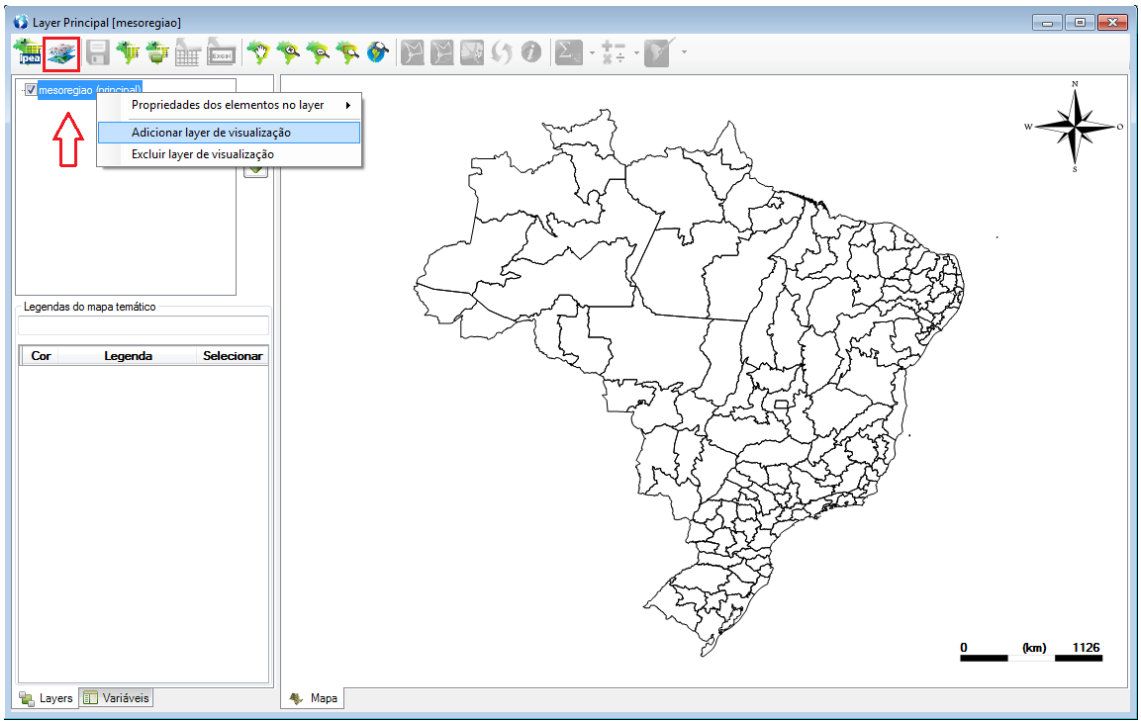
Visão Geral

No IpeaGEO, é possível importar diversas camadas (*layers*) para serem trabalhadas em conjunto com o arquivo principal da malha digital.

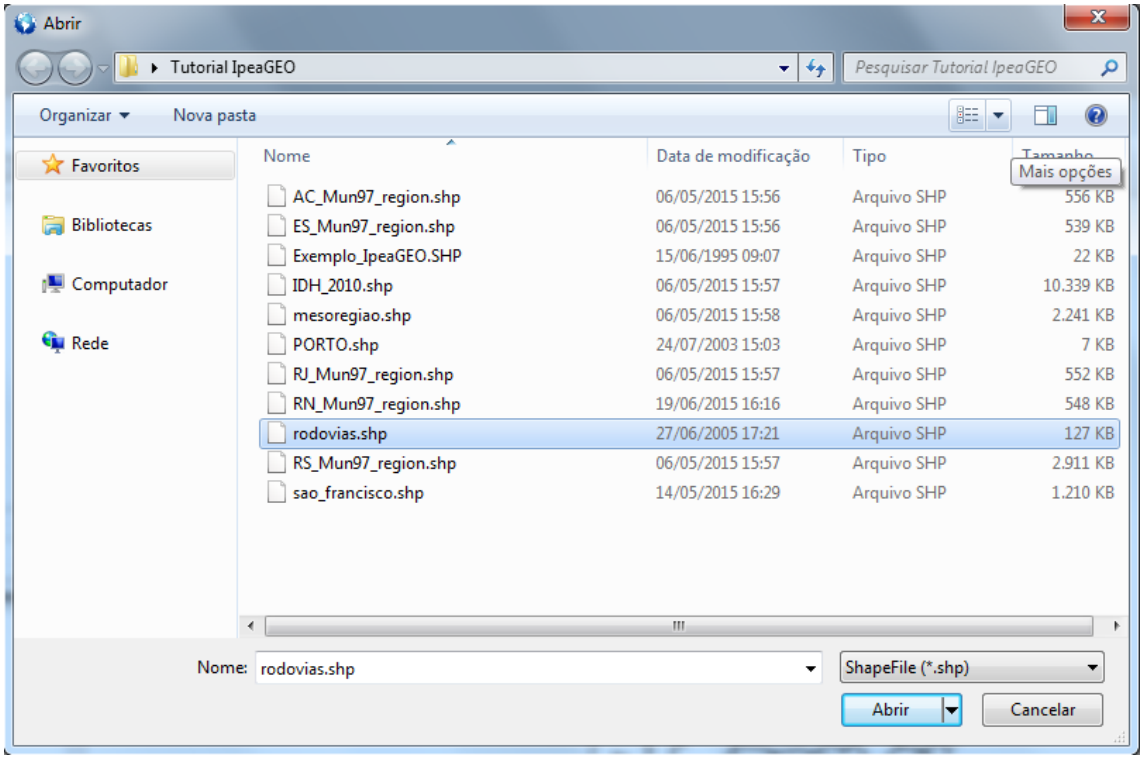
Exemplo

Após importar a malha digital por quaisquer das maneiras descritas no tópico "Importar Malha", é possível adicionar camadas (*layers*) auxiliares para serem trabalhadas conjuntamente com a malha digital principal. Existem duas maneiras diferentes para importação dessas camadas (*layers*):

Clique no botão "*Adição de layers auxiliares ao mapa*" , ou clique com o botão direito do mouse, sobre o mapa que representa a camada (layer) principal e, em seguida, clique em "*Adicionar layer de visualização*".

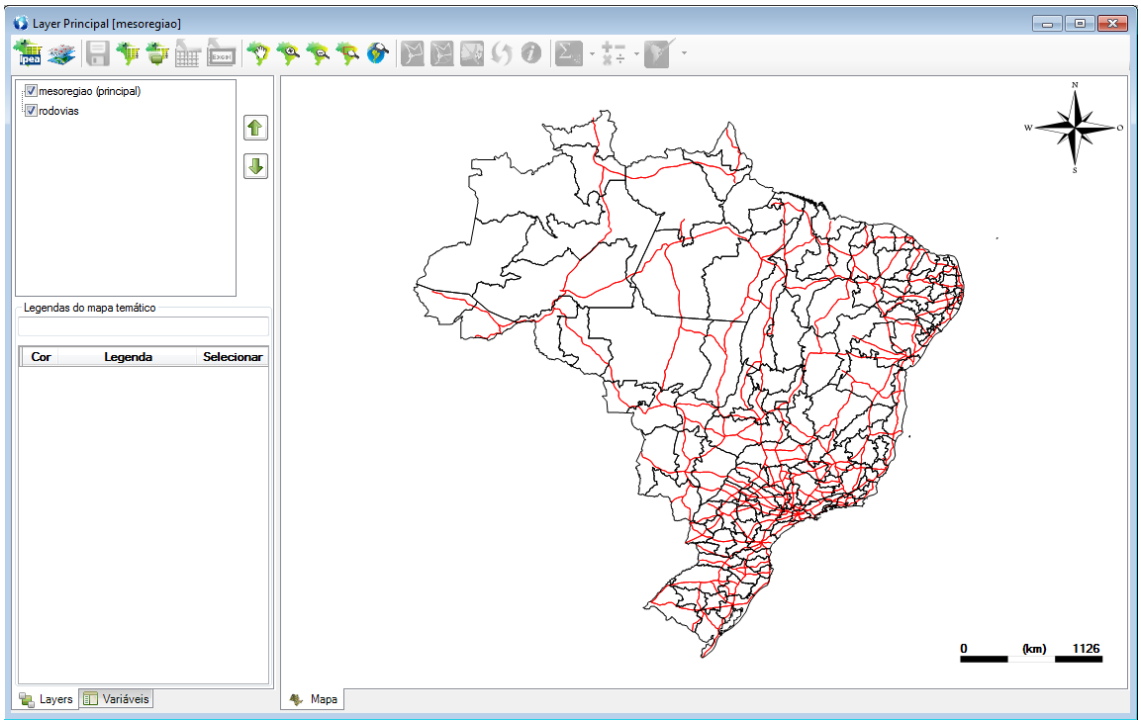


Em ambos os métodos para importação de malha digital auxiliar o seguinte formulário irá se abrir:



Selecione a pasta que contém a camada desejada, selecione o arquivo correspondente e clique em "*Abrir*". Somente é possível abrir uma camada auxiliar por vez; para abrir outras camadas é necessário repetir o mesmo procedimento.

Após abertas, as novas camadas serão sobrepostas à camada principal e os seus nomes aparecerão no lado esquerdo superior da janela. Caso deseje ocultar alguma das camadas, basta desmarcá-la clicando no marcador correspondente.



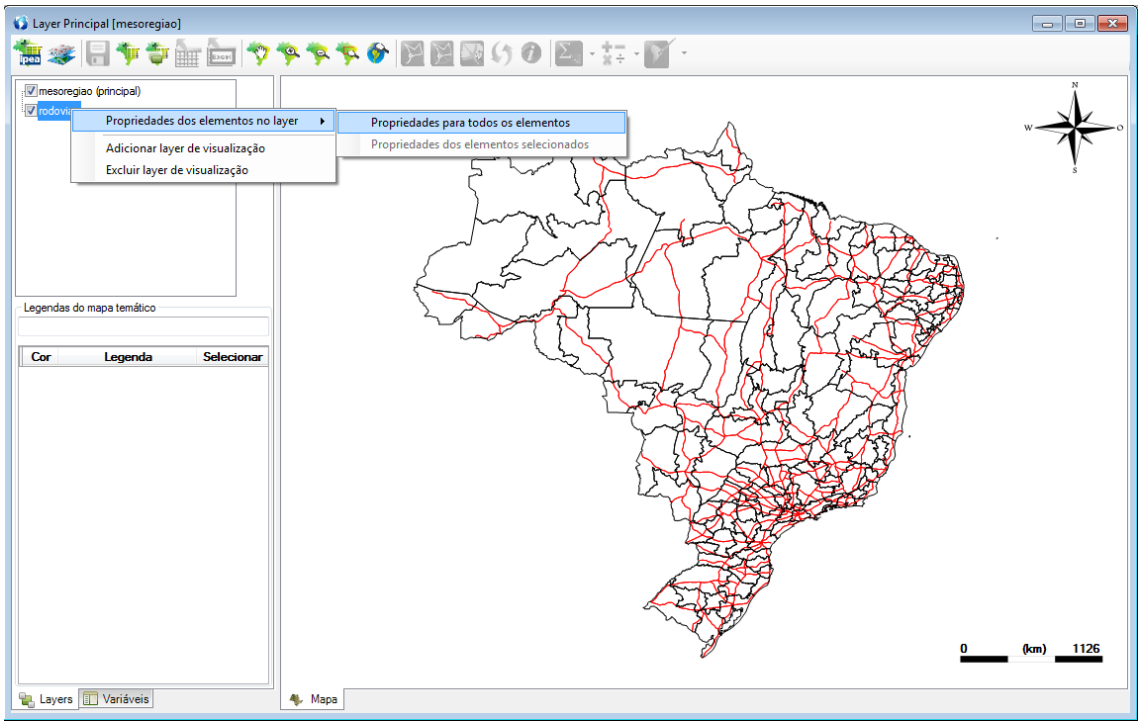
Alterando as Propriedades das Camadas (Layers)

Visão Geral

No IpeaGEO é possível alterar as propriedades das malhas digitais e das camadas auxiliares nas quais se está trabalhando.

Exemplo

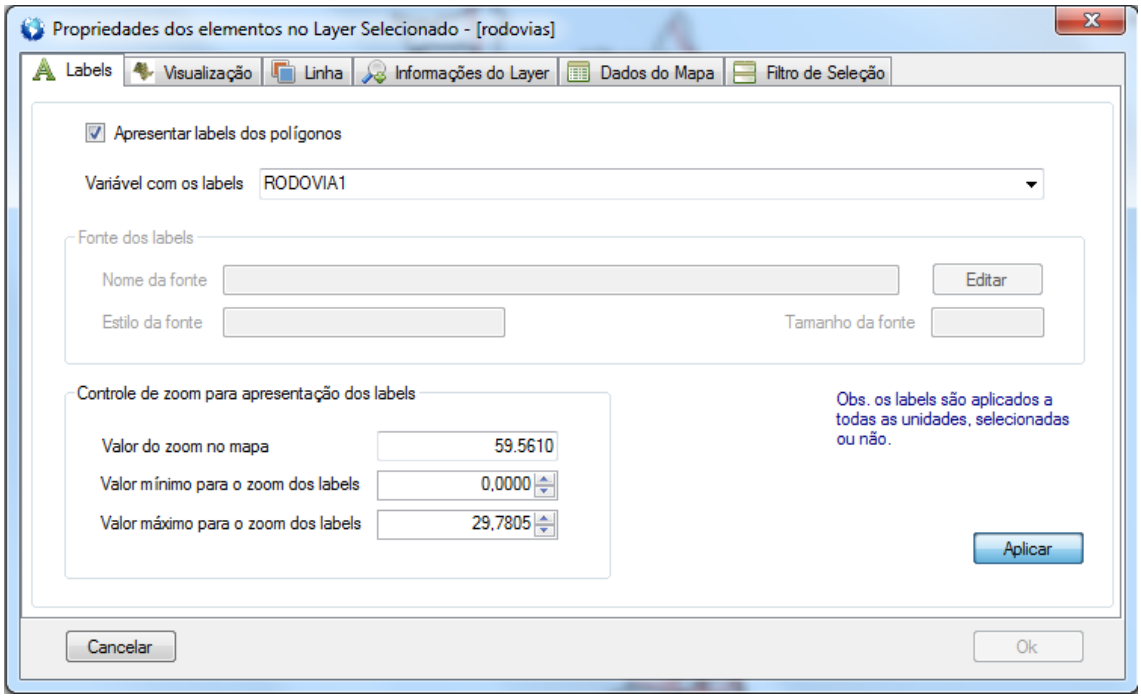
Após importar a malha digital por quaisquer das duas maneiras descritas em "Importar Malha" ou após adicionar as camadas auxiliares conforme descrito em "Adicionar *Layers* Auxiliares", é possível alterar as respectivas propriedades clicando sobre a camada correspondente com o botão direito do mouse, selecionando "Propriedades dos elementos do *layer*" e, em seguida, "Propriedades para todos os elementos".



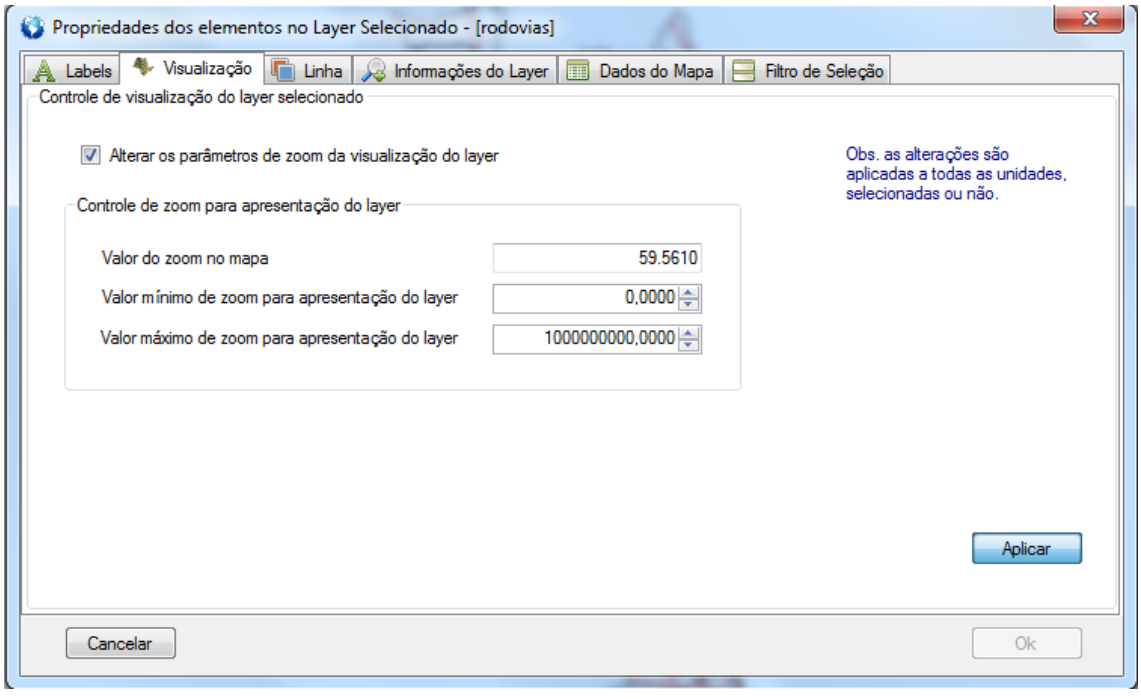
A seguir será exibido um formulário com as seguintes abas:

- Labels;
- Visualização;
- Linha;
- Informações do *Layer*;
- Dados do Mapa;
- Filtro de Seleção.

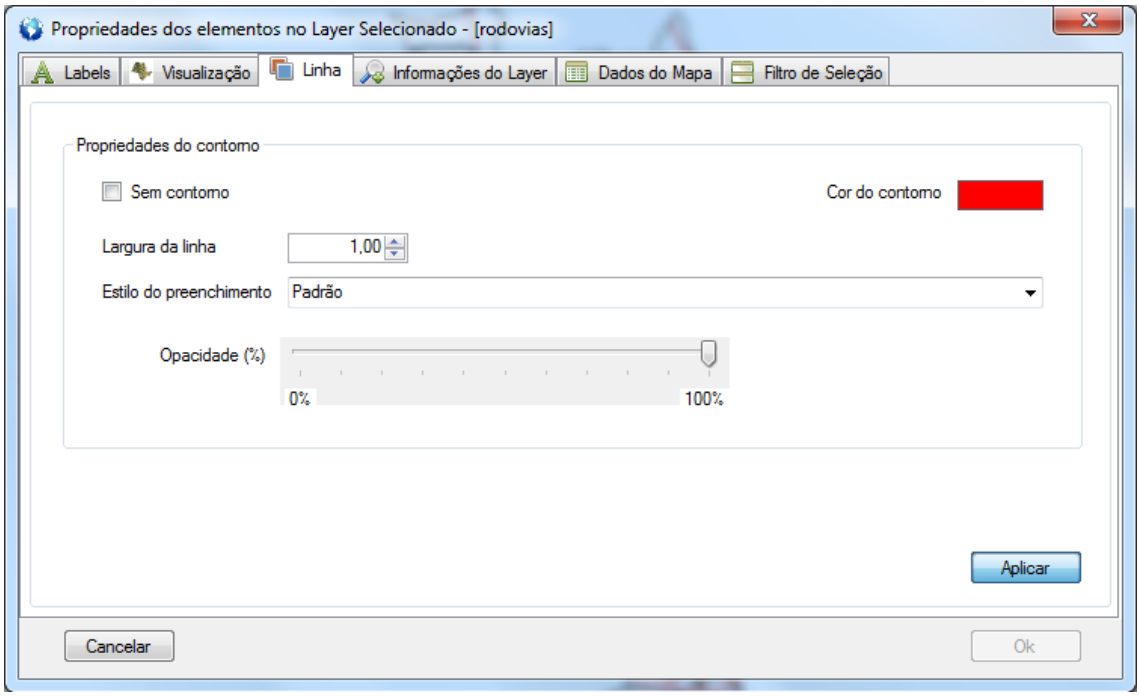
Na aba "Labels" é possível selecionar a opção "Apresentar labels dos polígonos" no mapa. Nesta aba também é possível alterar o tamanho, o tipo, a cor e o estilo da fonte usada para exibir os labels no mapa. Após definir os parâmetros clique em “Aplicar”.



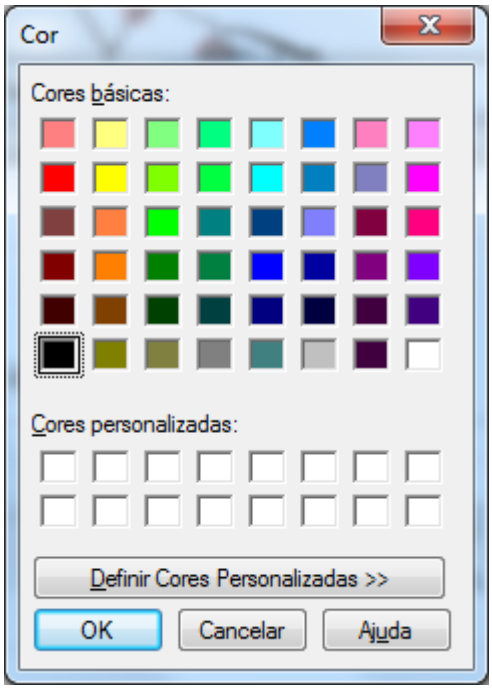
Na aba "Visualização" é permitido alterar os parâmetros de zoom do layer, bastando para isto selecionar a opção correspondente. É possível selecionar o valor atual, mínimo e máximo para o zoom na camada (layer) desejada. Para aplicar os parâmetros escolhidos, clique em “Aplicar”.



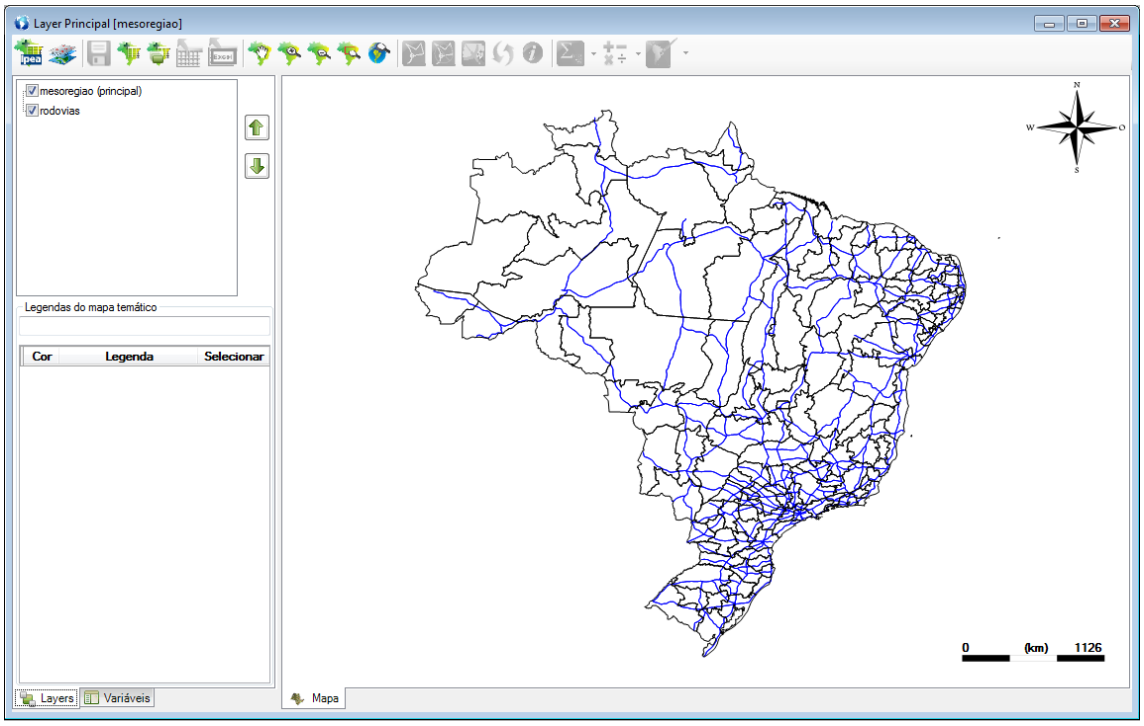
A aba "Linha" permite alterar as propriedades do contorno da camada tais como a cor, a largura da linha, opacidade e se o mesmo deve ser ocultado (para isto basta marcar a opção “Sem contorno”).



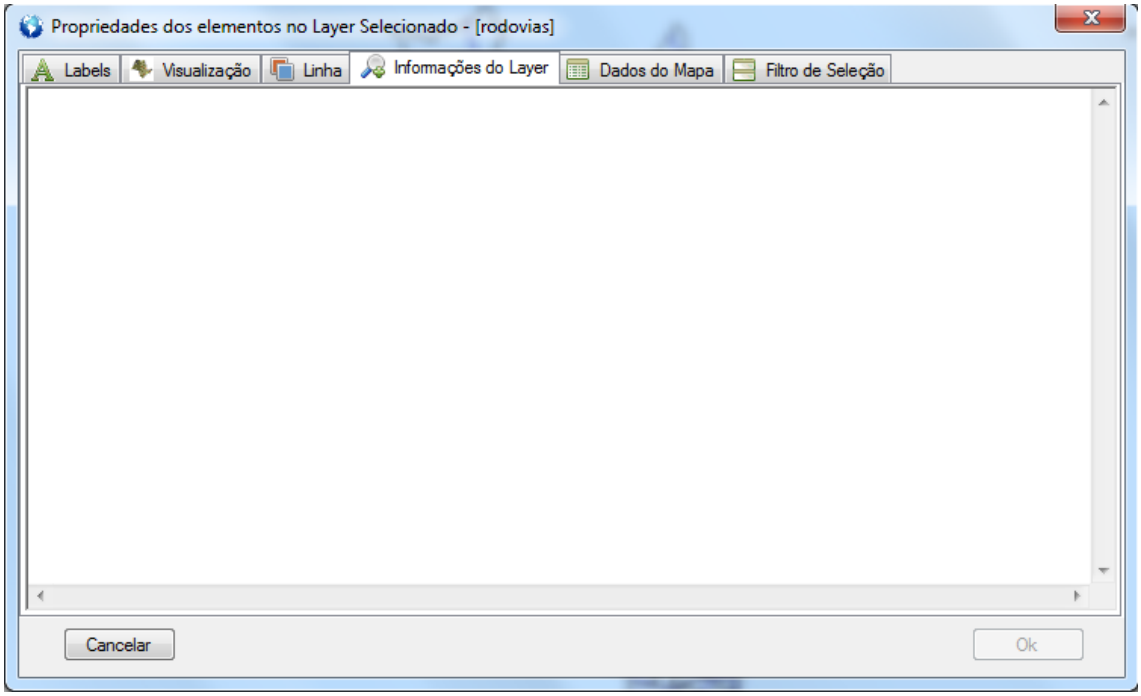
Para alterar a cor do contorno da camada, basta aplicar um duplo clique sobre a caixa de cores intitulada “Cor do contorno” para abrir o seguinte formulário:



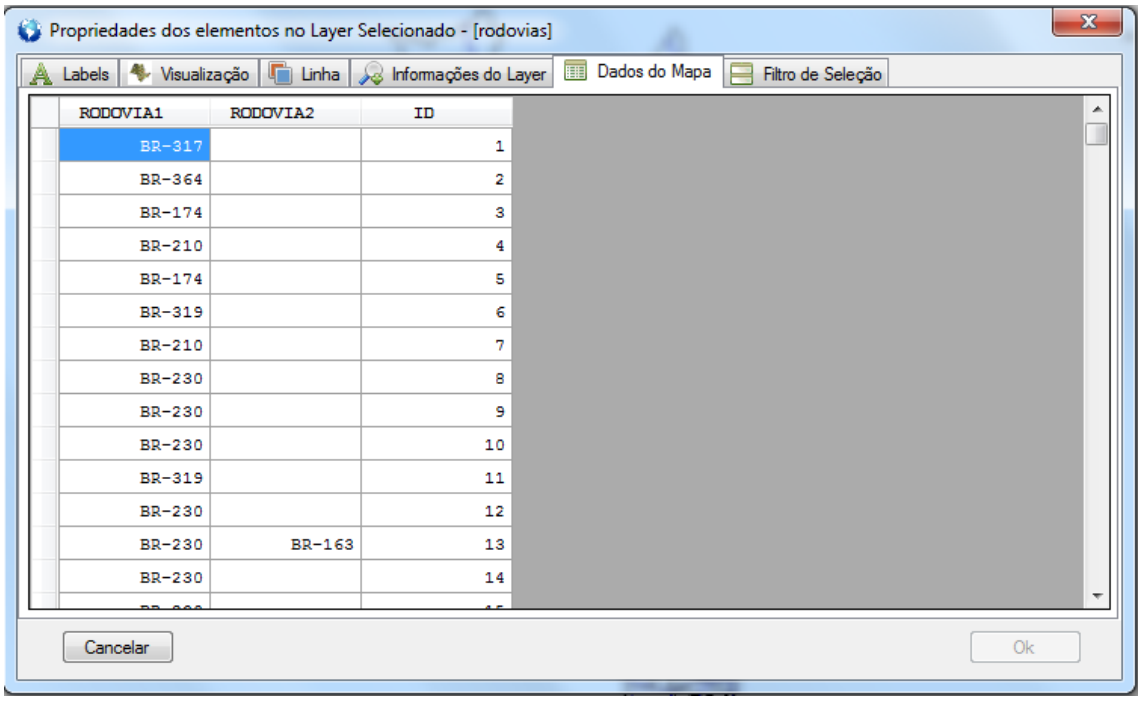
Após selecionar a cor desejada, clique em "OK" e em seguida em "Aplicar". No mapa abaixo, por exemplo, a camada das rodovias está representada em azul:



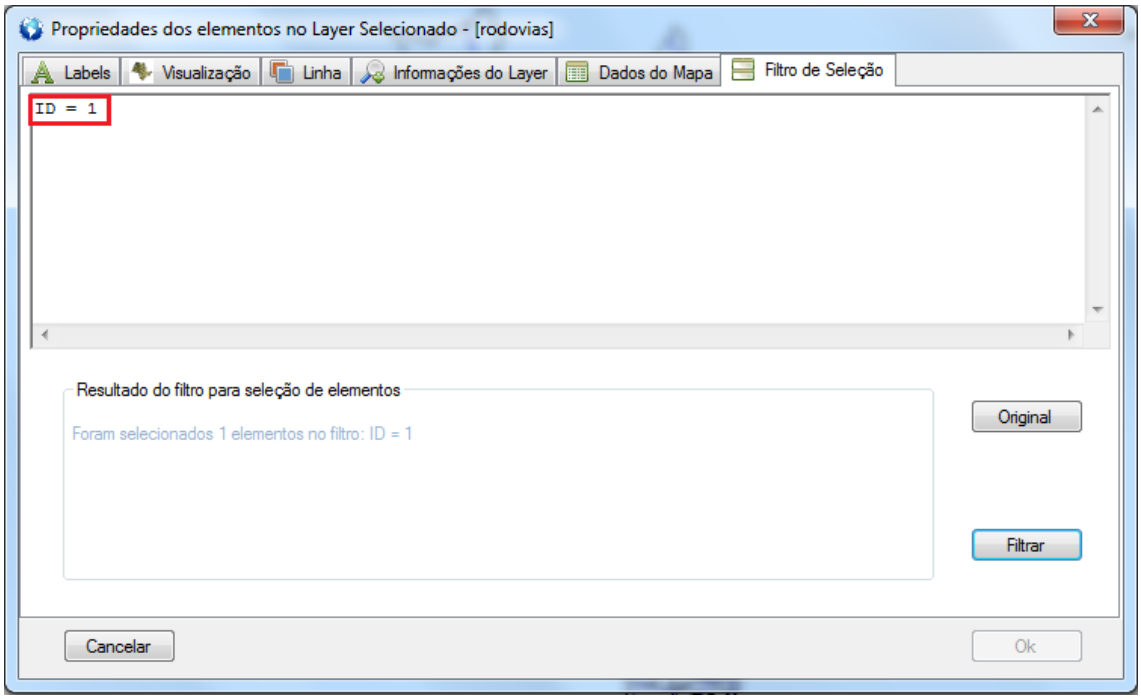
A aba "Informações do Layer" permite ao usuário escrever informações de seu interesse sobre a camada na qual se está trabalhando.



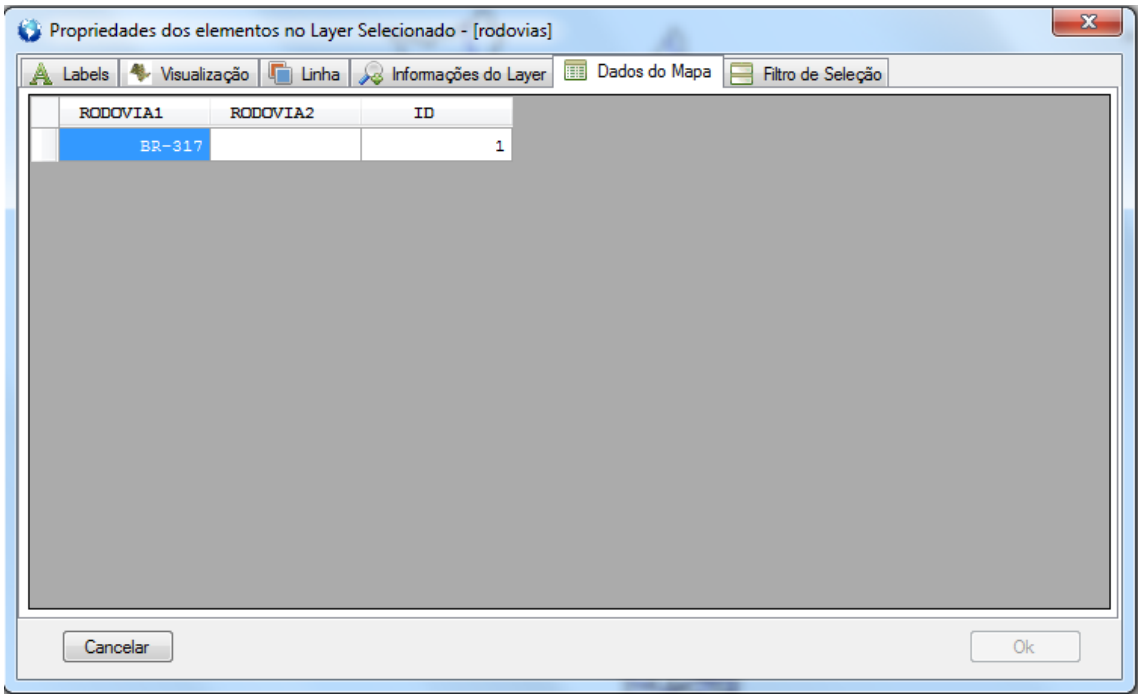
A aba "Dados do Mapa" mostra a base de dados contida na camada em questão.



A aba "Filtro de Seleção" permite filtrar os elementos da camada na qual se está trabalhando. Para isso, digite a expressão que seleciona os elementos desejados e clique em "Filtrar". O quadro "Resultado do filtro para seleção dos elementos" exibirá o resultado da aplicação do filtro.



Para visualizar os dados após a aplicação do filtro, clique na aba "Dados do Mapa".



Base de Dados

Importar Dados

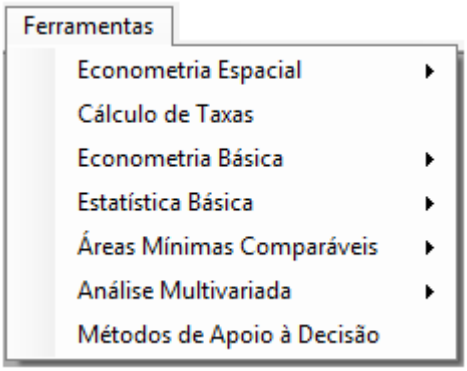
Visão Geral

No IpeaGEO, o usuário pode importar sua tabela de dados para trabalhar conjuntamente com um arquivo de malha digital ou ainda, caso prefira, pode realizar análises somente com sua tabela de dados.

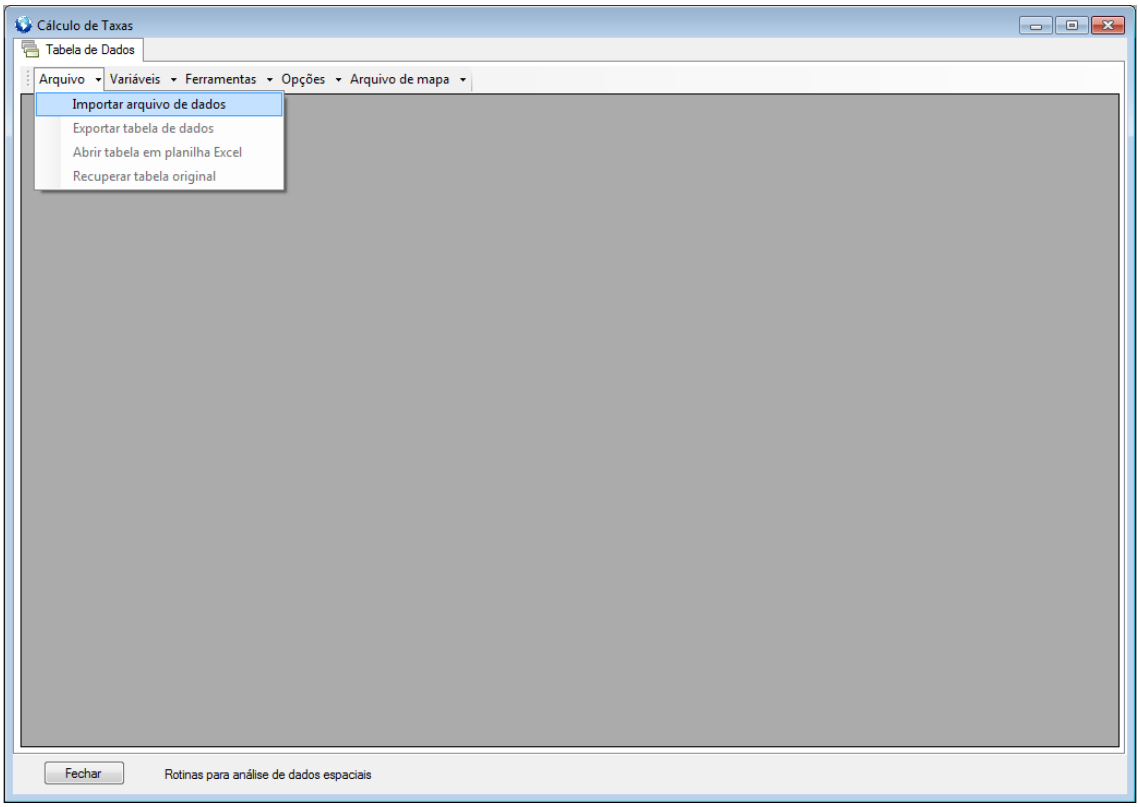
Exemplos

Importar somente a tabela de dados

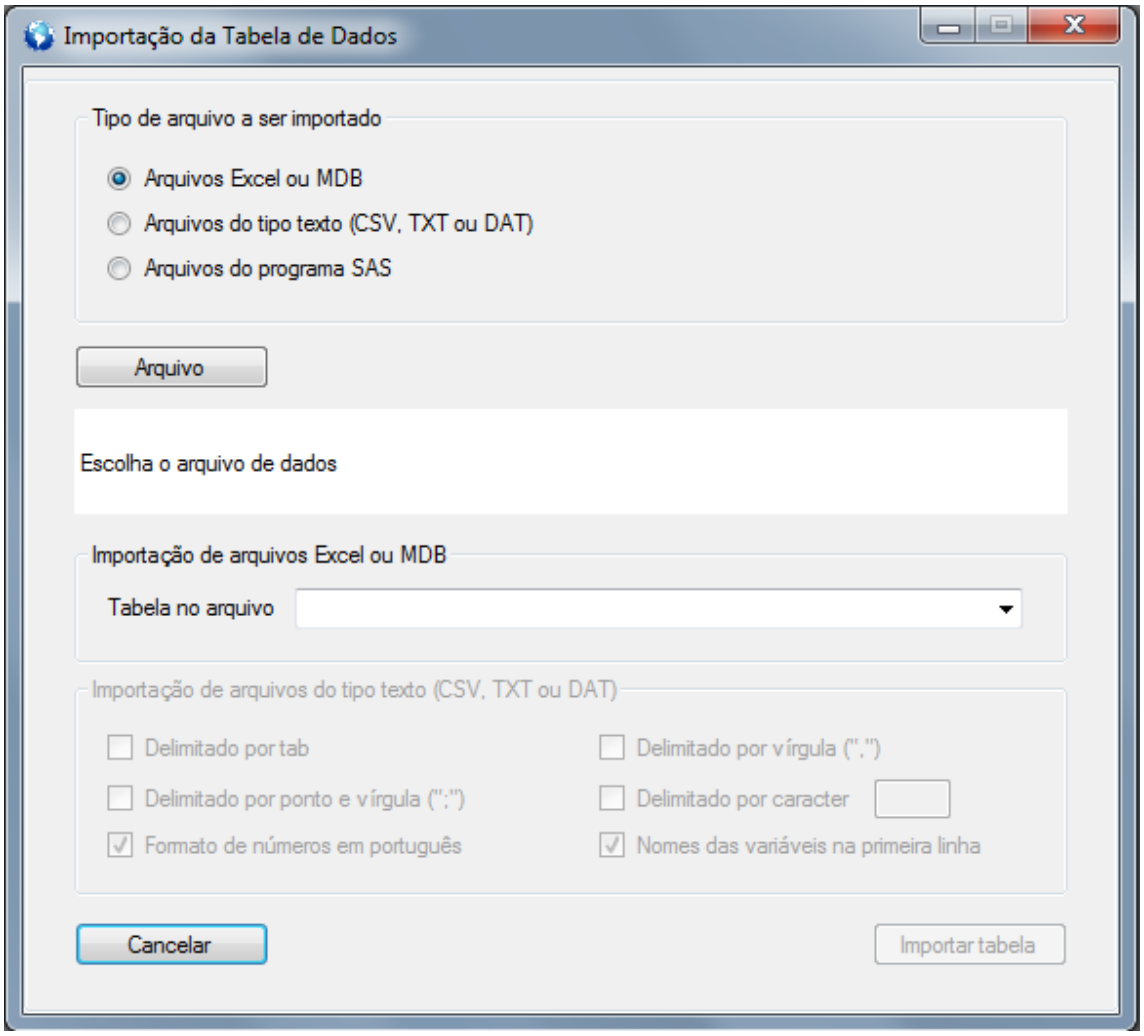
Caso o interesse seja realizar análises sem um arquivo de malha digital, inicialmente, o usuário deve escolher qual tipo de análise deseja realizar por meio do menu "Ferramentas":



Após selecionar a análise de interesse, clique em "Arquivo" e em "Importar arquivo de dados".



Escolha o formato do arquivo a ser importado dentre os disponíveis: Microsoft Excel (XLS), Microsoft Access (MDB), texto separado por vírgulas (CSV), texto separado por tabulações ou espaços (TXT ou DAT) ou tabela de dados do SAS (SAS7BDAT). Em seguida clique em “Arquivo”, selecione o arquivo desejado e em seguida clique em "Importar Tabela".



Irá abrir, após esse ultimo passo, o formulário da ferramenta de análise escolhida no início.

Importar uma tabela de dados para concatenar com uma malha (*shape*)

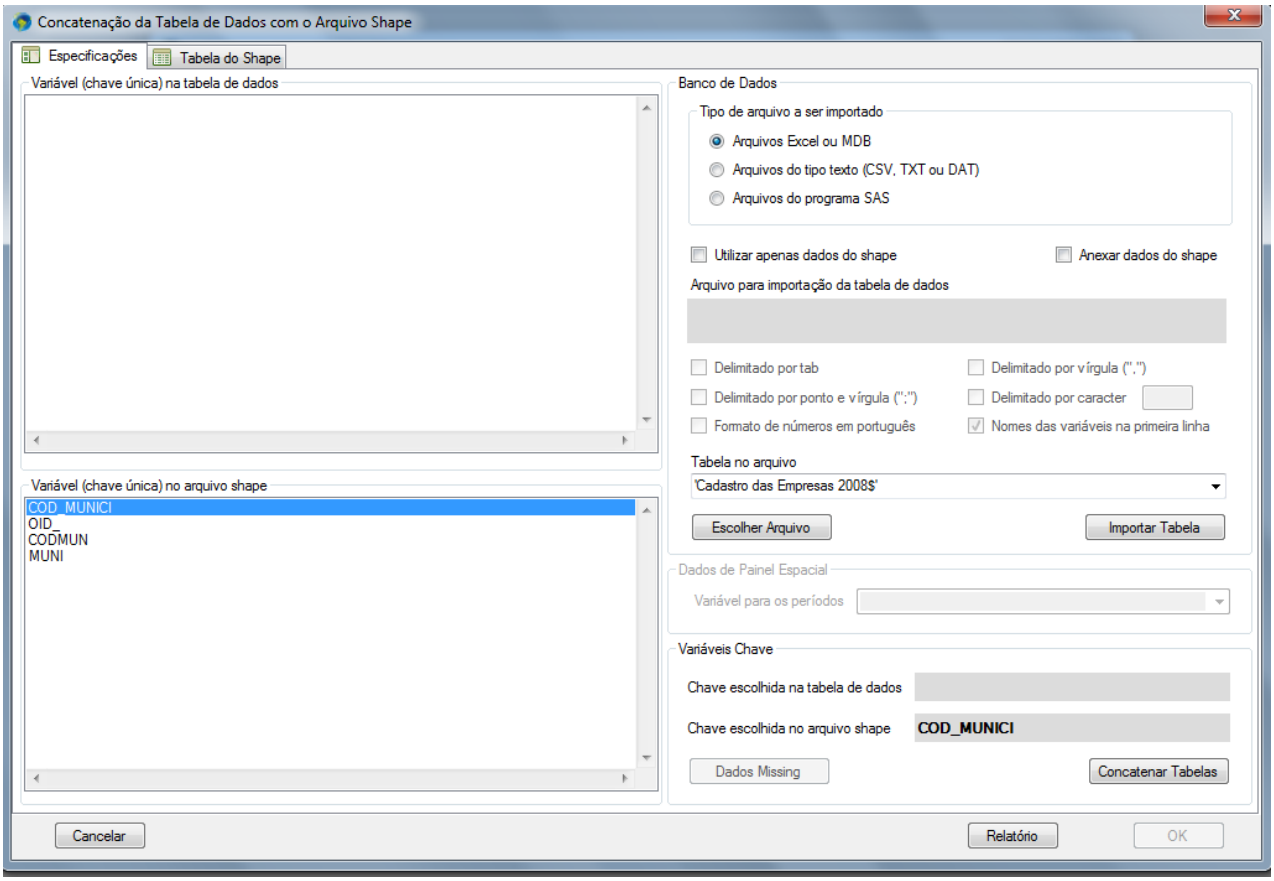
Para importar a tabela de dados e relacioná-la com um arquivo shape existem duas opções, relacionadas com a procedência da malha digital (para ver como abrir uma malha, vide página 9).

Importar uma tabela de dados própria

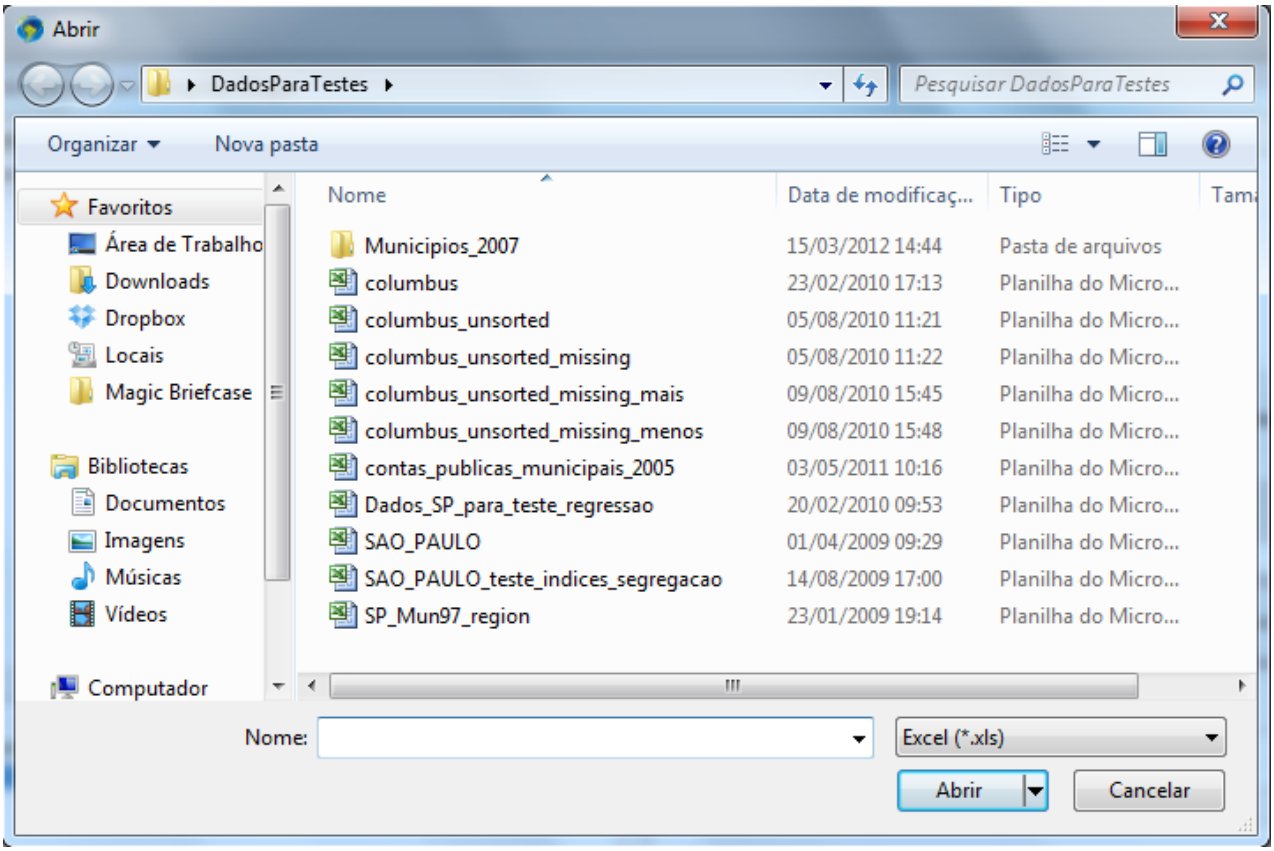
Independentemente da procedência da malha digital (arquivo externo ou IpeaGEO) é permitido ao usuário importar uma tabela de dados própria.

Para isso, com a malha digital já aberta, o usuário deve clicar no ícone

denominado "Adicionar tabela a partir de dados externos" . Neste momento o seguinte formulário se abre:



Para definir qual tabela se deseja importar, clique em "Escolher Arquivo". O seguinte formulário irá se abrir:

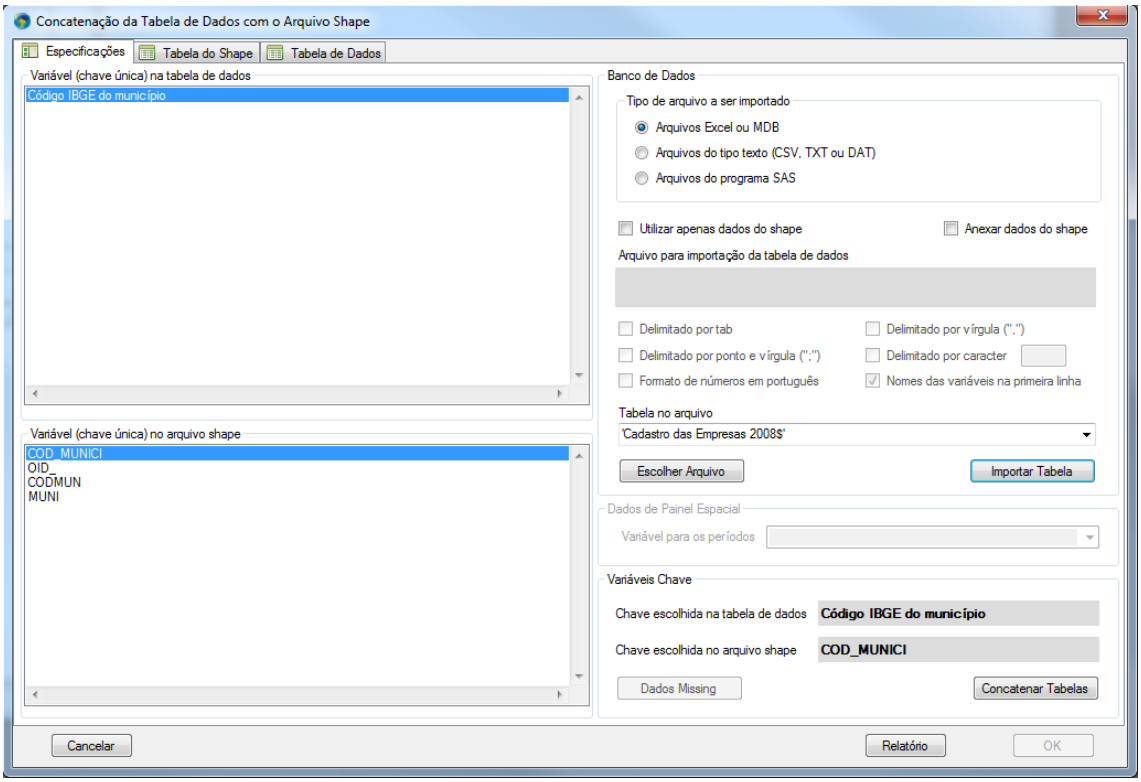


Selecione o arquivo desejado e clique em "Abrir".

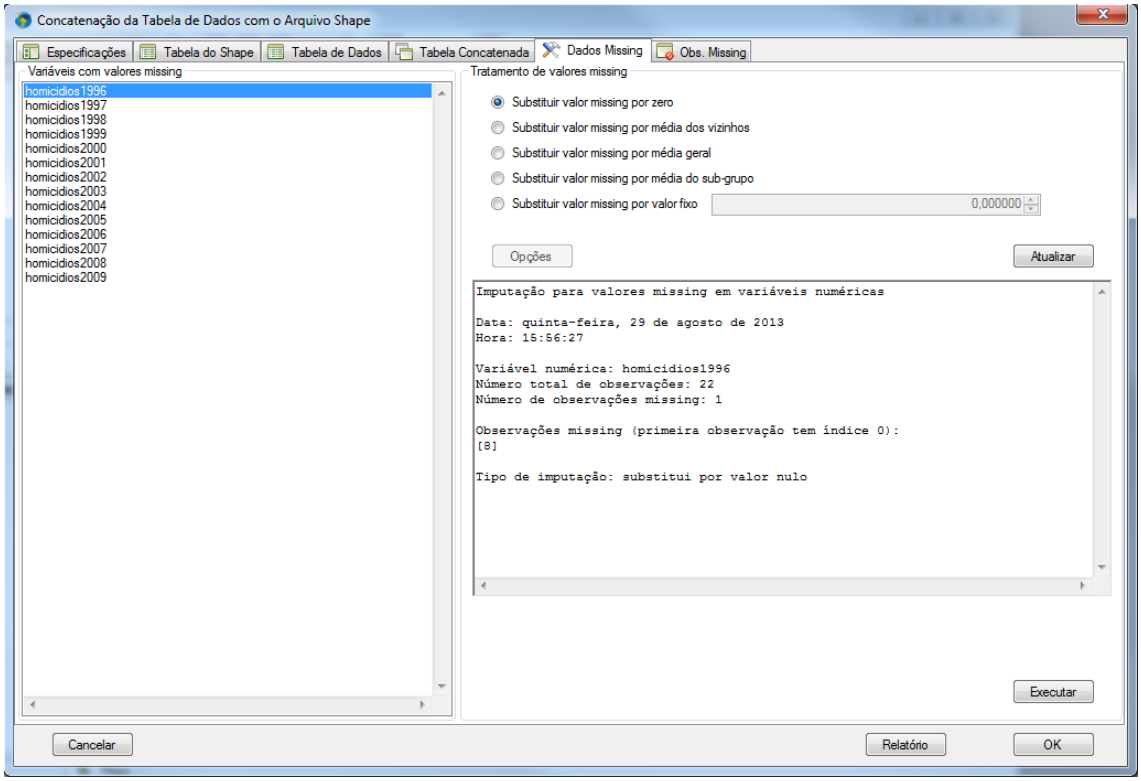
Após importar a tabela, é necessário definir as variáveis chaves para o cruzamento entre a malha e o arquivo de dados, isto é, uma variável comum entre a tabela importada e o arquivo de malha digital que permita que os dois arquivos sejam concatenados. Para isso selecione uma variável de cada quadro localizado ao lado esquerdo do formulário.

Repare que as variáveis selecionadas aparecerão no canto inferior direito; em seguida clique em "Concatenar Tabelas" e "OK".

É importante ressaltar que, para que a concatenação seja concluída com sucesso, as variáveis a serem utilizadas como chaves precisam ter o mesmo tamanho e formato.

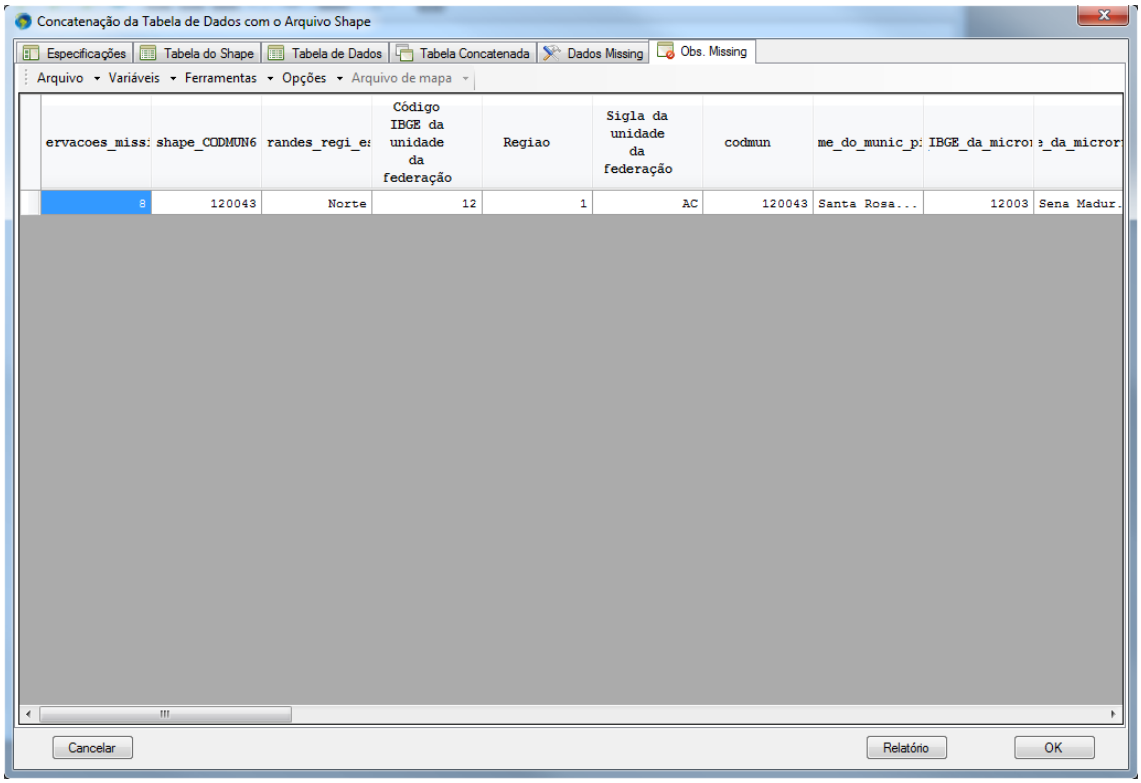


Caso a tabela contenha dados faltando (dados *missing*), irão aparecer duas novas abas: “Dados Missing” e “Obs. Missing”:

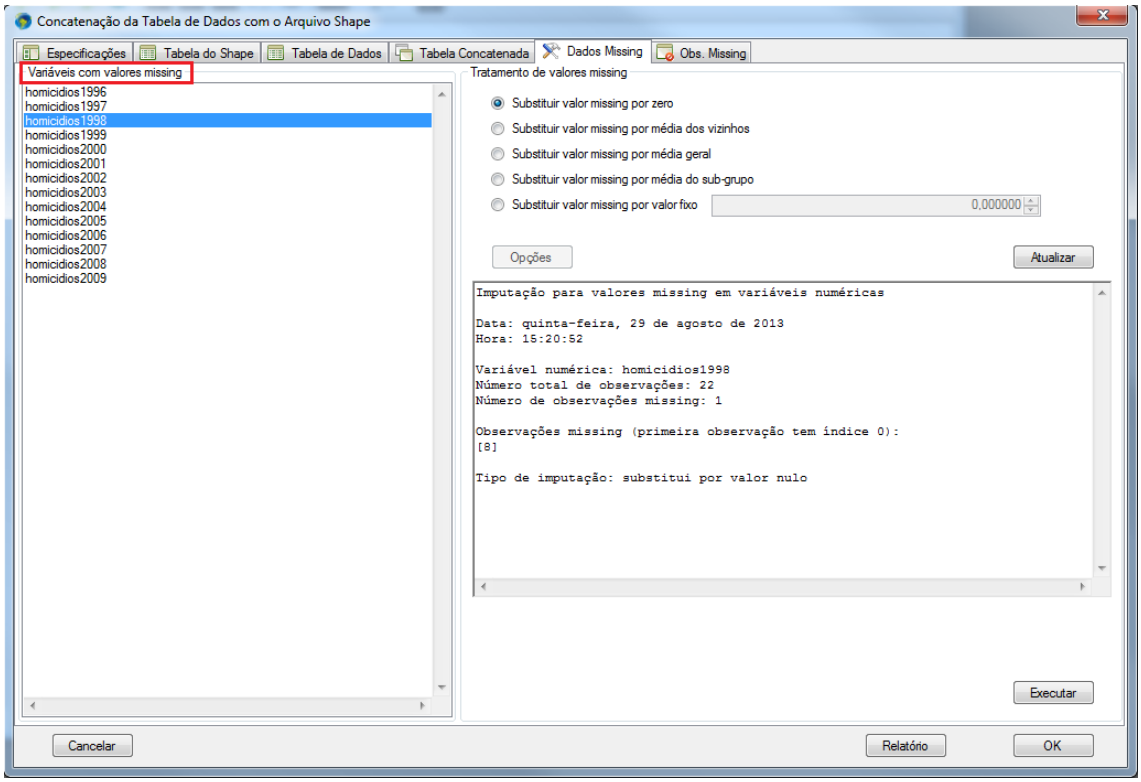


Neste formulário estão disponíveis as opções: “Variáveis com valores missing” e “Tratamento de valores missing”.

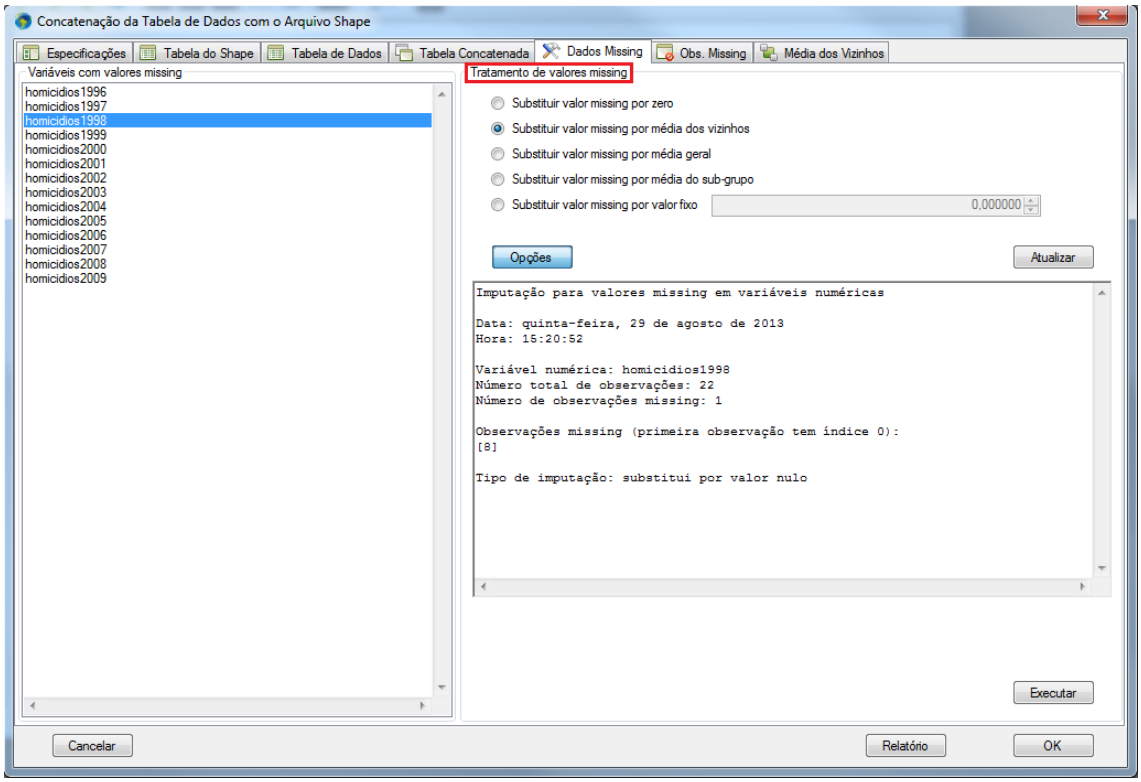
Na aba “Obs. Missing” se encontram as informações referentes às variáveis sem dados.



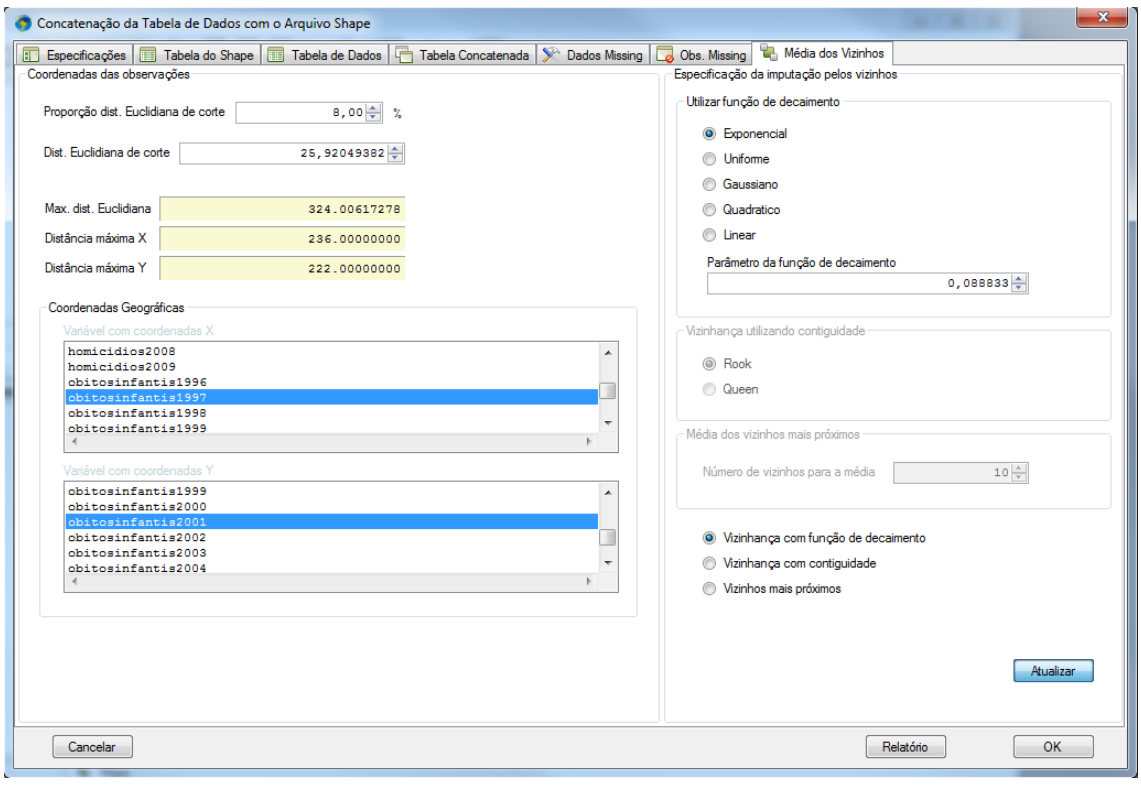
No formulário abaixo, selecione as variáveis que não possuem os valores desejados.



Em seguida, selecione o tipo de tratamento que deseja aplicar. As opções “Substituir valor *missing* por média dos vizinhos” e “Substituir valor missing por média do subgrupo” habilitam o botão “Opções”, como indicado na figura abaixo.

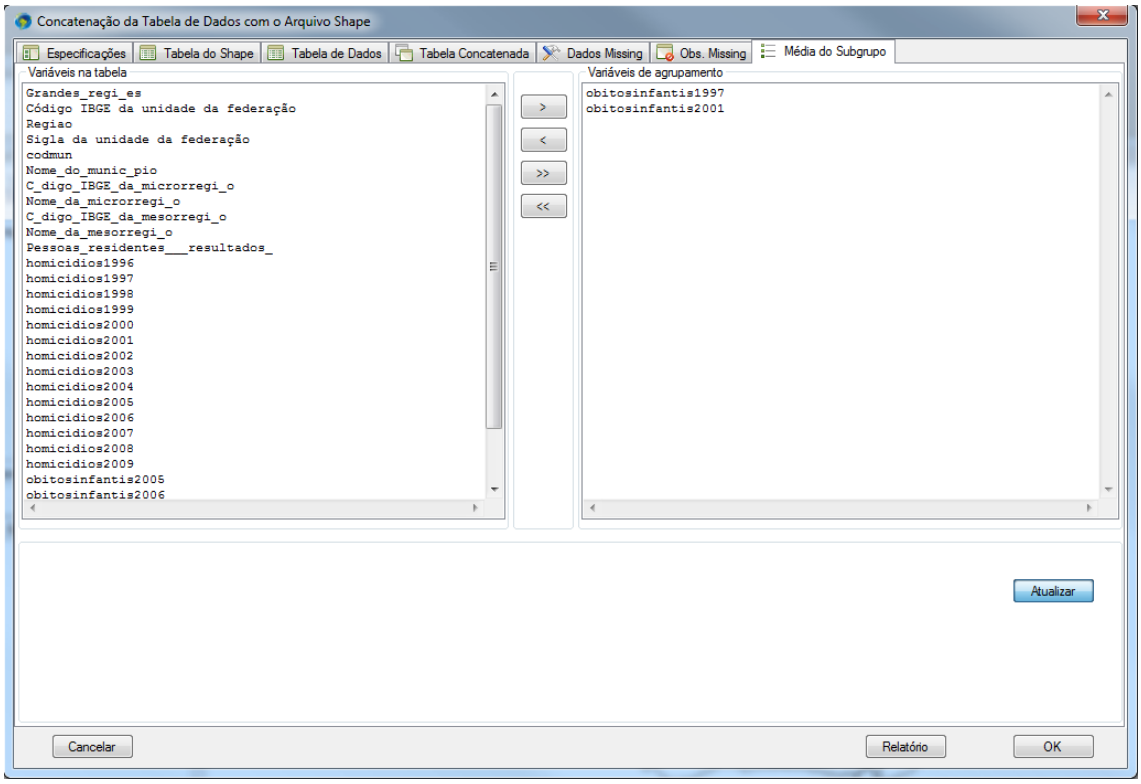


Caso a opção desejada seja “Substituir valor *missing* por média dos vizinhos”, o seguinte formulário contido na aba “Média dos vizinhos” irá se abrir.



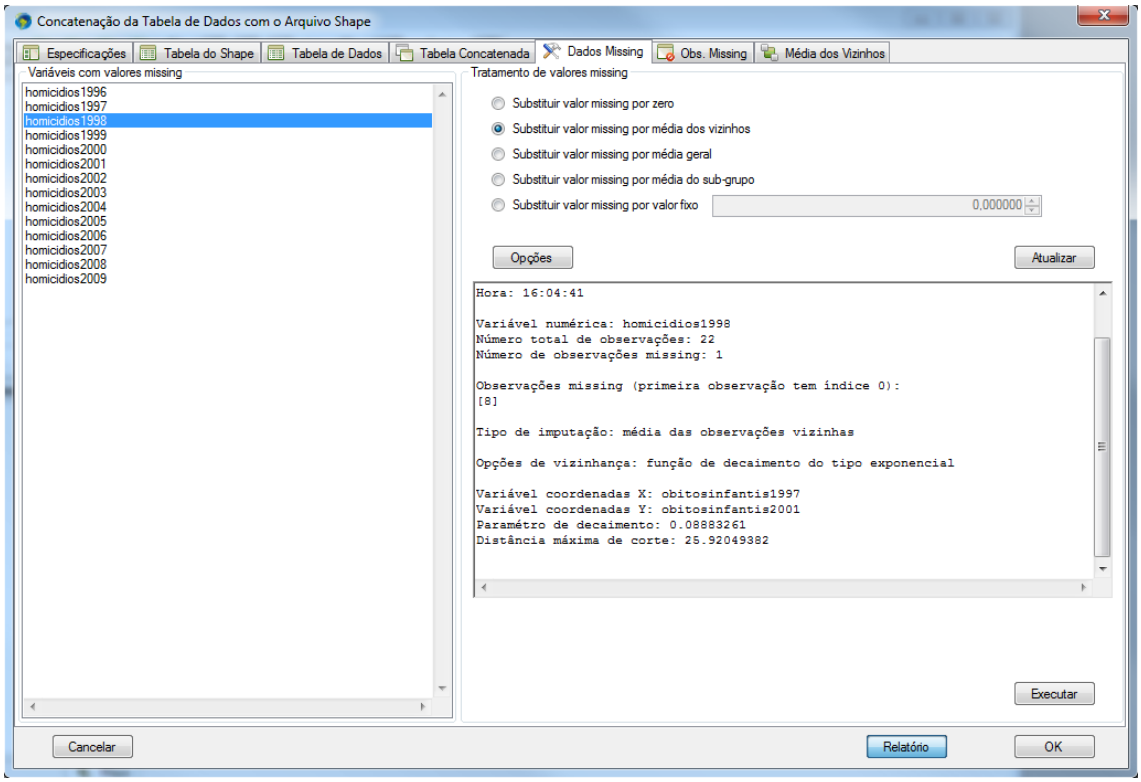
Neste formulário, após definir as opções para manipulação das variáveis que contenham dados *missing*, clique no botão “Atualizar”.

Caso a opção desejada seja “Substituir valor *missing* por média do sub-grupo”, o seguinte formulário, contido na aba “Média do Subgrupo” irá se abrir.

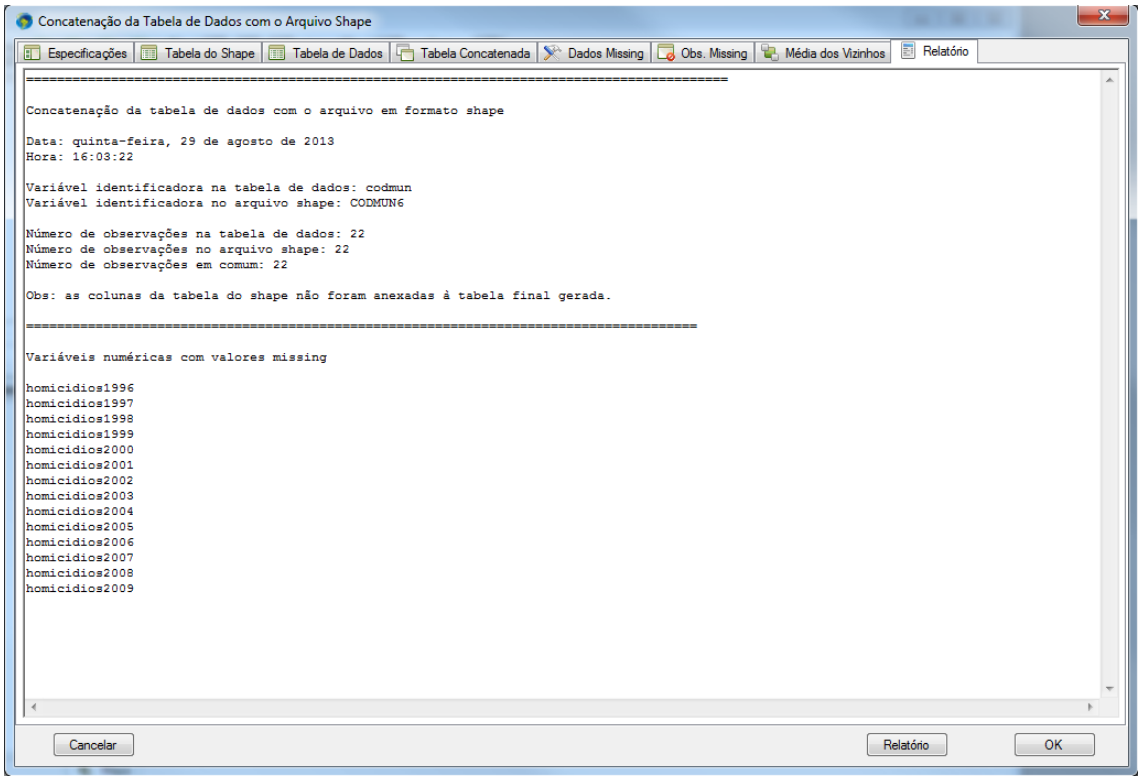


Esse formulário permite ao usuário definir as variáveis a serem utilizadas como “Variáveis de agrupamento”, escolhidas a partir das “Variáveis na tabela”. Após selecionadas as opções desejadas, clique em “Atualizar”.

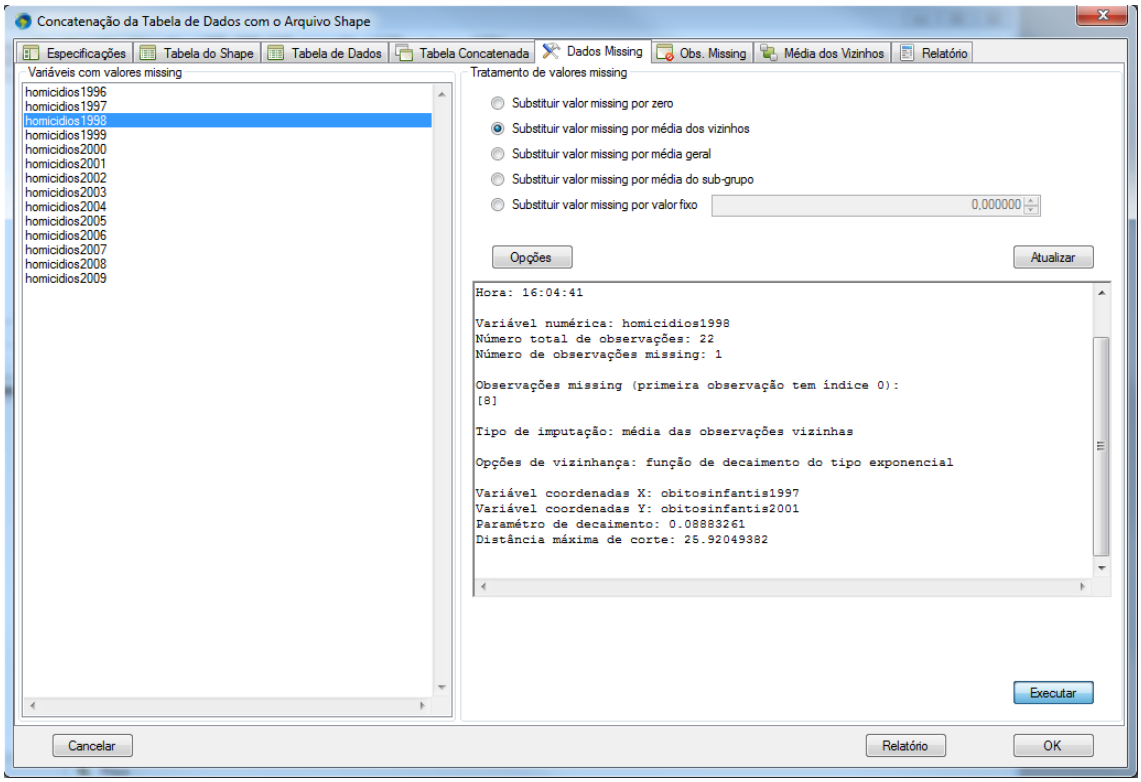
O quadro localizado na aba “Dados *Missing*” indicará as atualizações realizadas.



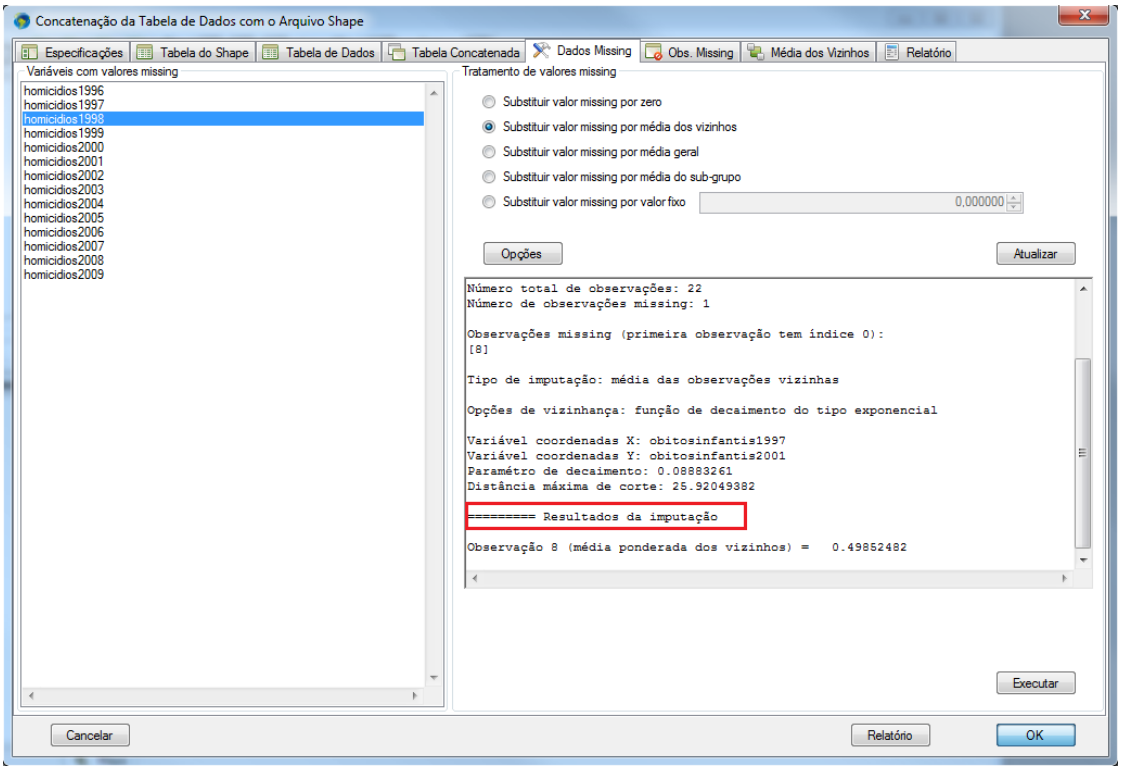
Ao clicar no botão “Relatório” uma nova aba de mesmo nome se abrirá, contendo as informações definidas anteriormente.



Por fim, clique no botão “Atualizar” para atualizar o relatório de acordo com o tratamento escolhido e em seguida clique em “Executar”.



Na caixa de texto do formulário aparecerá uma nova seção chamada “Resultados da imputação” com o resultado final obtido. Ao término, clique no botão “OK” para voltar ao formulário contendo o mapa da região escolhida.



Importar uma tabela do IpeaGEO

Caso a malha digital seja proveniente do IpeaGEO, também é permitido importar uma tabela de dados que acompanha o programa.

Para isso clique no ícone denominado "Adicionar tabela a partir de dados do IpeaGEO" . O formulário descrito no tópico anterior irá se abrir.

O próximo passo é definir qual a tabela que se deseja importar. Para isso, escolha uma dentre as tabelas disponíveis em "Tabela no Arquivo" e clique em "Importar Tabela". Após a importação, é necessário selecionar as variáveis chaves e realizar o mesmo procedimento descrito no tópico anterior.

Exportar Dados

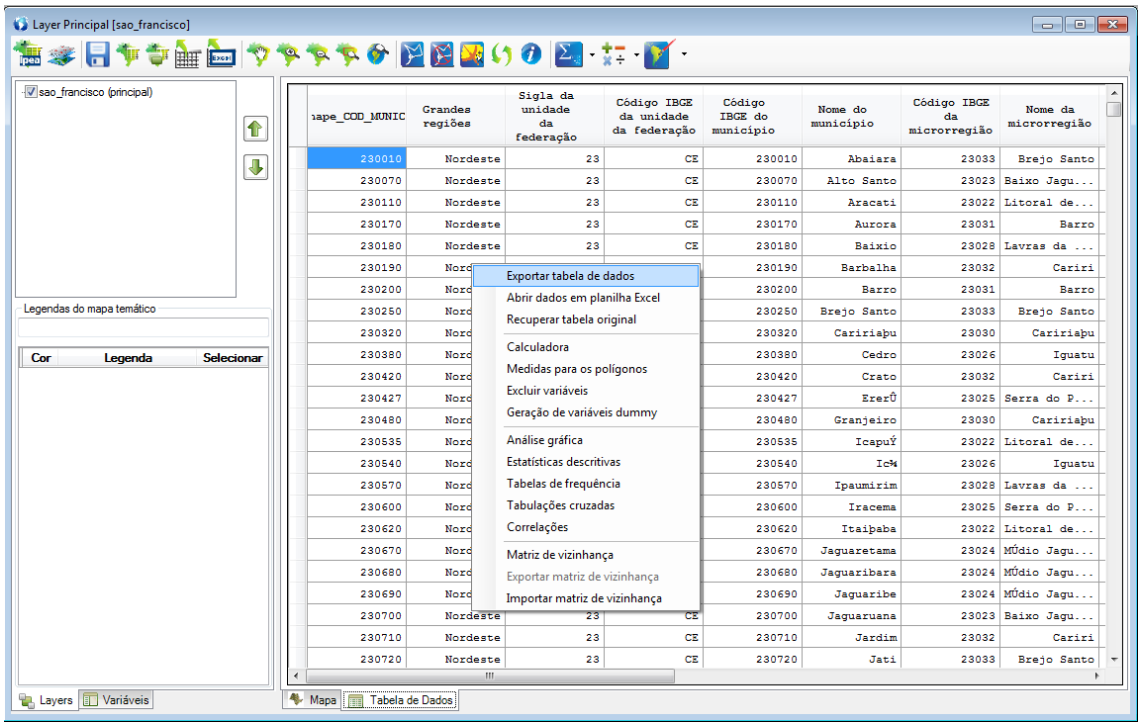
Visão Geral

O IpeaGEO oferece a opção de exportar a tabela de dados que está se trabalhando a qualquer momento, seja após alterar sua base original ou realizar os resultados de uma análise estatística nessa base.

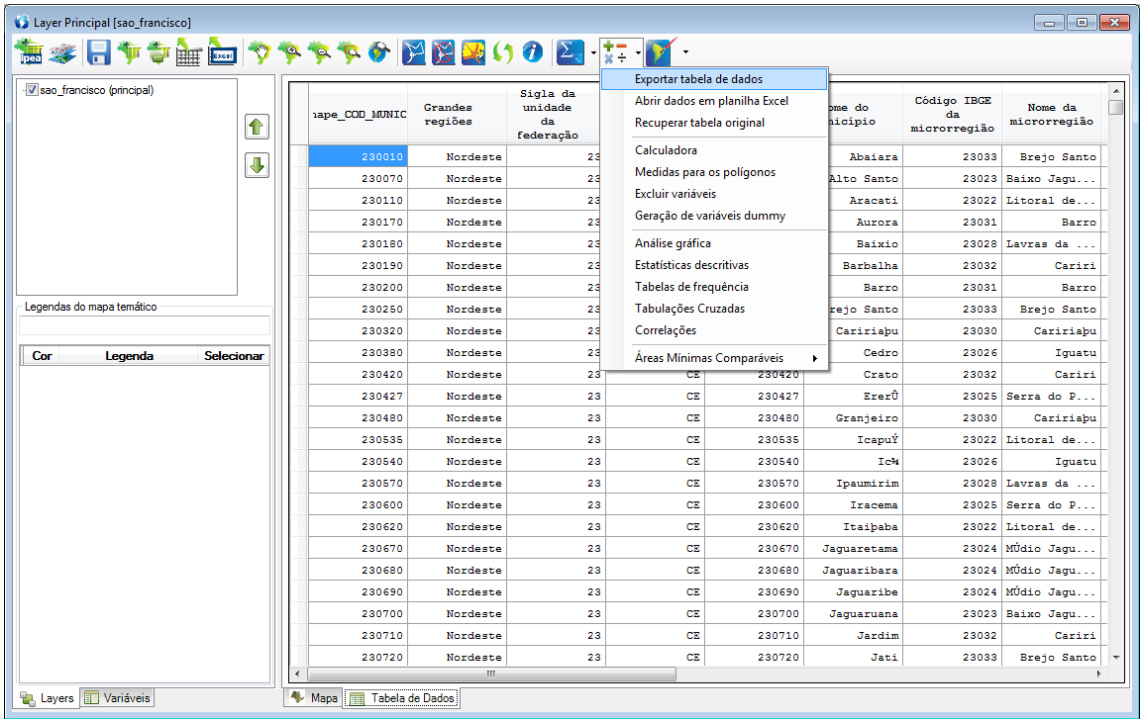
Exemplos

É possível exportar a base de dados de duas maneiras:

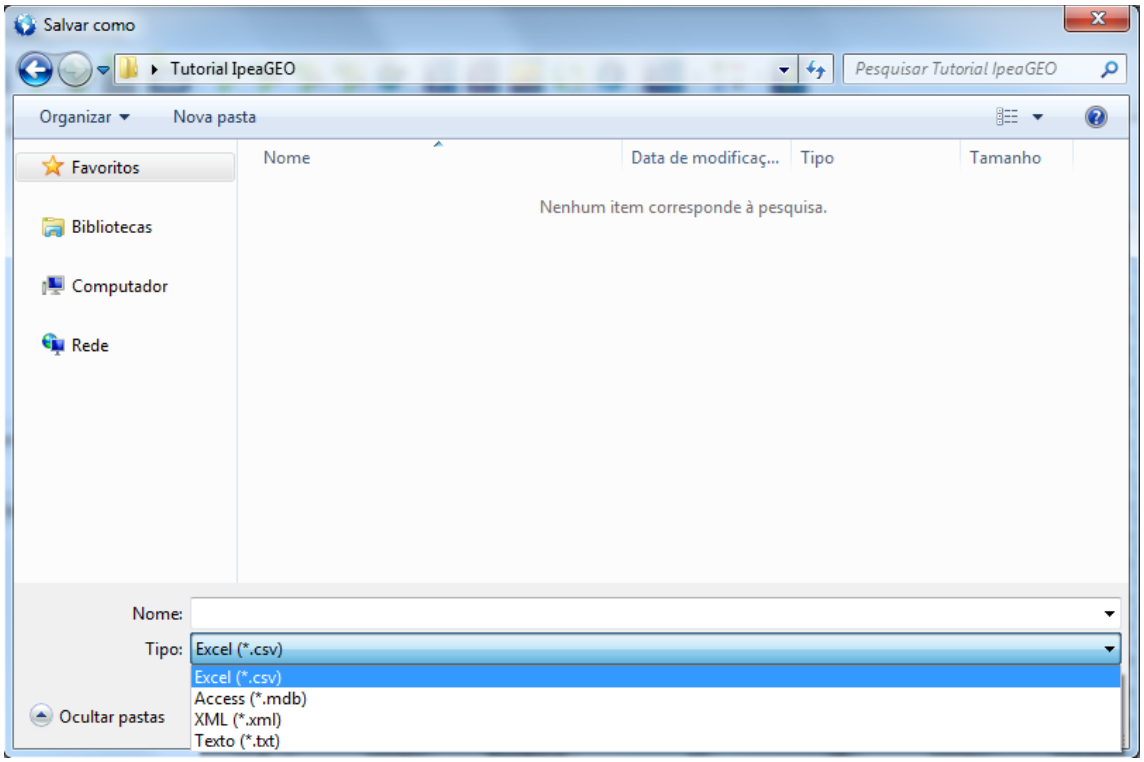
Opção 1: Clique com o botão direito do mouse sobre a tabela e, em seguida, em "Exportar tabela de dados".



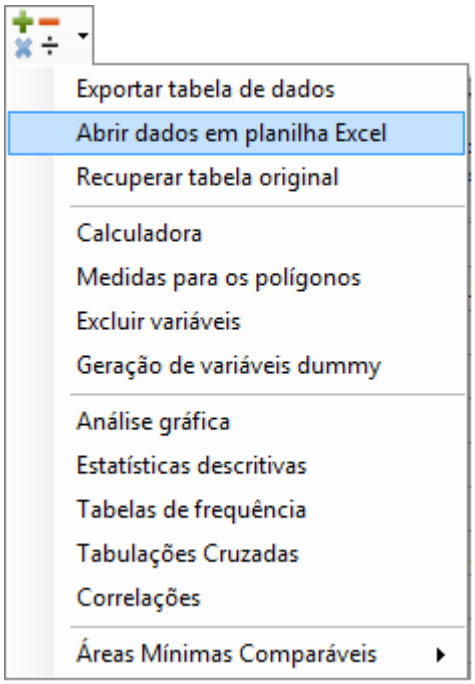
Opção 2: Clique no botão "Análise Exploratória e Manipulação de Dados" e, em seguida, em "Exportar tabela de dados"



O passo seguinte é escolher o local no qual os dados serão salvos e qual o formato do arquivo. O IpeaGEO permite exportar os dados nos formatos Microsoft Excel (.csv), Microsoft Access (.mdb), .xml ou texto (.txt).



Outra opção disponível é visualizar a tabela no formato Microsoft Excel sem a necessidade de exportá-la. Para isso, basta seguir uma das opções descritas acima e, em seguida, clicar em "Abrir dados em planilha Excel", disponível no ícone .



Funcionalidades

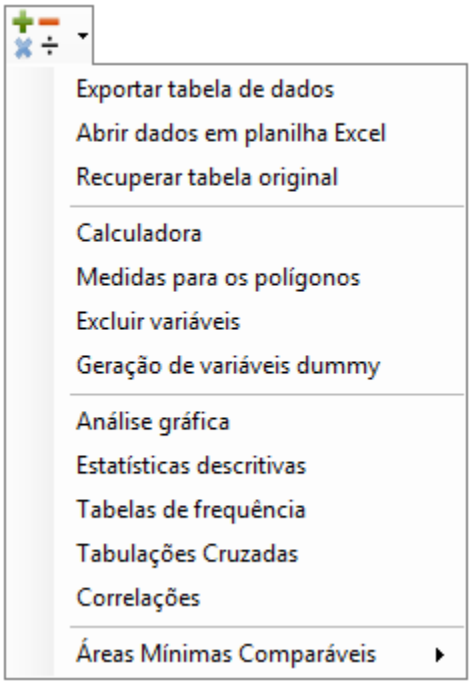
Visão Geral

Após o usuário abrir um arquivo de malha digital e concatenar com uma tabela de dados, os seguintes ícones são habilitados:

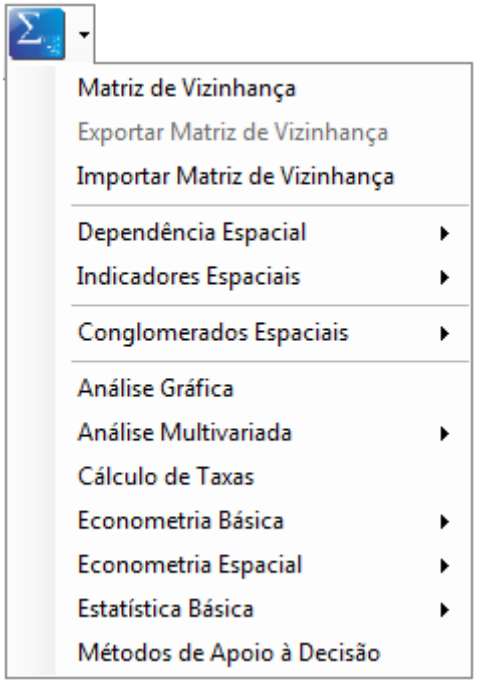
Os significados e funcionalidades de cada ícone/atalho são:

Ícone	Descrição
	Adicionar uma tabela a partir de dados já existentes no IpeaGEO.
	Adições de layers auxiliares ao mapa – Adiciona camadas (layers) que podem ser sobrepostos ao mapa.

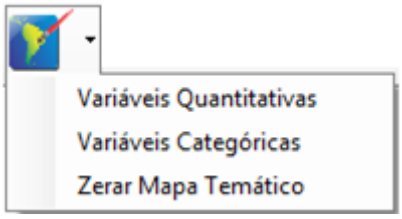
	Salvar Projeto.
	Adicionar uma tabela a partir de dados externos.
	Inserir dados em um painel espacial.
	Exportar tabela de dados – Exporta a tabela de dados para os formatos Microsoft Excel (*.xls), Microsoft Access (*.mdb), XML (*.xml) e texto (*.txt).
	Abrir tabela de dados em planilha Excel - Abre automaticamente a planilha de dados no Excel.
	Atualizar tabela e mapa.
	Mover o mapa – Permite mover o mapa a todas as direções.
	Adiciona Zoom – Aplica zoom no mapa inteiro.
	Remove Zoom – Remove zoom no mapa inteiro.
	Zoom – Aplica zoom na área selecionada do mapa.
	Padrão – Retorna o mapa ao formato original quanto ao zoom.
	Seleciona em conjunto – Permite selecionar áreas do mapa conjuntamente.
	Remove seleção – Retorna o mapa ao formato original quanto à seleção.
	Seleciona individualmente – Permite selecionar áreas do mapa individualmente.
	Informações sobre o polígono – Retorna todas as informações presentes na tabela concatenada referente ao polígono clicado.



Análise Exploratória e Manipulação de Dados – Possibilita a seleção de ferramentas e operações estatísticas a serem aplicadas aos dados concatenados.



Ferramentas – Ferramentas estatísticas que podem ser aplicadas aos dados concatenados.



Mapa Temático – Permite a geração de mapas de acordo com o interesse do usuário.

Mapa Temático

Visão Geral

Mapas temáticos são mapas personalizados que têm como objetivo identificar a distribuição espacial de eventos físicos, políticos, culturais, econômicos, sociológicos ou qualquer outro aspecto que se deseja investigar, de uma determinada área geográfica de interesse, como, por exemplo, municípios, microrregiões, unidades de federação, etc.

O mapa é dividido em classes, não necessariamente com o mesmo tamanho e amplitude, no qual cada classe possui a sua cor.

Os métodos disponíveis para gerar os mapas temáticos são:

Quantil*:* Intervalos são selecionados de modo que o número de observações em cada intervalo é igual.

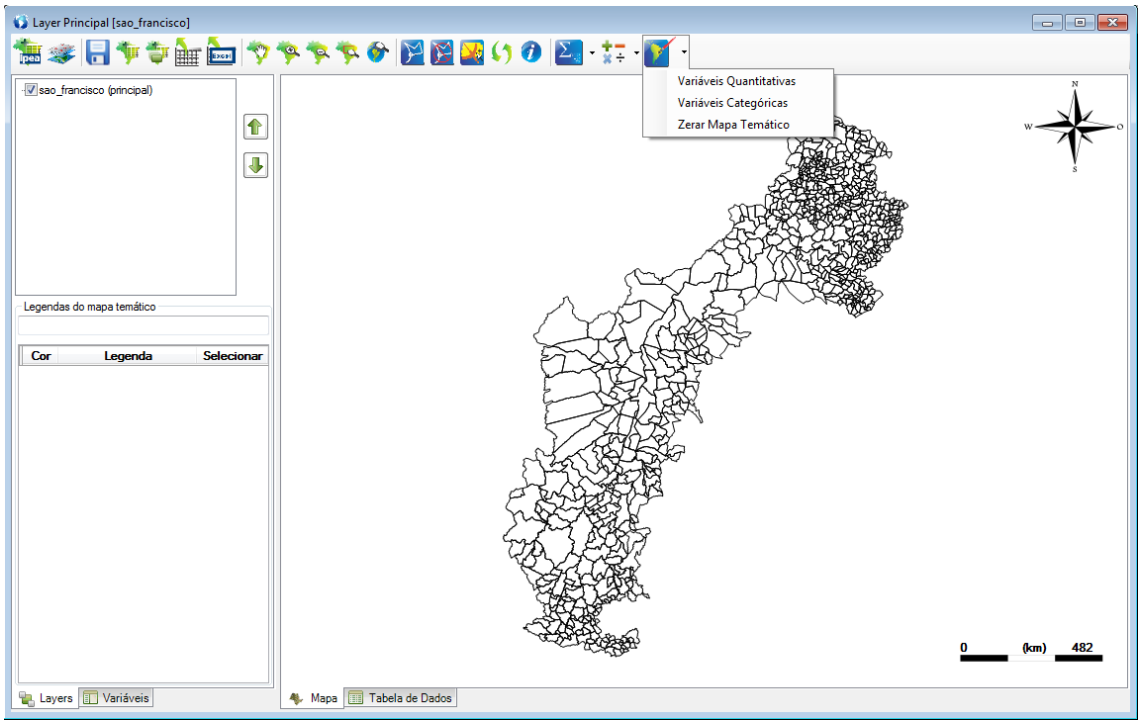
Equal: Os valores são divididos em k classes com intervalos iguais com o mesmo tamanho (w) em que w = Amplitude k.

Jenks *(Natural breaks):* É o método mais utilizado nos pacotes de análise espacial. Sua composição tem como objetivo criar classes com a menor variância interna e maior variabilidade externa. Pode demandar muito tempo de cálculos. No IpeaGEO os cálculos do método são feitos na forma logarítmica.

Valores Únicos: Cada valor é considerado uma classe e apresentado no mapa com uma cor diferente.

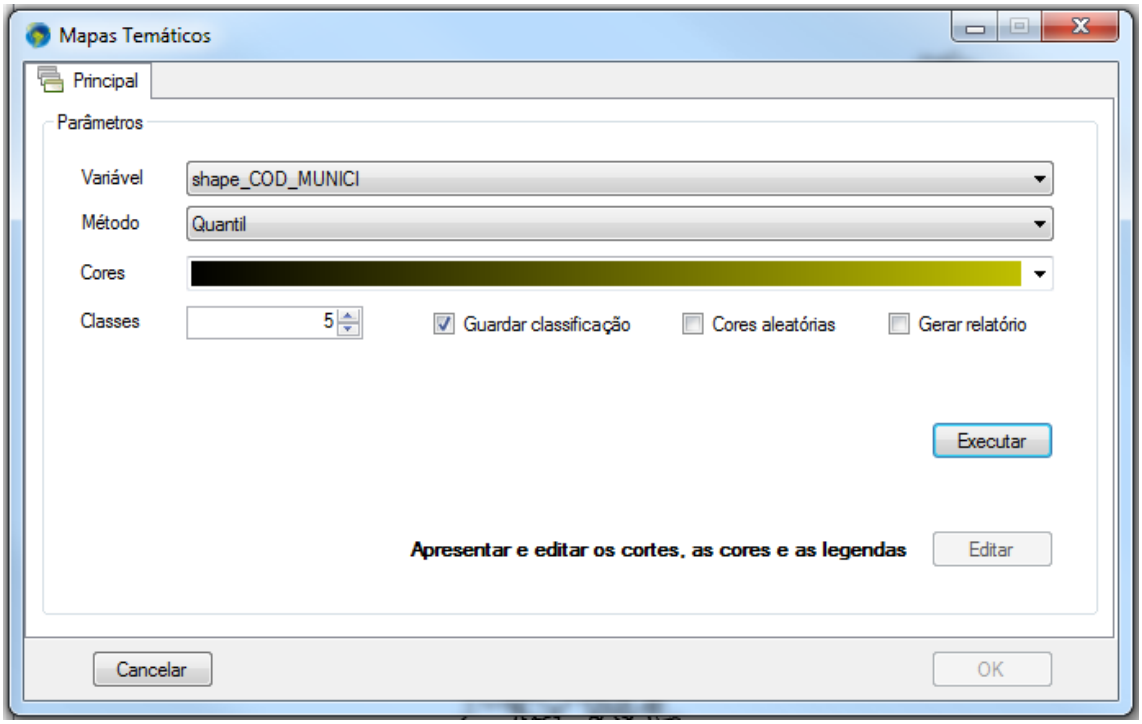
Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) e concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique no botão “Mapas Temáticos” e selecione uma das três opções exibidas abaixo.

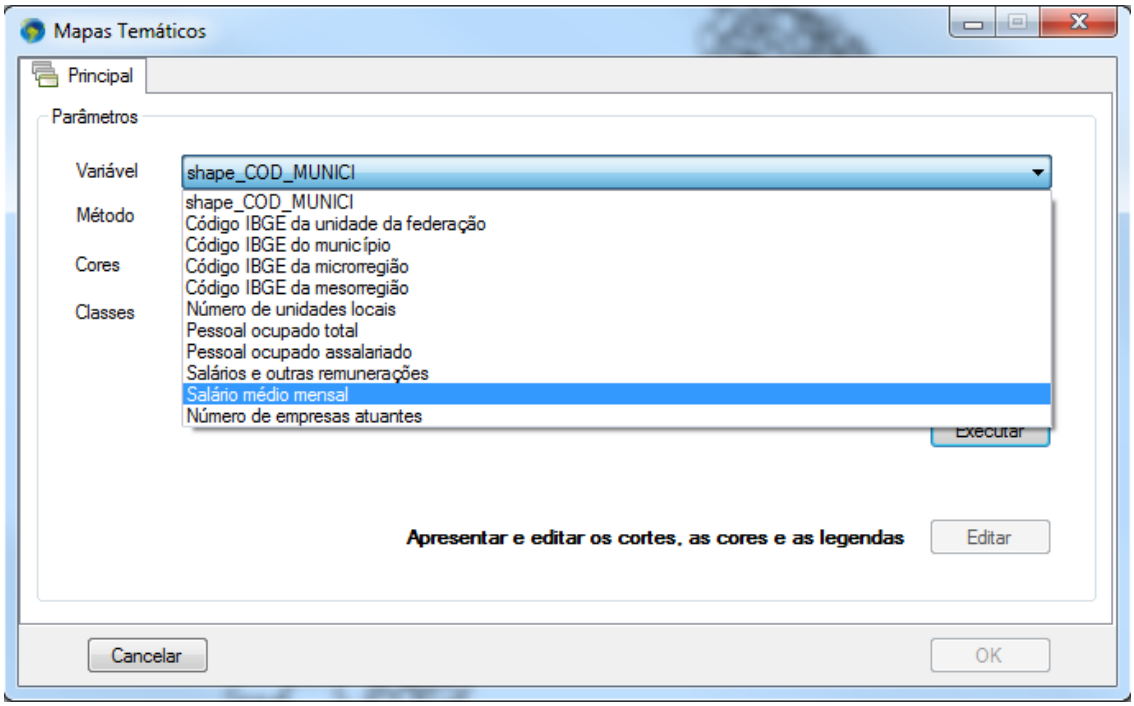


- **Variáveis Quantitativas**

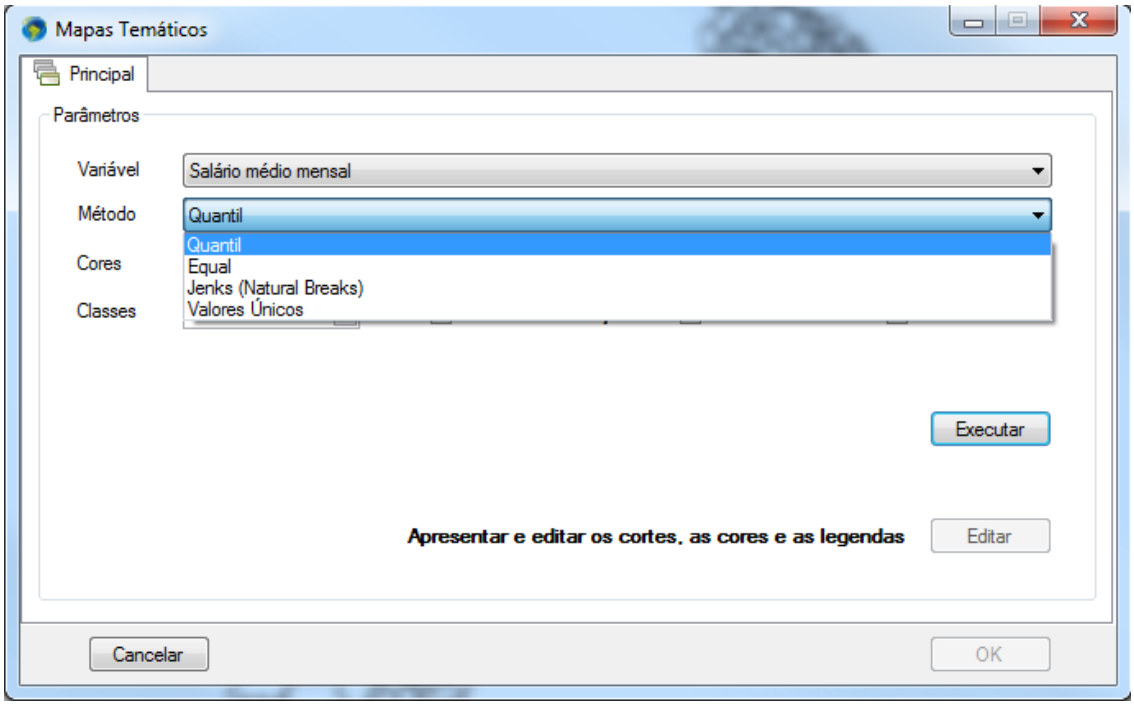
  Ao selecionar a opção “Variáveis Quantitativas”, abrirá a nova janela.



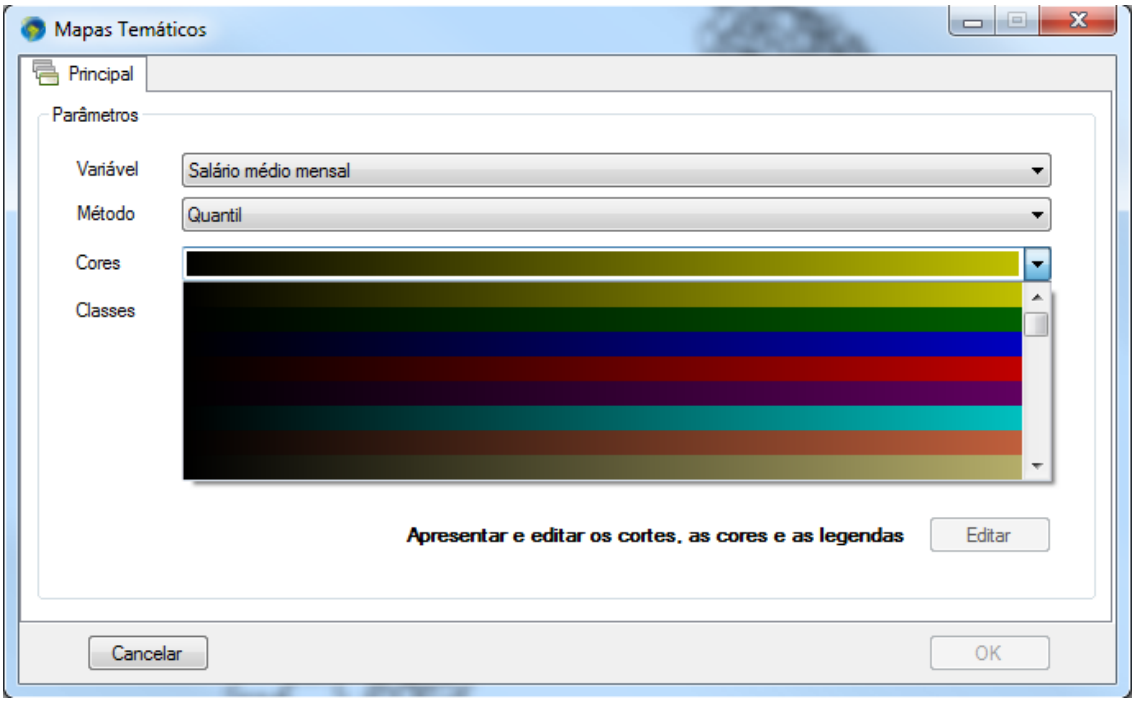
Na opção “Variável”, selecione a variável a ser visualizada no mapa.



Após escolher a variável desejada, selecione o método (Quantil, Equal, Jenks ou Valores Únicos) a ser aplicado para a categorização da variável selecionada.

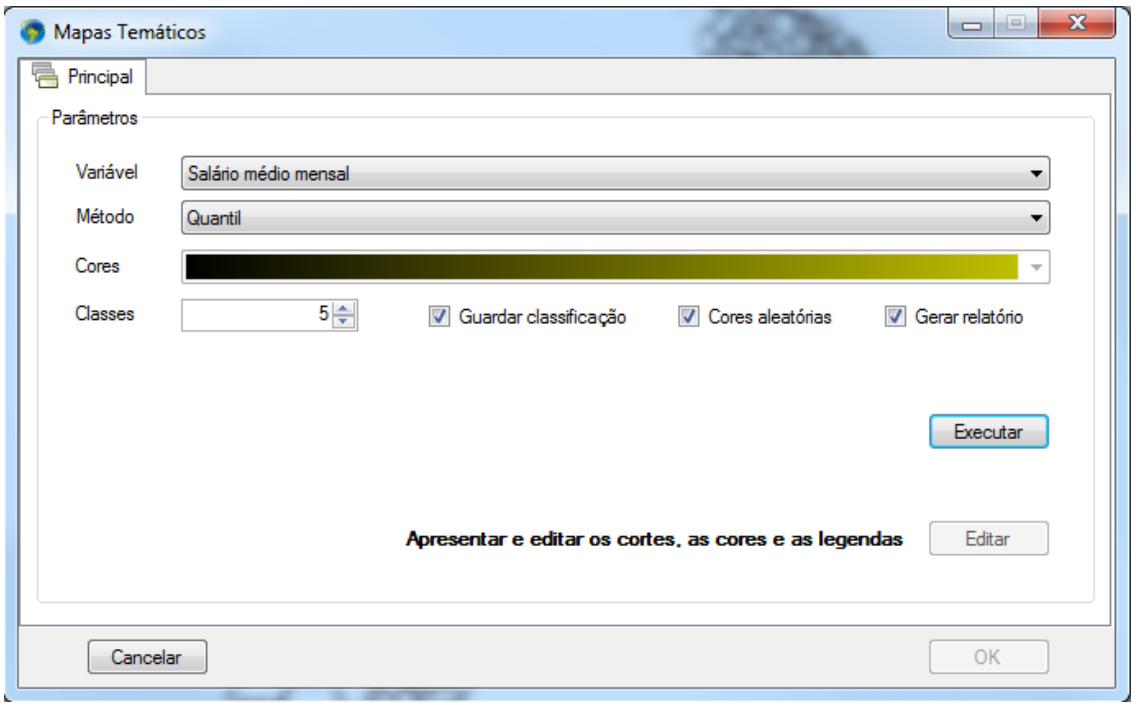


Escolha um padrão de cores que será utilizado para representar as categorias no mapa.

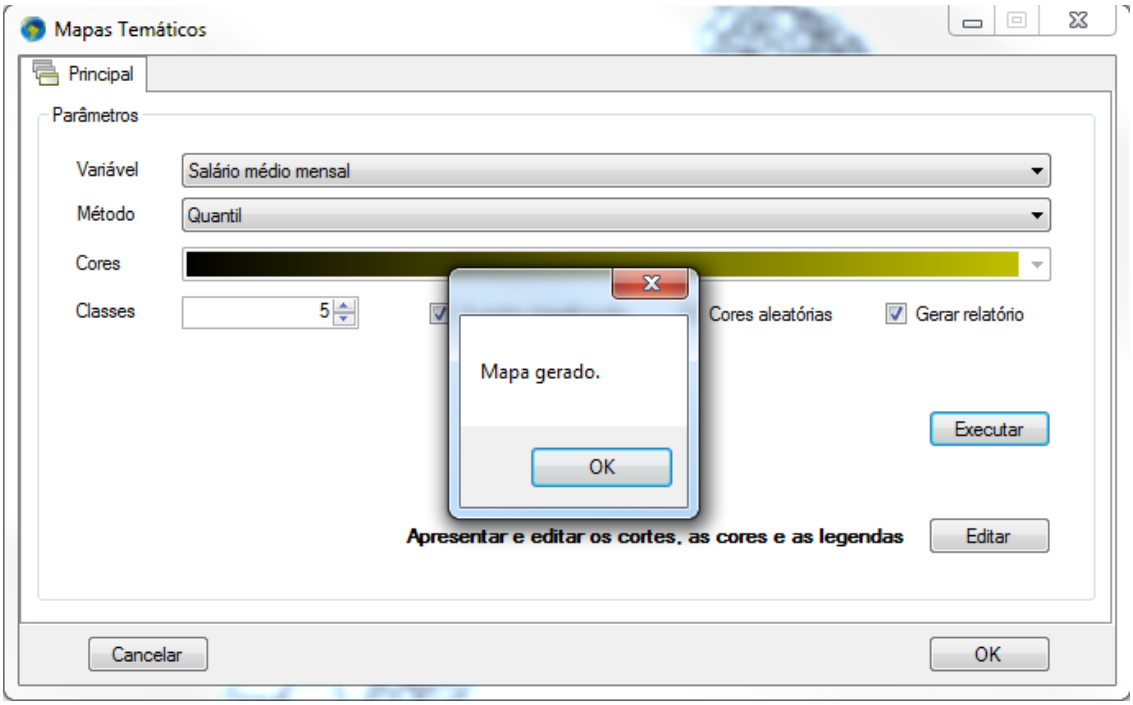


Após a escolha das opções “Variável”, “Método” e “Cores”, selecione o número de classes/categorias a serem criadas. Também é possível selecionar as opções:

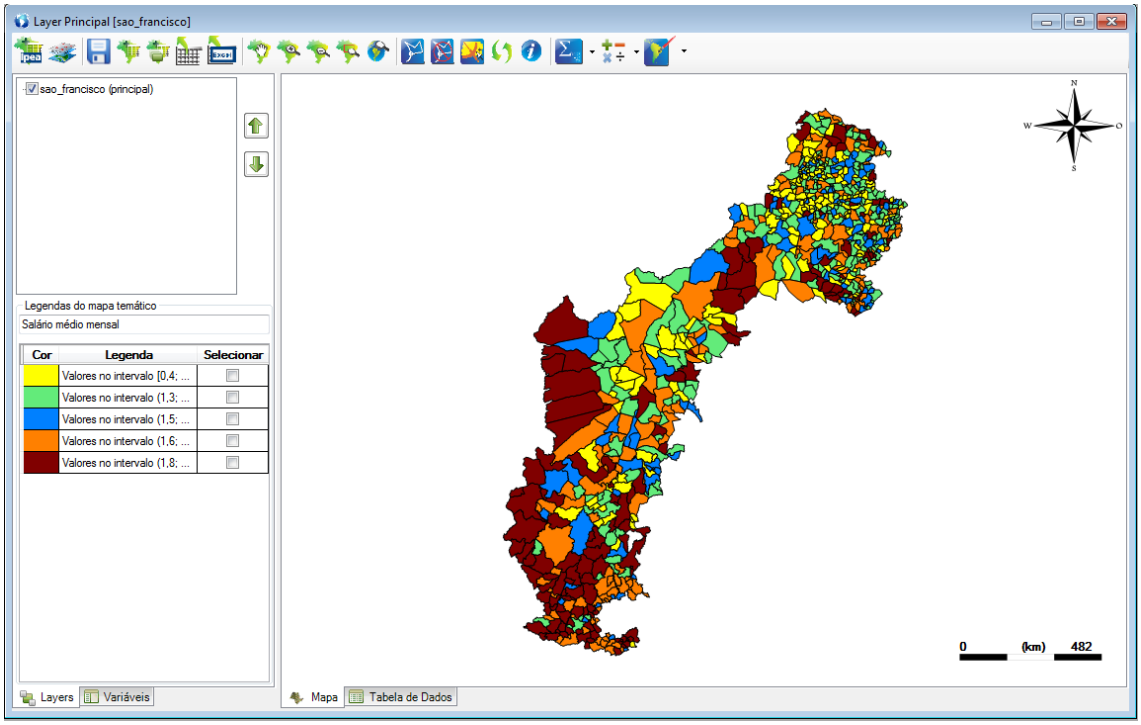
- “Guardar classificação”: Uma nova coluna denominada “Mapa Temático” é inserida na tabela de dados contendo identificação das categorias criadas;
- “Cores aleatórias”: O IpeaGEO utiliza cores aleatórias para representar as categorias no mapa;
- “Gerar relatório”: é gerado um relatório em formato PDF contendo o mapa e as categorias geradas.



Depois de selecionadas as opções desejadas, clique no botão “Executar”. Imediatamente surgirá uma nova janela informando a criação do novo mapa. Clique em “OK” duas vezes para visualizar o mapa gerado.



Mapa gerado utilizando a ferramenta de cores aleatórias:

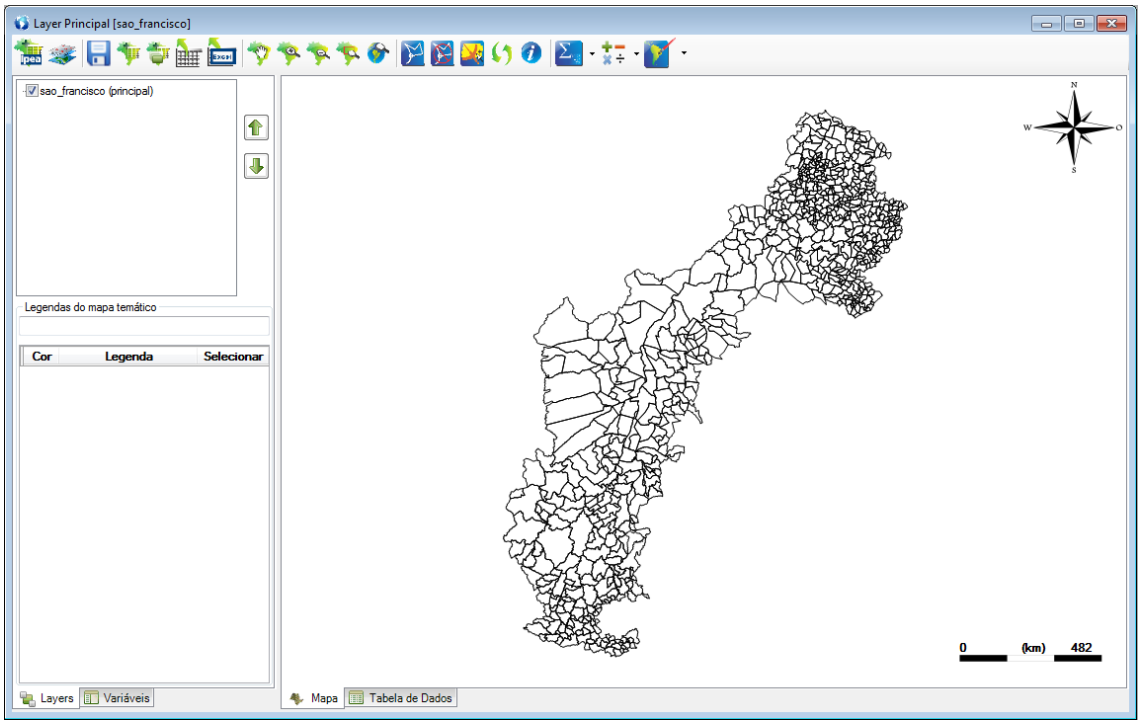


- **Variáveis Categóricas**

Para gerar um Mapa Temático para “Variáveis Categóricas”, basta repetir o mesmo procedimento descrito para as “Variáveis Quantitativas” onde o único método disponível para exibição dos dados no mapa é o método dos “Valores Únicos”.

- **Zerar Mapa Temático**

Reinicia o arquivo malha, apagando os mapas temáticos gerados em análises anteriores.



Funções Estatísticas do IpeaGEO

Análise Exploratória dos Dados

Ao clicar com o botão direito do mouse sobre a tabela de dados, são disponibilizadas ao usuário uma série de ferramentas para realização de analise exploratória dos dados.

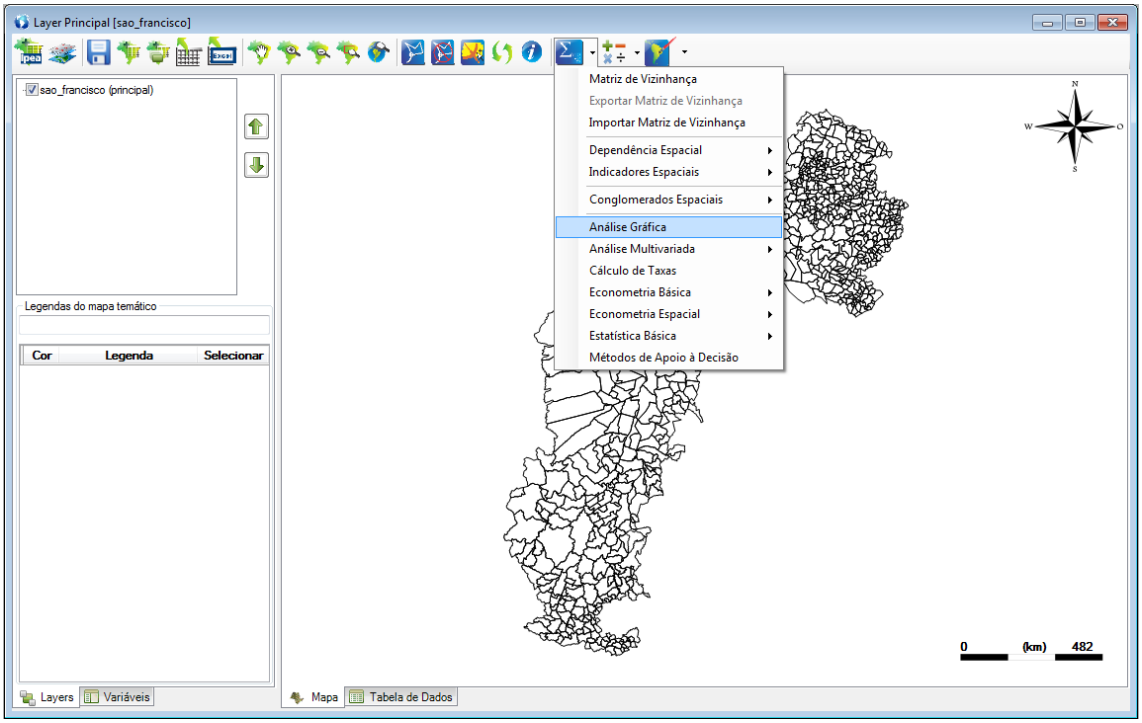
Análise Gráfica

Visão Geral

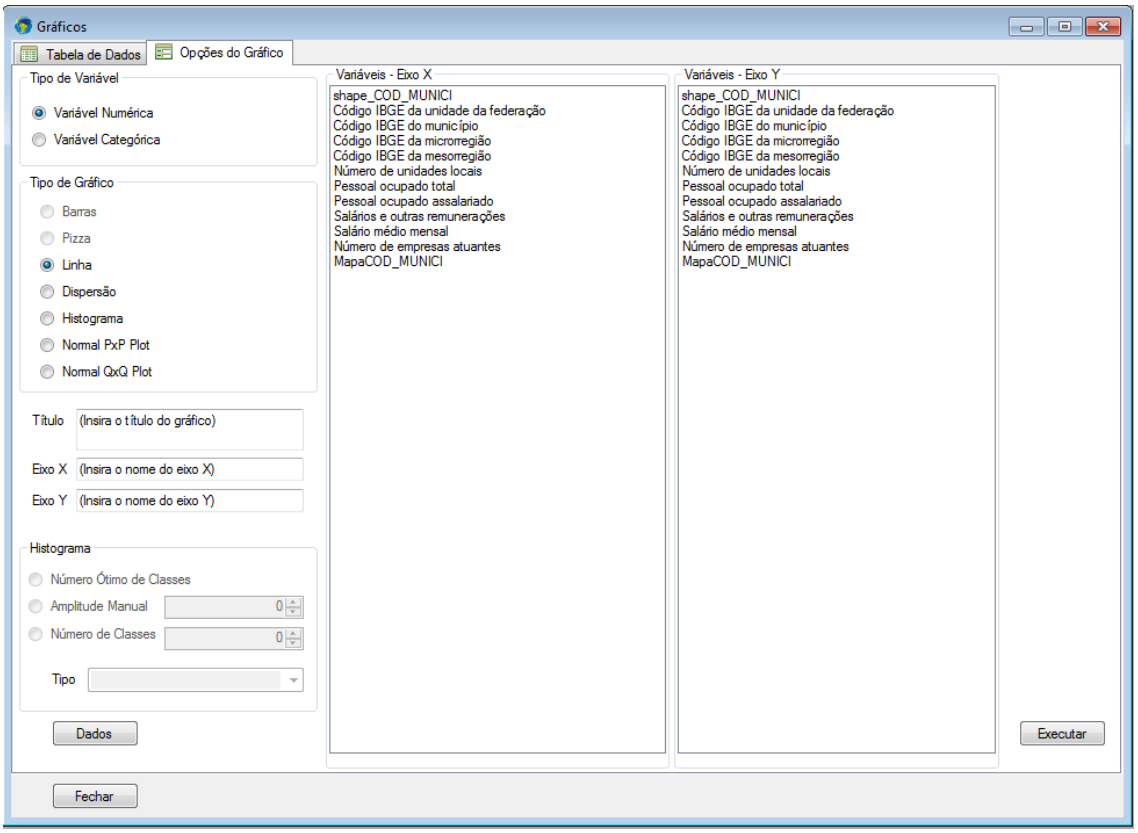
A análise gráfica é uma ferramenta que auxilia na análise exploratória dos dados, onde o usuário pode verificar, por meio de gráficos, o comportamento dos dados em estudo.

Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique na opção “Análise Exploratória e manipulação de dados” e, em seguida, na opção “Análise Gráfica”.

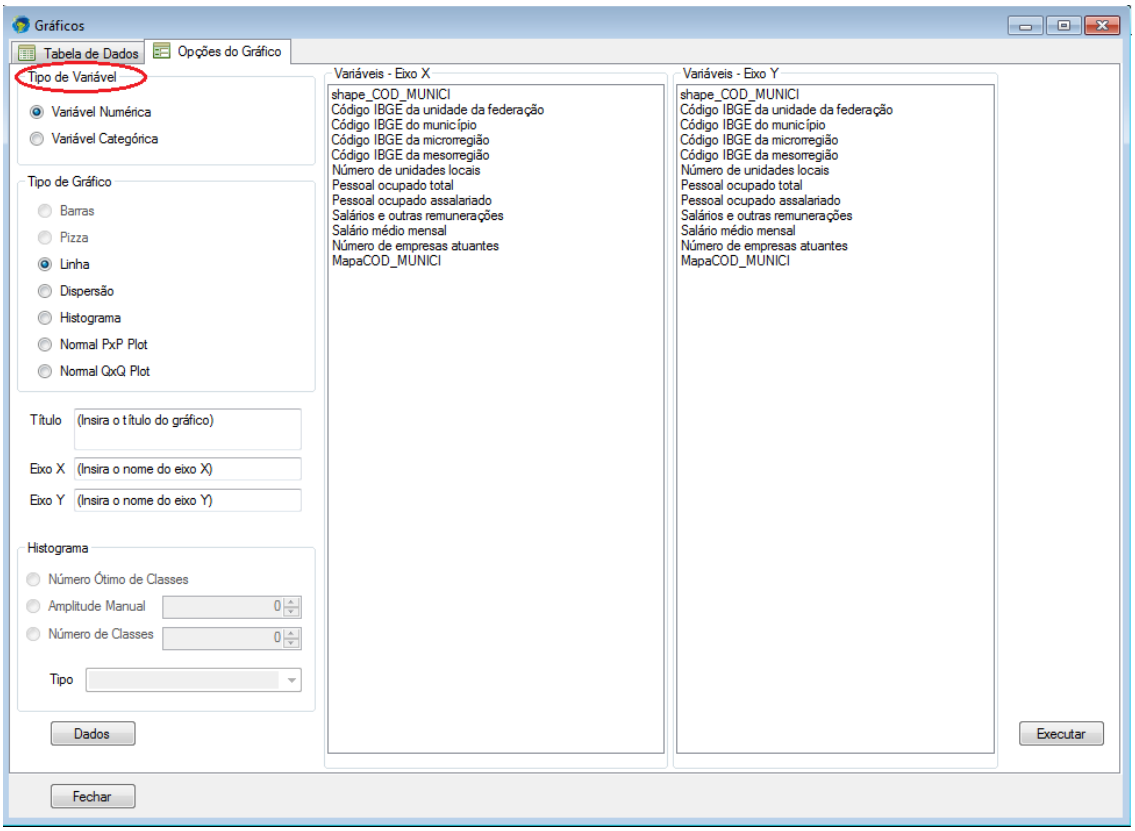


O seguinte formulário irá se abrir:



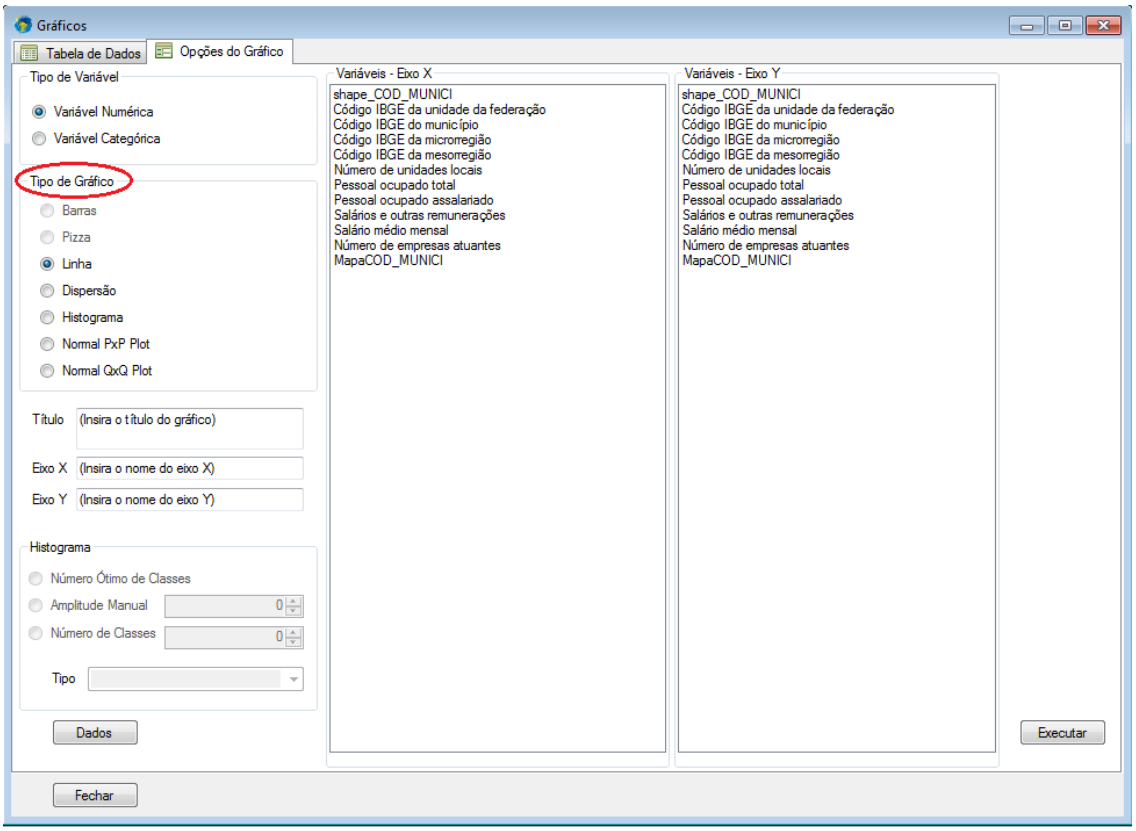
Escolha o tipo de variável a ser analisada:

- Variável Numérica;
- Variável Categórica.

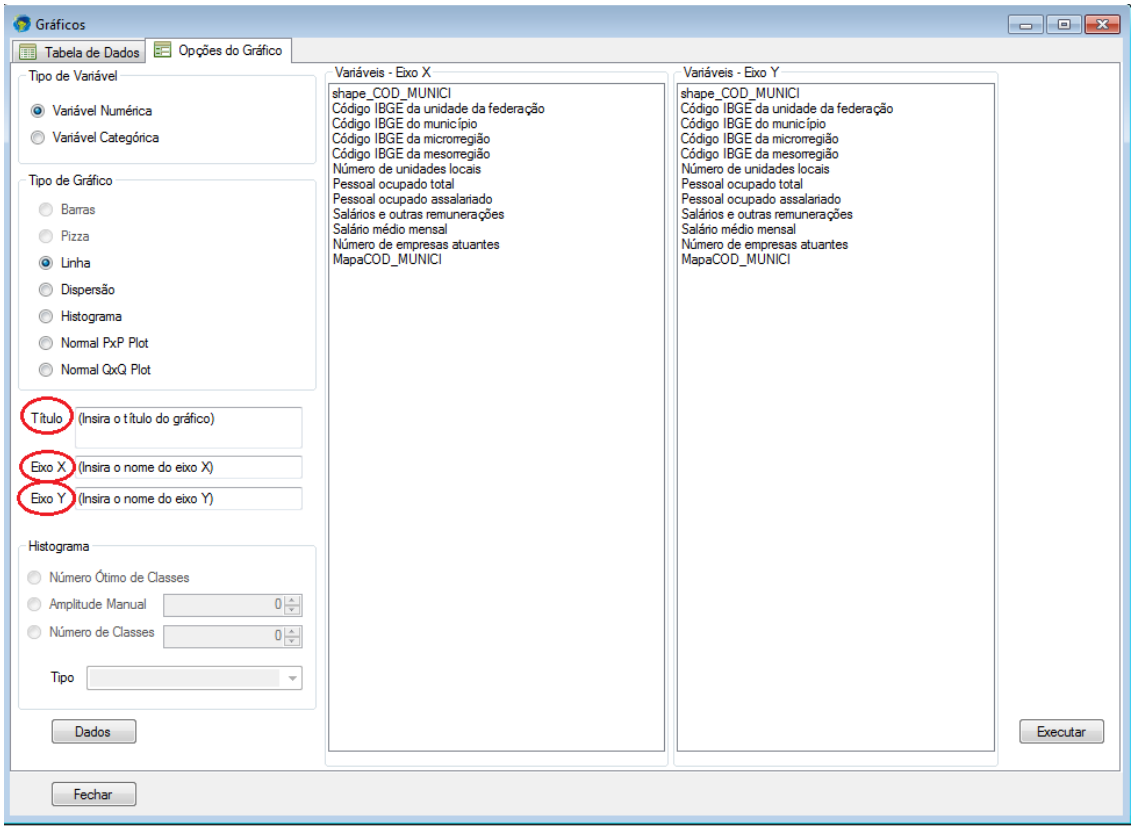


Os tipos de gráficos disponíveis para variáveis numéricas são: gráfico de linha, diagrama de dispersão, histograma, PP plot e QQ plot (esses últimos para a distribuição normal).

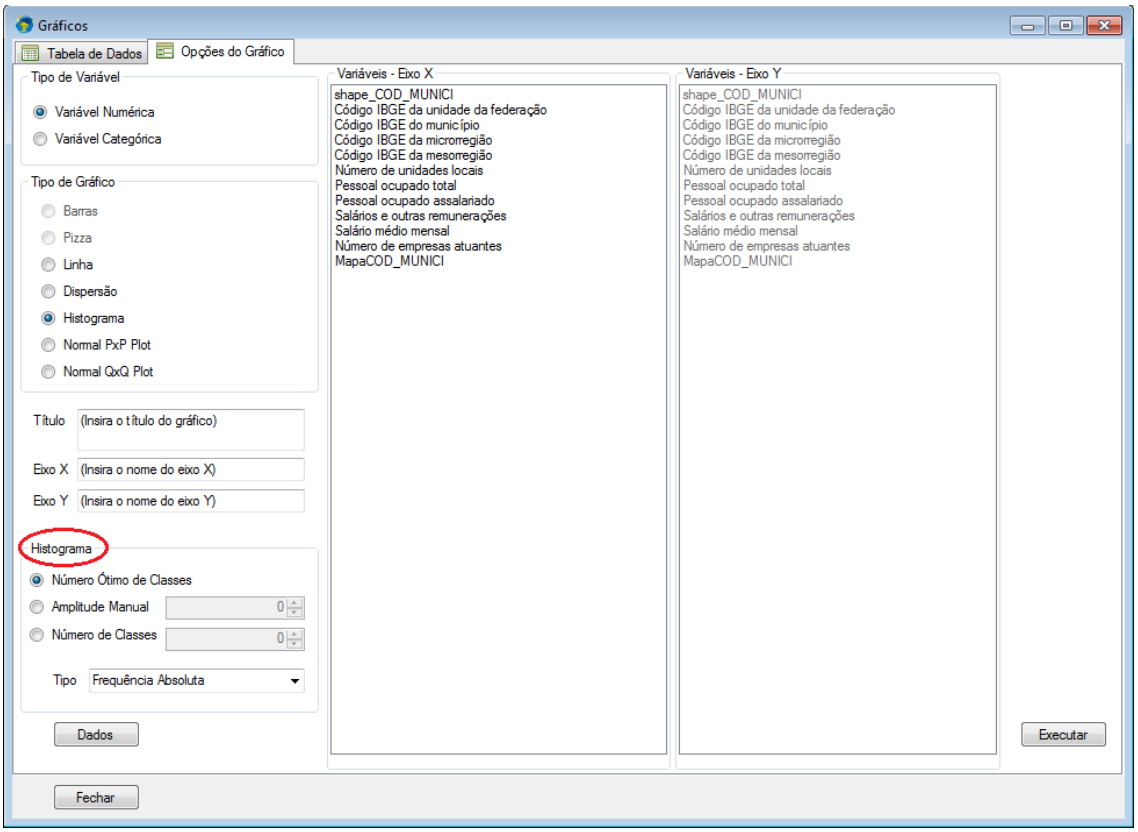
Os tipos de gráficos disponíveis para variáveis categóricas são o gráfico de barras e o de setores ou pizza.



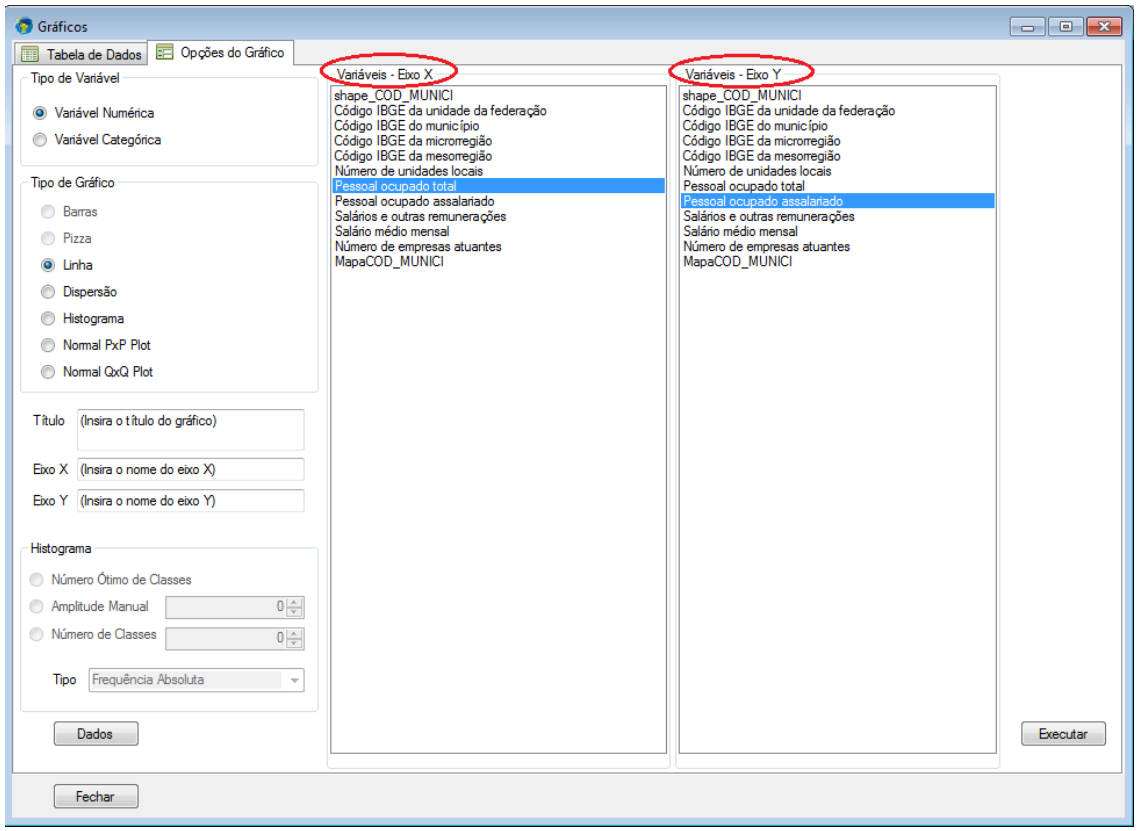
Para qualquer tipo de variável é possível inserir o título do gráfico (Título) e os nomes dos eixos das abscissas (Nome do Eixo X) e ordenadas (Nome do Eixo Y).



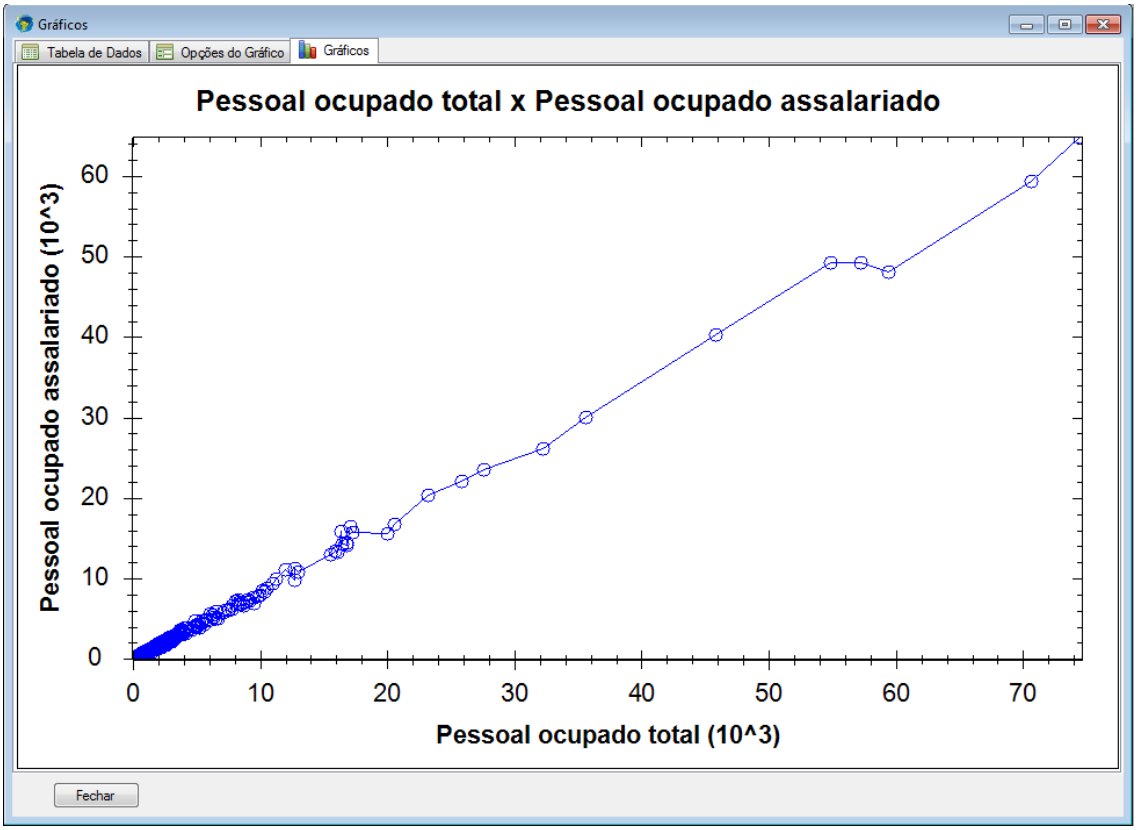
Para o gráfico do tipo histograma é possível selecionar uma variável categórica para o eixo das abscissas (Variáveis - Eixo X); definir as classes por meio da quantidade ótima de classes (Número Ótimo de Classes), da amplitude de cada classe (Amplitude Manual) ou por uma quantidade fixa de classes (Número de Classes); além de escolher se a frequência observada em cada classe será exibida como frequência absoluta ou relativa.



Após escolha do tipo de gráfico, selecionam-se as variáveis dos eixos das abscissas (X) e das ordenadas (Y).



Ao clicar em “Executar” o gráfico selecionado é gerado em uma nova aba nomeada “Gráficos”.

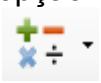


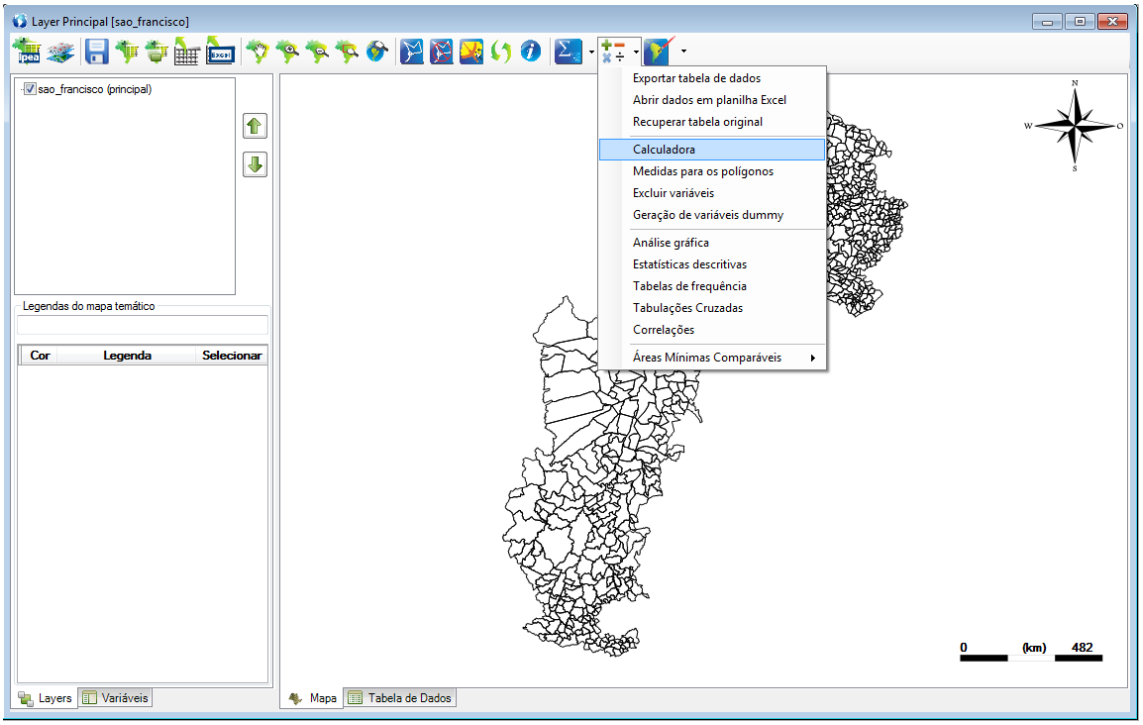
Calculadora

Visão Geral

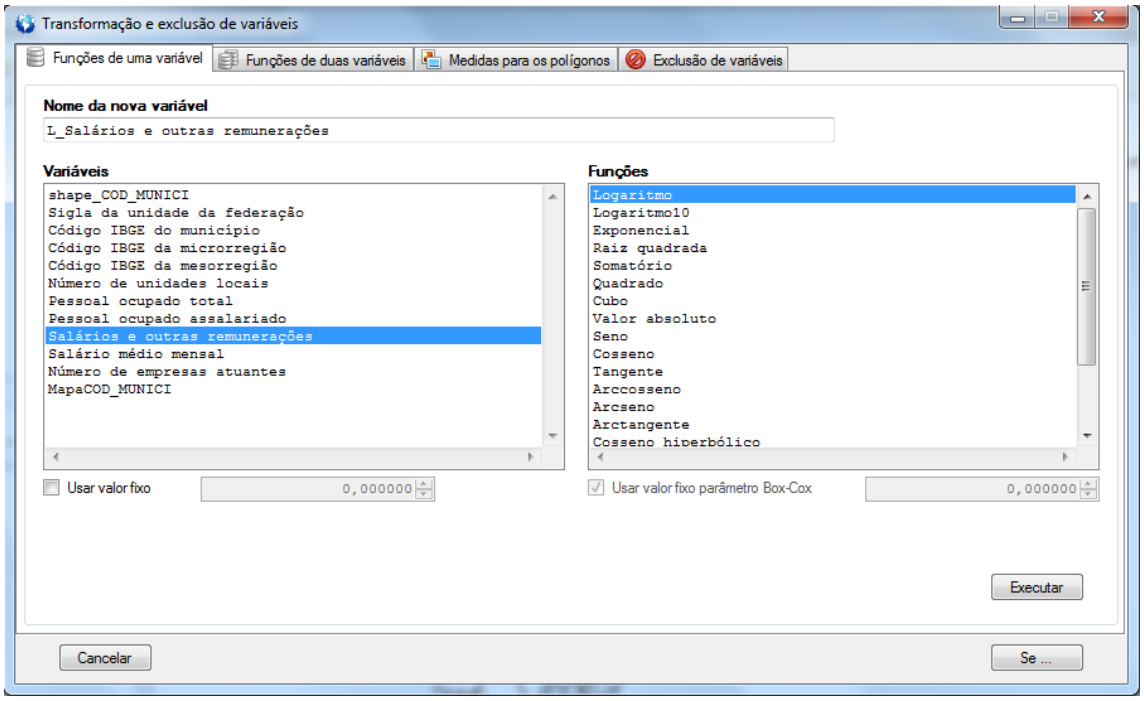
A Calculadora do IpeaGEO permite a aplicação de diversas funções em um conjunto de dados.

Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique na opção “Calculadora” localizada no botão "Análise Exploratória e Manipulação de Dados"  .



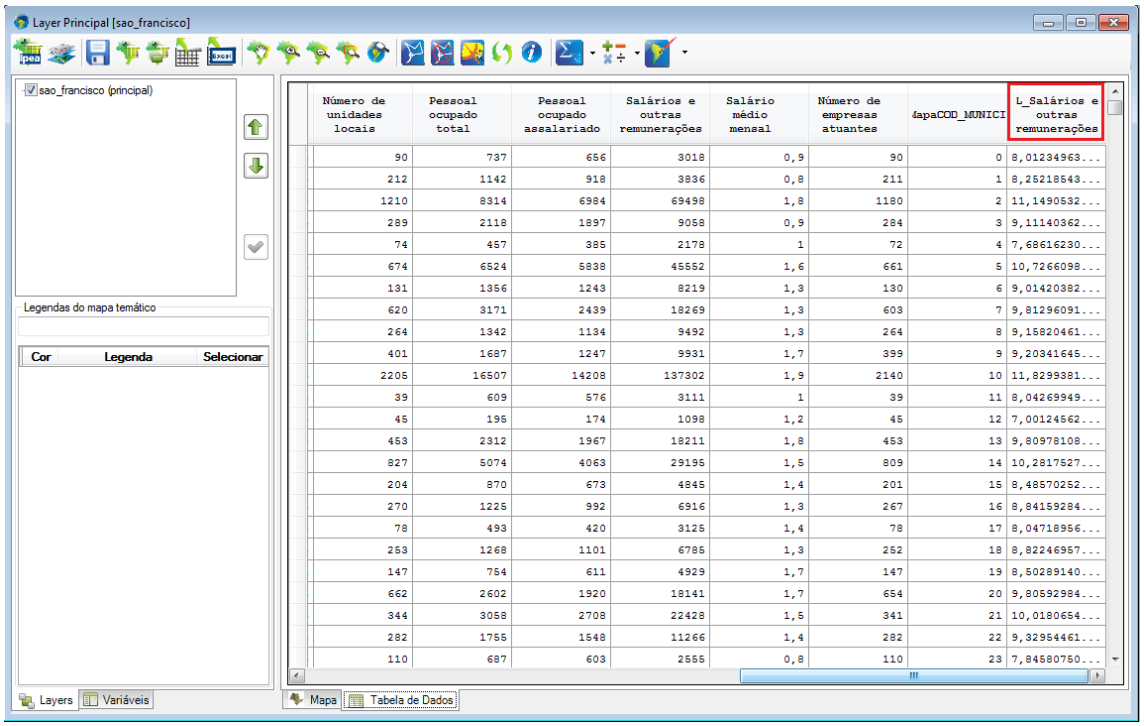
O seguinte formulário irá se abrir:



As funções disponíveis para aplicação e transformação de uma variável são logaritmo natural (Logaritmo), Logaritmo na base 10, Exponencial, Raiz Quadrada, elevar ao quadrado (Quadrado), elevar ao cubo (Cubo), Valor Absoluto, as funções trigonométricas Seno, Cosseno, Tangente, arco seno (Arcseno), arco cosseno (Arccosseno), arco tangente (Arctangente), Seno Hiperbólico, Cosseno Hiperbólico e Tangente Hiperbólica. Também é possível calcular lags espaciais (Lag Espacial) e a transformação de Box-Cox (Transformação Box-Cox).

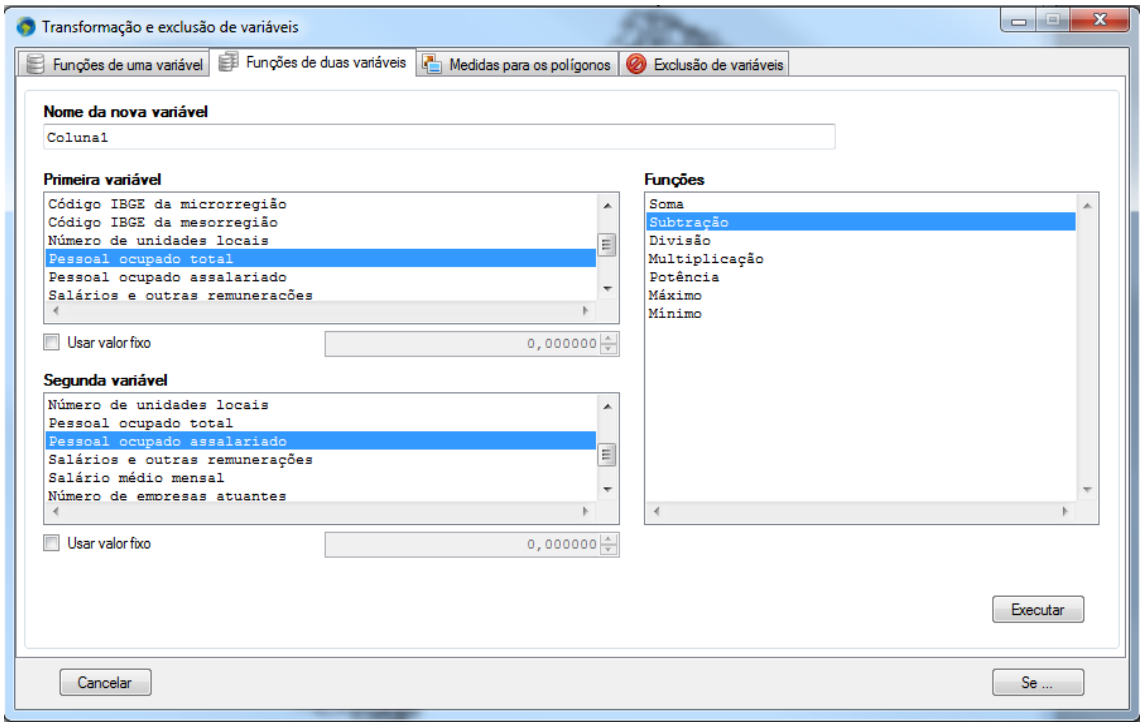
O usuário também dispõe das opções de transformação de variáveis "Usar valor fixo" e "Usar valor fixo parâmetro Box-Cox" (esta última fica disponível quando a opção Transformação Box-Cox estiver selecionada).

Ao clicar no botão "Executar", na tabela de dados principal será acrescentada uma nova coluna com o resultado da operação matemática realizada.



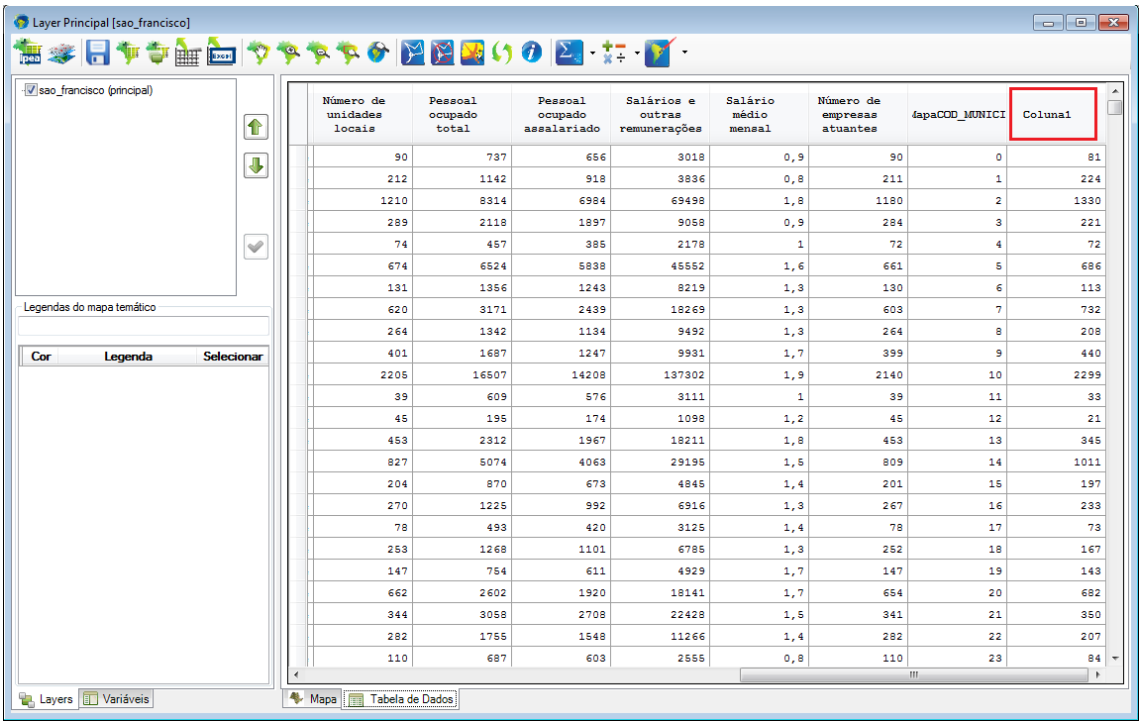
Número de unidades locais	Pessoal ocupado total	Pessoal ocupado assalariado	Salários e outras remunerações	Salário médio mensal	Número de empresas atuantes	fapaCOD_MUNICI	L_Salários e outras remunerações
90	737	656	3018	0,9	90	0	8,01234963...
212	1142	918	3836	0,8	211	1	8,25218543...
1210	8314	6984	69498	1,8	1180	2	11,1490532...
289	2118	1897	9058	0,9	284	3	9,11140362...
74	457	385	2178	1	72	4	7,68616230...
674	6524	5838	45552	1,6	661	5	10,7266098...
131	1356	1243	8219	1,3	130	6	9,01420382...
620	3171	2439	18269	1,3	603	7	9,81296091...
264	1342	1134	9492	1,3	264	8	9,15820461...
401	1687	1247	9931	1,7	399	9	9,20341645...
2205	16507	14208	137302	1,9	2140	10	11,8299381...
39	609	576	3111	1	39	11	8,04269949...
45	195	174	1098	1,2	45	12	7,00124562...
453	2312	1967	18211	1,8	453	13	9,80978108...
827	5074	4063	29195	1,5	809	14	10,2817527...
204	870	673	4845	1,4	201	15	8,48570252...
270	1225	992	6916	1,3	267	16	8,84159284...
78	493	420	3125	1,4	78	17	8,04718956...
253	1268	1101	6785	1,3	252	18	8,82246957...
147	754	611	4929	1,7	147	19	8,50289140...
662	2602	1920	18141	1,7	654	20	9,80592984...
344	3058	2708	22428	1,5	341	21	10,0180654...
282	1755	1548	11266	1,4	282	22	9,32954461...
110	687	603	2555	0,8	110	23	7,84580750...

Na aba "Funções de duas variáveis" se encontra o seguinte formulário:



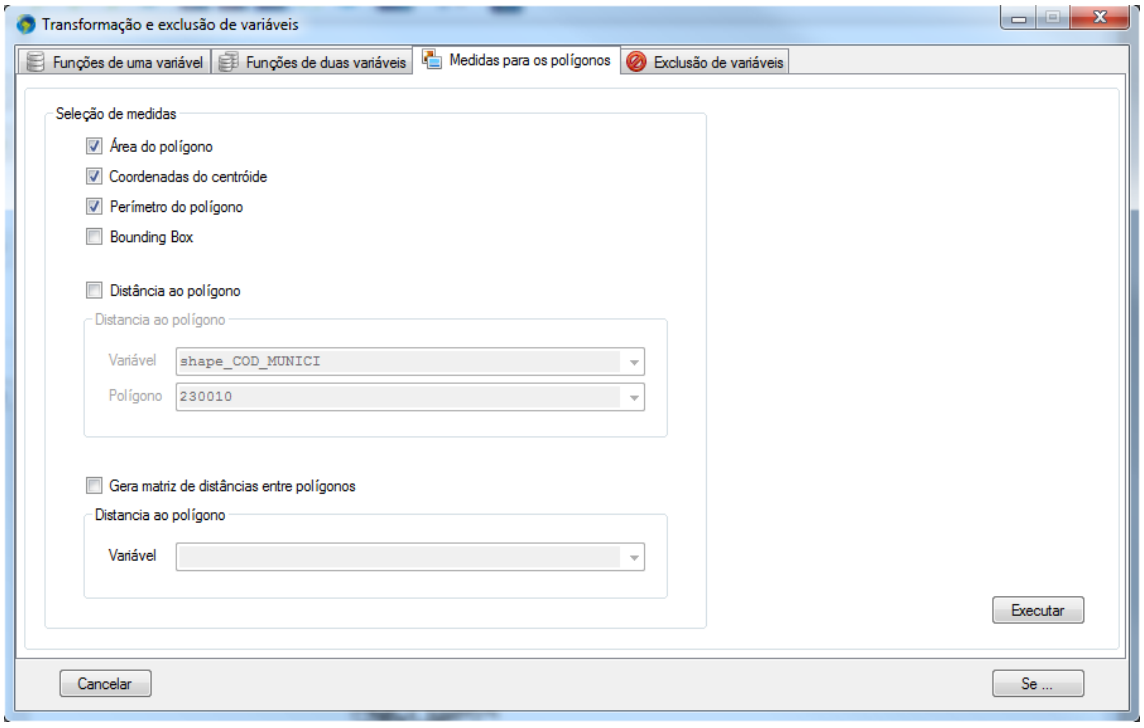
As funções disponíveis a serem aplicadas a duas variáveis são: soma, subtração, divisão, multiplicação, potência, valor máximo e valor mínimo.

Ao clicar no botão "Executar", será acrescentada na tabela de dados principal uma nova coluna com a operação matemática realizada.



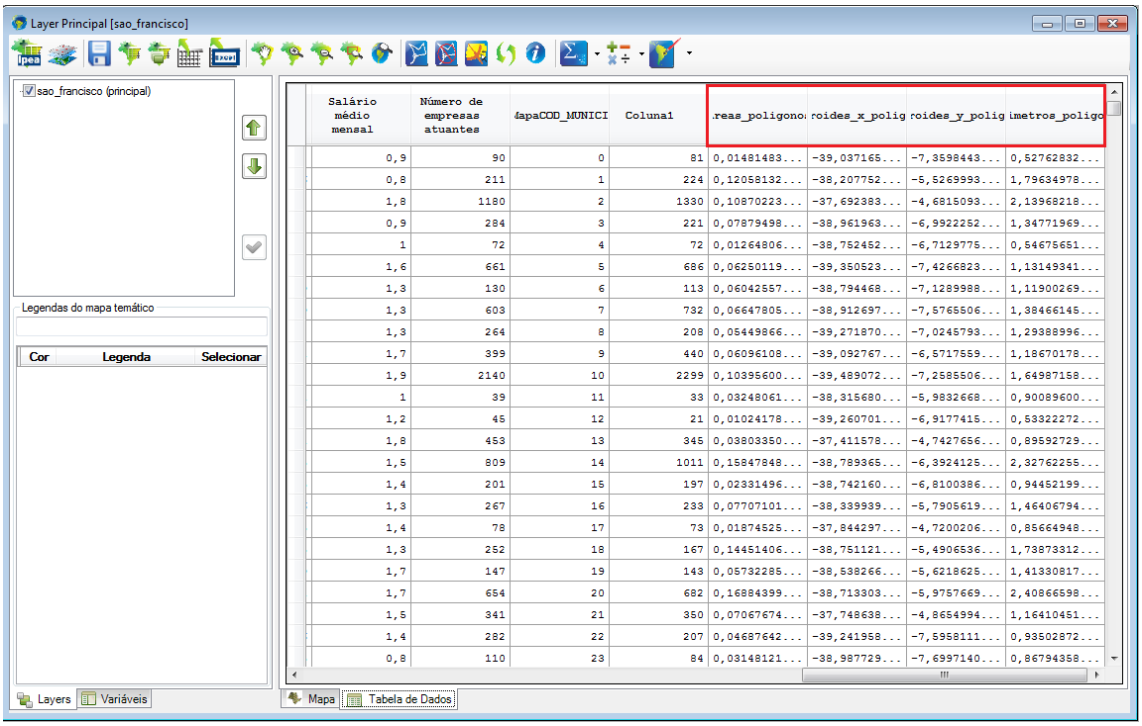
Número de unidades locais	Pessoal ocupado total	Pessoal ocupado assalariado	Salários e outras remunerações	Salário médio mensal	Número de empresas atuantes	&apaCOD_MUNICI	Coluna1
90	737	656	3018	0,9	90	0	81
212	1142	918	3836	0,8	211	1	224
1210	8314	6984	69498	1,8	1180	2	1330
289	2118	1897	9058	0,9	284	3	221
74	457	385	2178	1	72	4	72
674	6524	5838	45552	1,6	661	5	686
131	1356	1243	8219	1,3	130	6	113
620	3171	2439	18269	1,3	603	7	732
264	1342	1134	9492	1,3	264	8	208
401	1687	1247	9931	1,7	399	9	440
2205	16507	14208	137302	1,9	2140	10	2299
39	609	576	3111	1	39	11	33
45	195	174	1098	1,2	45	12	21
453	2312	1967	18211	1,8	453	13	345
827	5074	4063	29195	1,5	809	14	1011
204	870	673	4845	1,4	201	15	197
270	1225	992	6916	1,3	267	16	233
78	493	420	3125	1,4	78	17	73
253	1268	1101	6785	1,3	252	18	167
147	754	611	4929	1,7	147	19	143
662	2602	1920	18141	1,7	654	20	682
344	3058	2708	22428	1,5	341	21	350
282	1755	1548	11266	1,4	282	22	207
110	687	603	2555	0,8	110	23	84

Na aba "Medidas para polígonos" está contido o seguinte formulário:



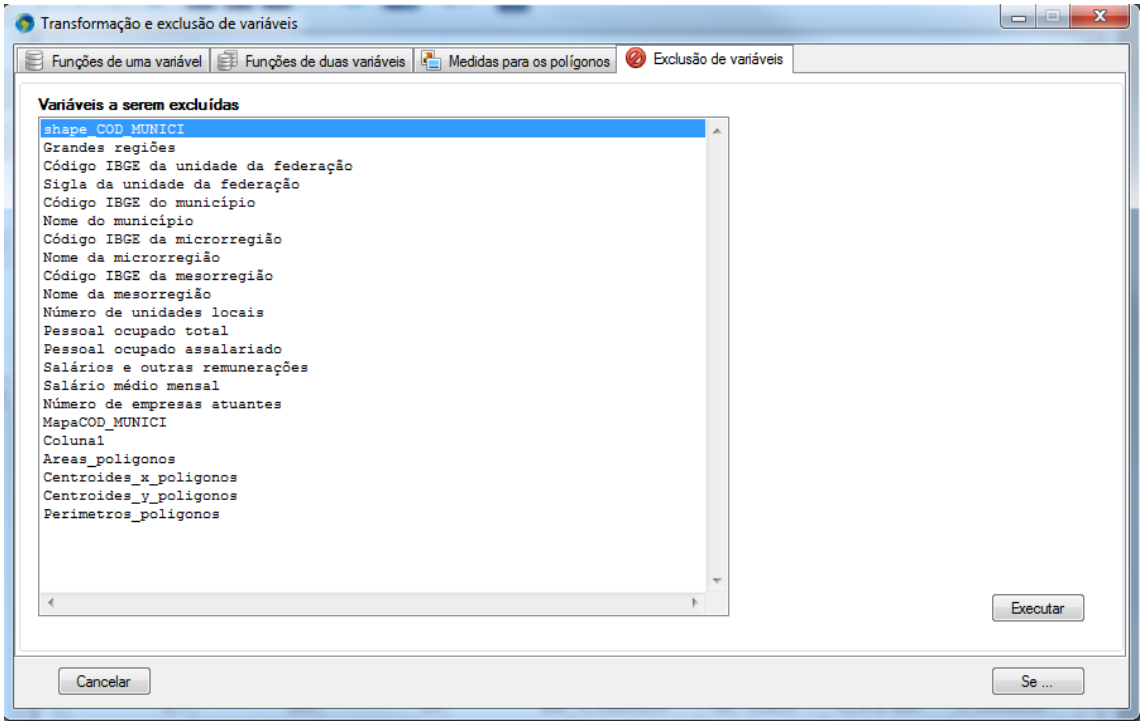
Caso o usuário esteja trabalhando com os dados concatenados com uma malha digital, é possível obter as seguintes medidas para os polígonos: Área do Polígono, Perímetro do Polígono, Coordenada do Centroide, as coordenadas do retângulo limite do polígono (Bounding Box), a distância entre uma determinada variável e o polígono escolhido; também é permitido gerar a matriz de distâncias entre polígonos.

Ao clicar no botão "Executar", serão acrescentadas na tabela de dados principal novas colunas com as seleções de medidas feitas.

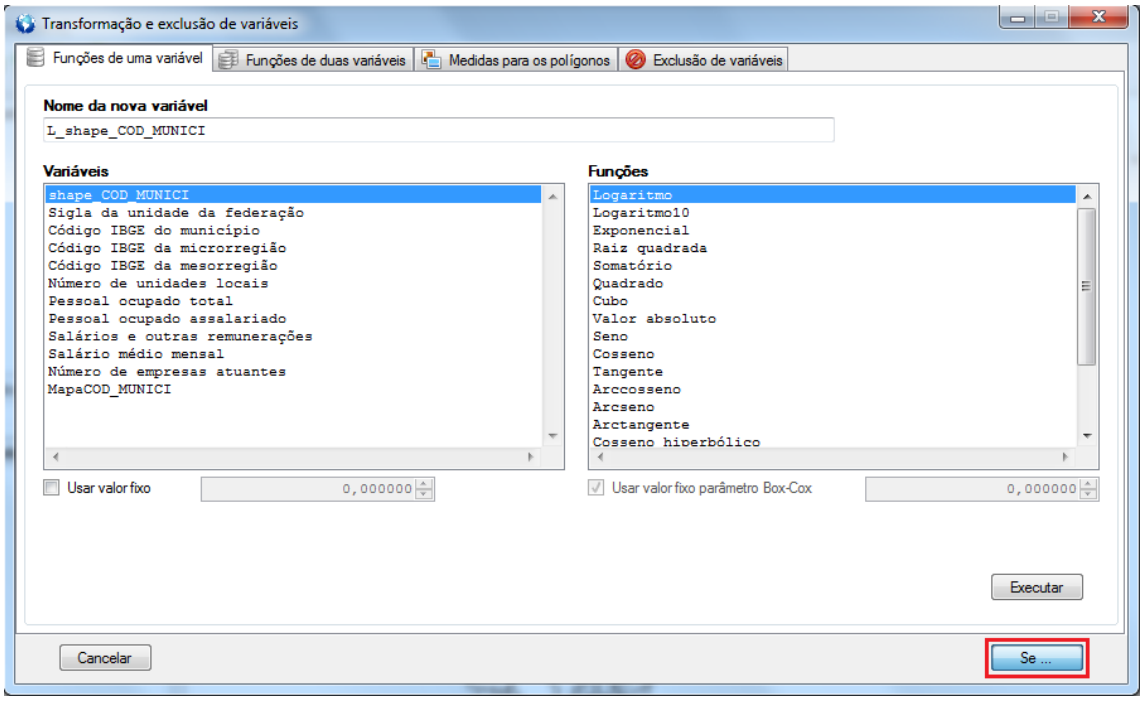


Salário médio mensal	Número de empresas atuantes	MapaCOD_MUNICI	Coluna1	reas_poligono	roides_x_polig	roides_y_polig	imetros_poligo
0,9	90	0	81	0,01481483...	-39,037165...	-7,3598443...	0,52762832...
0,8	211	1	224	0,12058132...	-38,207752...	-5,5269993...	1,79634978...
1,8	1180	2	1330	0,10870223...	-37,692383...	-4,6815093...	2,13968218...
0,9	284	3	221	0,07879498...	-38,961963...	-6,9922252...	1,34771969...
1	72	4	72	0,01264806...	-38,752452...	-6,7129775...	0,54675651...
1,6	661	5	686	0,06250119...	-39,350523...	-7,4266823...	1,13149341...
1,3	130	6	113	0,06042557...	-38,794468...	-7,1289988...	1,11900269...
1,3	603	7	732	0,06647805...	-38,912697...	-7,5765506...	1,38466145...
1,3	264	8	208	0,05449866...	-39,271870...	-7,0245793...	1,29388996...
1,7	399	9	440	0,06096108...	-39,092767...	-6,5717559...	1,18670178...
1,9	2140	10	2299	0,10395600...	-39,489072...	-7,2585506...	1,64987158...
1	39	11	33	0,03248061...	-38,315680...	-5,9832668...	0,90089600...
1,2	45	12	21	0,01024178...	-39,260701...	-6,9177415...	0,53322272...
1,8	453	13	345	0,03803350...	-37,411578...	-4,7427656...	0,89592729...
1,5	809	14	1011	0,15847848...	-38,789365...	-6,3924125...	2,32762255...
1,4	201	15	197	0,02331496...	-38,742160...	-6,8100386...	0,94452199...
1,3	267	16	233	0,07707101...	-38,339939...	-5,7905619...	1,46406794...
1,4	78	17	73	0,01874525...	-37,844297...	-4,7200206...	0,85664948...
1,3	252	18	167	0,14451406...	-38,751121...	-5,4906536...	1,73873312...
1,7	147	19	143	0,05732285...	-38,538266...	-5,6218625...	1,41330817...
1,7	654	20	682	0,16884399...	-38,713303...	-5,9757669...	2,40866598...
1,5	341	21	350	0,07067674...	-37,748638...	-4,8654994...	1,16410451...
1,4	282	22	207	0,04687642...	-39,241958...	-7,5958111...	0,93502872...
0,8	110	23	84	0,03148121...	-38,987729...	-7,6997140...	0,86794358...

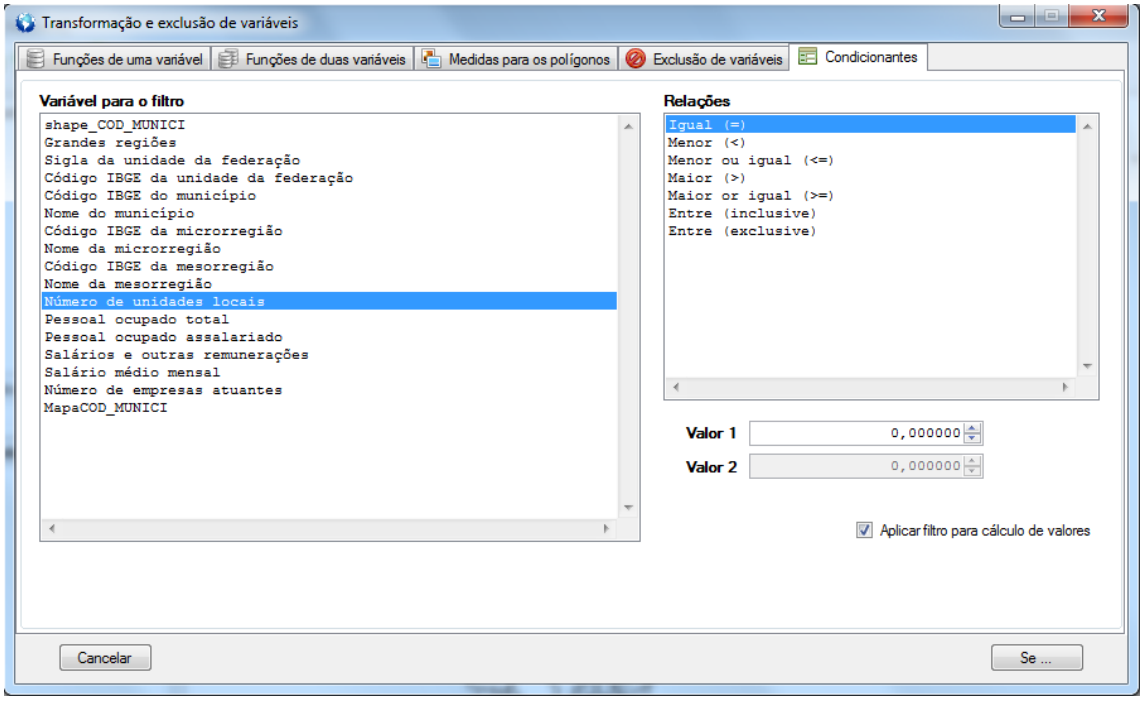
A Calculadora também permite ao usuário excluir variáveis da tabela de dados na aba “Exclusão de variáveis”, que apresenta o seguinte formulário:



Em todas as abas do formulário Calculadora encontra-se a disposição do usuário o botão “Se” localizado no canto inferior direito da janela.



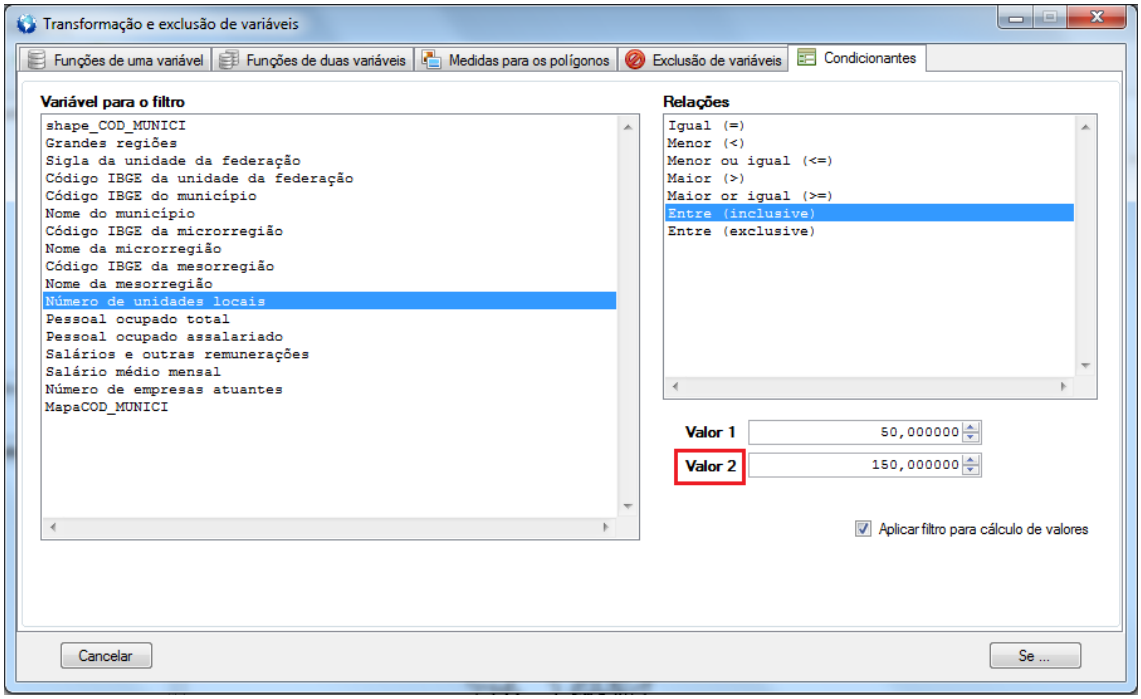
Ao selecionar este botão a aba “Condicionantes” ficará disponível para o usuário realizar as operações desejadas. As opções que podem ser escolhidas são: Igual, Menor, Menor ou igual, Maior, Maior ou igual, Entre (inclusive) e Entre (exclusive).



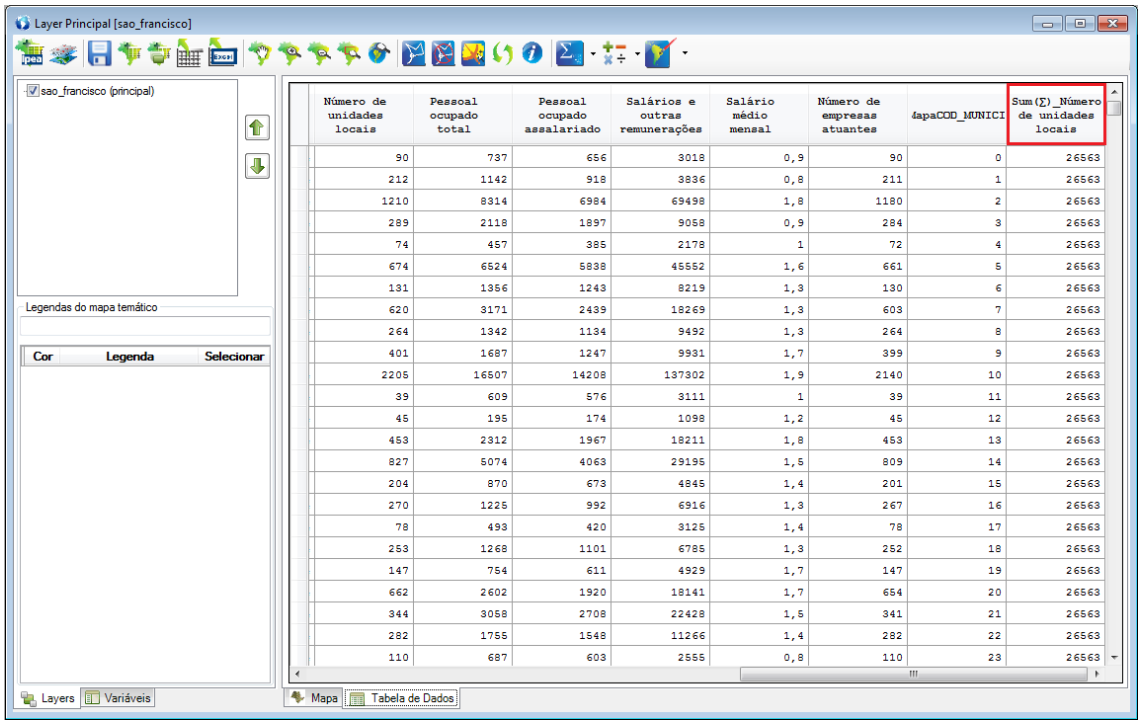
Na opção “Valor 1” o usuário deve escolher o número a ser utilizado como base para realizar o cálculo. Ao selecionar as operações “Entre (inclusive)” ou “Entre (exclusive)” a opção “Valor 2” também se torna disponível para o preenchimento.

Ressalta-se que para utilizar o filtro estabelecido na nova variável gerada é

necessário que a opção localizada no canto inferior direito “Aplicar filtro para cálculo de valores” esteja marcada.



Após serem escolhidos os parâmetros o usuário deve retornar à aba em que estava realizando a operação para finalizar o procedimento clicando no botão “Executar”. O exemplo ilustrado abaixo apresenta o somatório para a variável “Número de unidades locais” condicionado ao fato dos valores das variáveis estarem entre 50 e 150, conforme foi determinado na imagem acima.



Número de unidades locais	Pessoal ocupado total	Pessoal ocupado assalariado	Salários e outras remunerações	Salário médio mensal	Número de empresas atuantes	MapaCOD_MUNICI	Sum (Σ) _Número de unidades locais
90	737	656	3018	0,9	90	0	26563
212	1142	918	3836	0,8	211	1	26563
1210	8314	6984	69498	1,8	1180	2	26563
289	2118	1897	9058	0,9	284	3	26563
74	457	385	2178	1	72	4	26563
674	6524	5838	45552	1,6	661	5	26563
131	1356	1243	8219	1,3	130	6	26563
620	3171	2439	18269	1,3	603	7	26563
264	1342	1134	9492	1,3	264	8	26563
401	1687	1247	9931	1,7	399	9	26563
2205	16507	14208	137302	1,9	2140	10	26563
39	609	576	3111	1	39	11	26563
45	195	174	1098	1,2	45	12	26563
453	2312	1967	18211	1,8	453	13	26563
827	5074	4063	29195	1,5	809	14	26563
204	870	673	4845	1,4	201	15	26563
270	1225	992	6916	1,3	267	16	26563
78	493	420	3125	1,4	78	17	26563
253	1268	1101	6785	1,3	252	18	26563
147	754	611	4929	1,7	147	19	26563
662	2602	1920	18141	1,7	654	20	26563
344	3058	2708	22428	1,5	341	21	26563
282	1755	1548	11266	1,4	282	22	26563
110	687	603	2555	0,8	110	23	26563

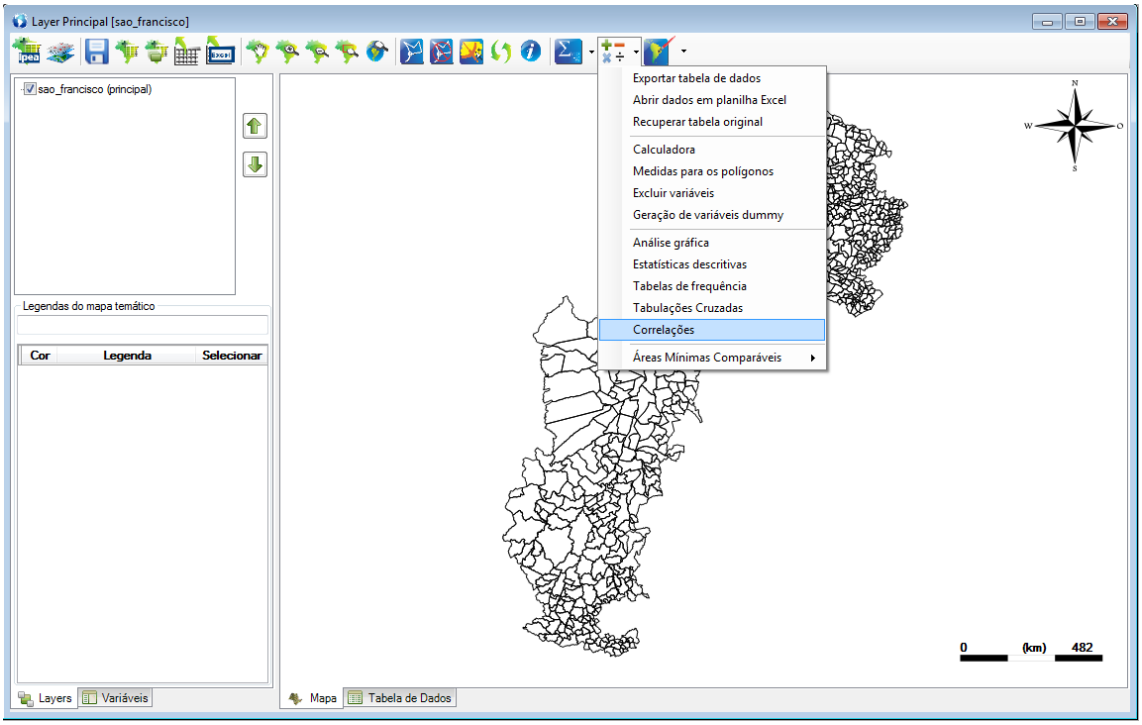
Correlações

Visão Geral

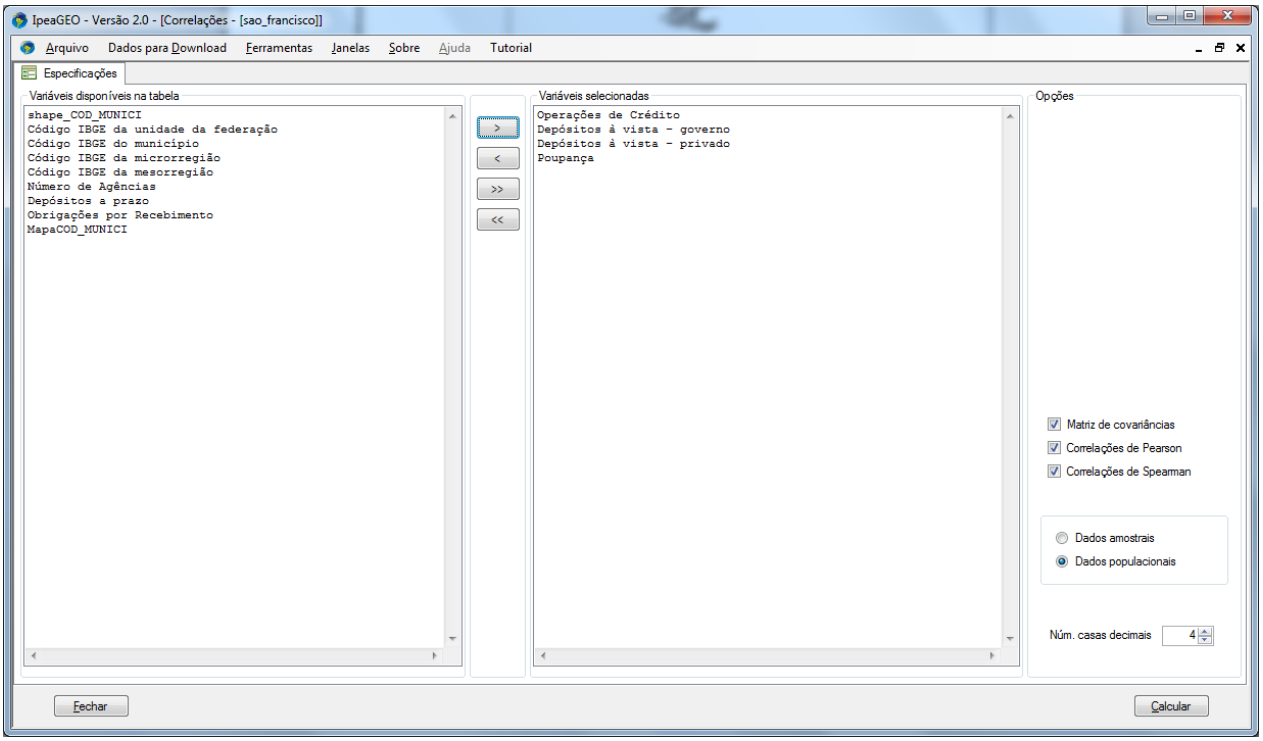
O formulário de correlações do IpeaGEO permite ao usuário verificar a existência de correlações entre as variáveis disponíveis na tabela de dados utilizada.

Exemplo

Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com sua tabela de dados (ver tópico de concatenação) clicando na opção “Análise Exploratória e manipulação de dados” e, em seguida, em “Correlações”.



O seguinte formulário irá se abrir:

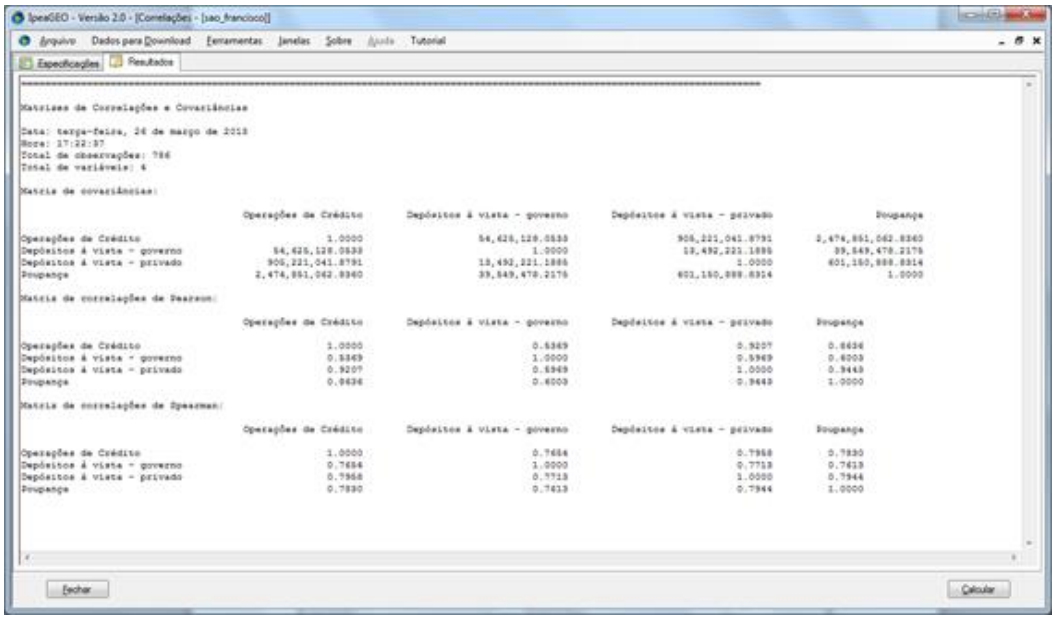


Na aba “Especificações” selecionar, dentre as variáveis disponíveis, as que deseja utilizar para verificar as correlações e clique no botão [>]. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis selecionadas.

O formulário de correlações do IpeaGEO permite calcular a matriz de variância-covariância (Matriz de Covariância) e as matrizes de correlação de Pearson e de Spearman.

O IpeaGEO também realiza a correção do viés do estimador da covariância para dados amostrais, selecionando a opção "Dados Amostrais". Se os dados forem provenientes de uma população selecione a opção “Dados Populacionais”.

Defina a quantidade de casas decimais a serem apresentadas nos resultados e clique em "Calcular" para que sejam exibidos os resultados da análise:



Estatísticas Descritivas

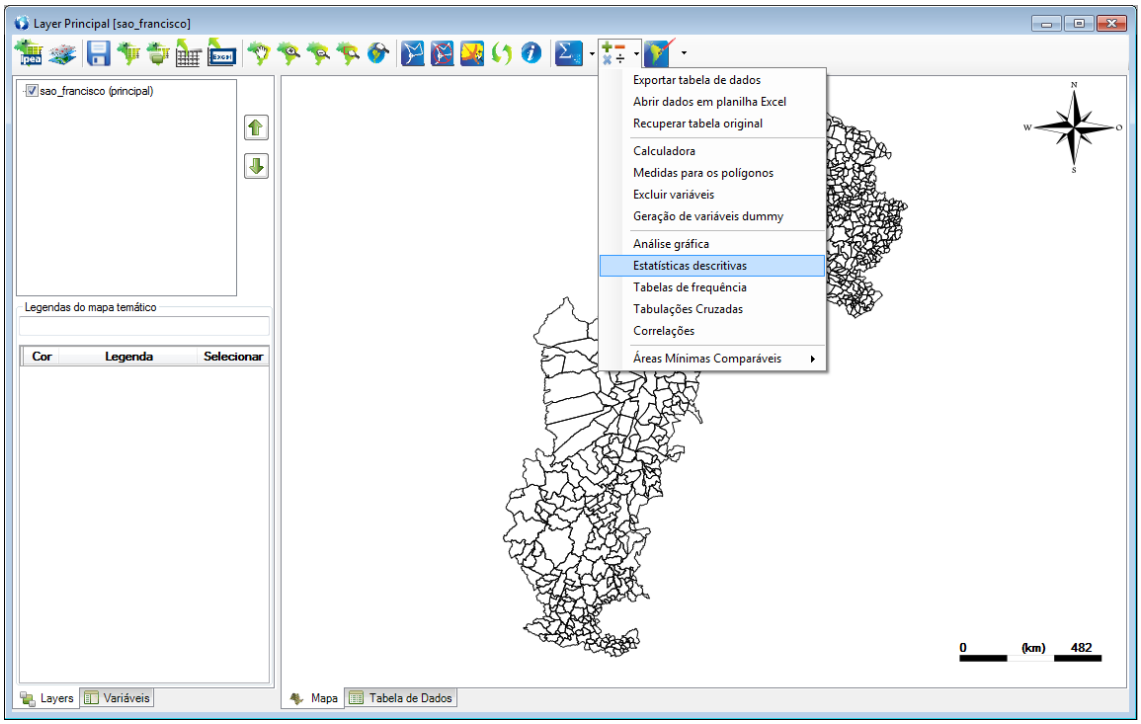
Visão Geral

As estatísticas descritivas são medidas resumo importantes para a análise exploratória dos dados. Elas permitem, por exemplo, que o pesquisador conheça o comportamento dos dados que estão sendo analisados e identifique inconsistências, presença de valores extremos (outliers), dentre outros problemas comuns que se pode ter em um conjunto de dados.

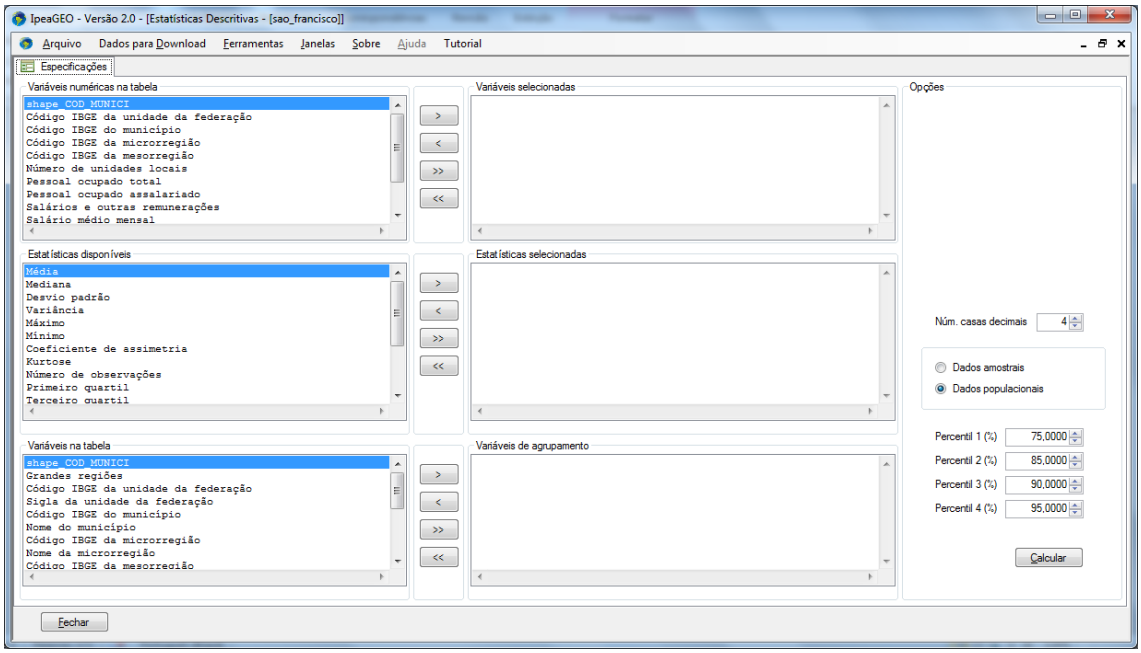
As estatísticas disponíveis no IpeaGEO são: Média, Mediana, Desvio Padrão, Variância, Máximo, Mínimo, Coeficiente de Assimetria, Coeficiente de Curtose, o tamanho amostral (Número de Observações), o primeiro (1º quartil) e terceiro (3º quartil) quartil (tendo em vista que a mediana é o segundo quartil) e o Intervalo interquartílico, que é a diferença entre o terceiro e o primeiro quartil. O IpeaGEO ainda permite ao usuário definir quatro percentis à livre escolha, chamados de Percentil 1, Percentil 2, Percentil 3 e Percentil 4.

Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar a malha digital com a tabela de dados (ver tópico de concatenação), clique na opção “Estatísticas Descritivas” localizada no botão "Análise Exploratória e Manipulação de Dados" .



O seguinte formulário irá se abrir:



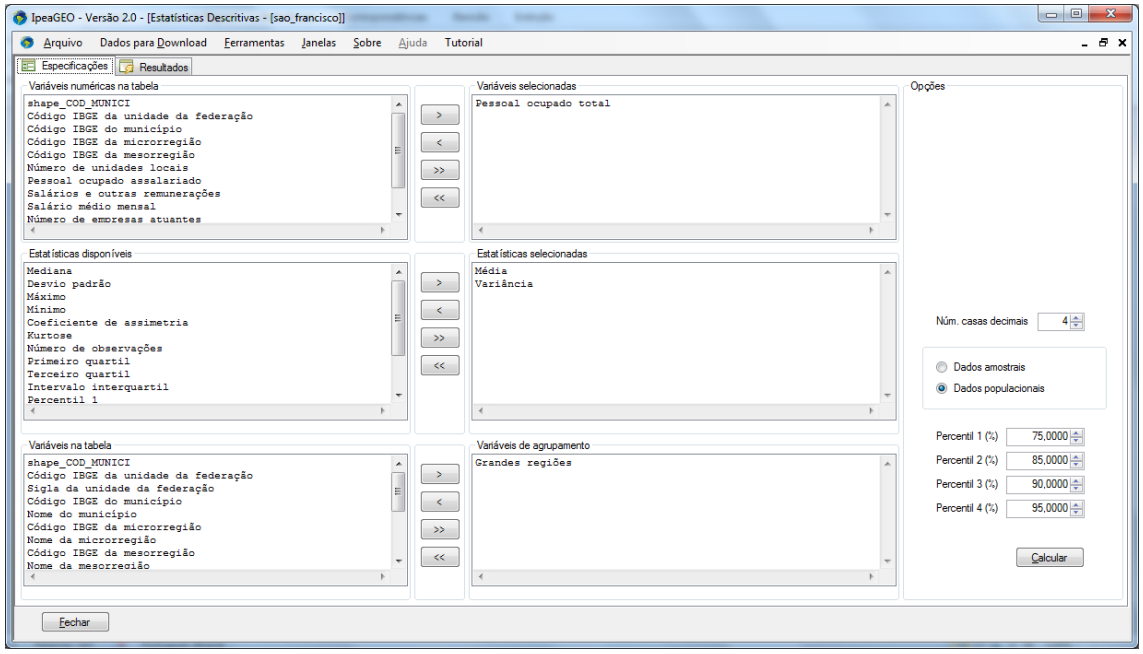
Na aba “Especificações” selecione, dentre as variáveis numéricas disponíveis, as que deseja utilizar para a análise e clique no botão [>]. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis selecionadas.

Selecione também estatísticas que deseja calcular.

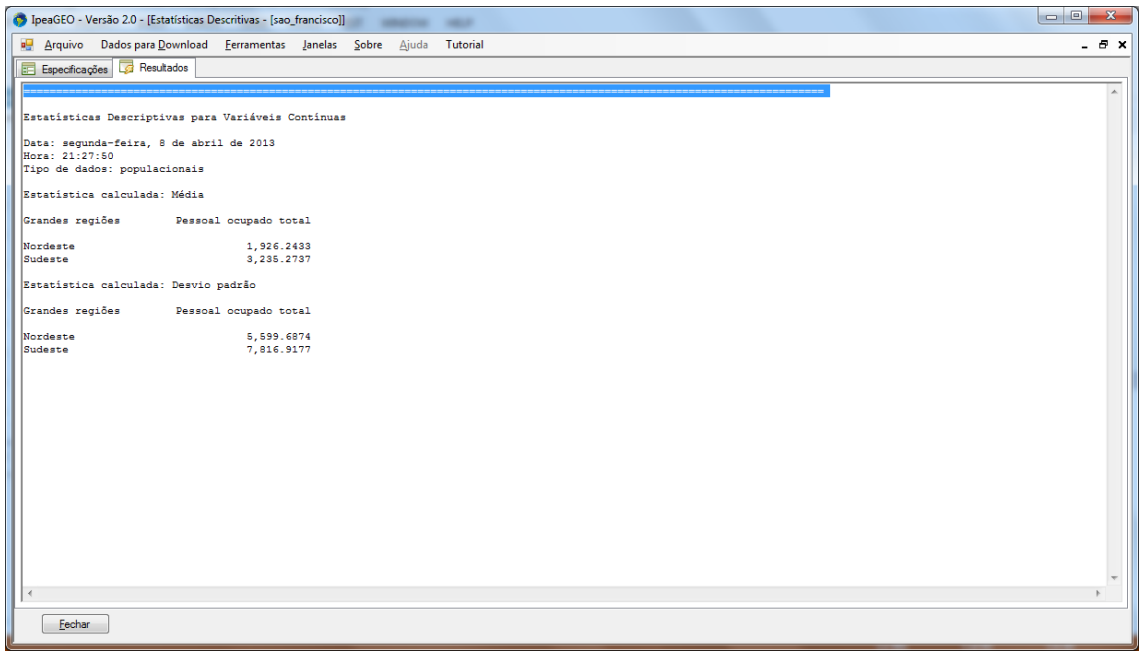
Caso seja interesse, é possível selecionar uma ou mais variáveis para serem “variáveis de agrupamento”.

Também é possível definir o número de casas decimais a serem usadas para

apresentar o resultado, indicar se os dados são oriundos de uma amostra ou de uma população (com o objetivo de correção de viés nos cálculos da variância e do desvio padrão), além de indicar percentis que se deseja calcular (entre 0 e 100), associados às estatísticas “Percentil 1” a “Percentil 4”.



Depois de selecionadas as opções desejadas, clique no botão “Calcular” para realizar as Estatísticas escolhidas. Uma nova aba chamada “Resultados” irá abrir contendo os resultados da analise.



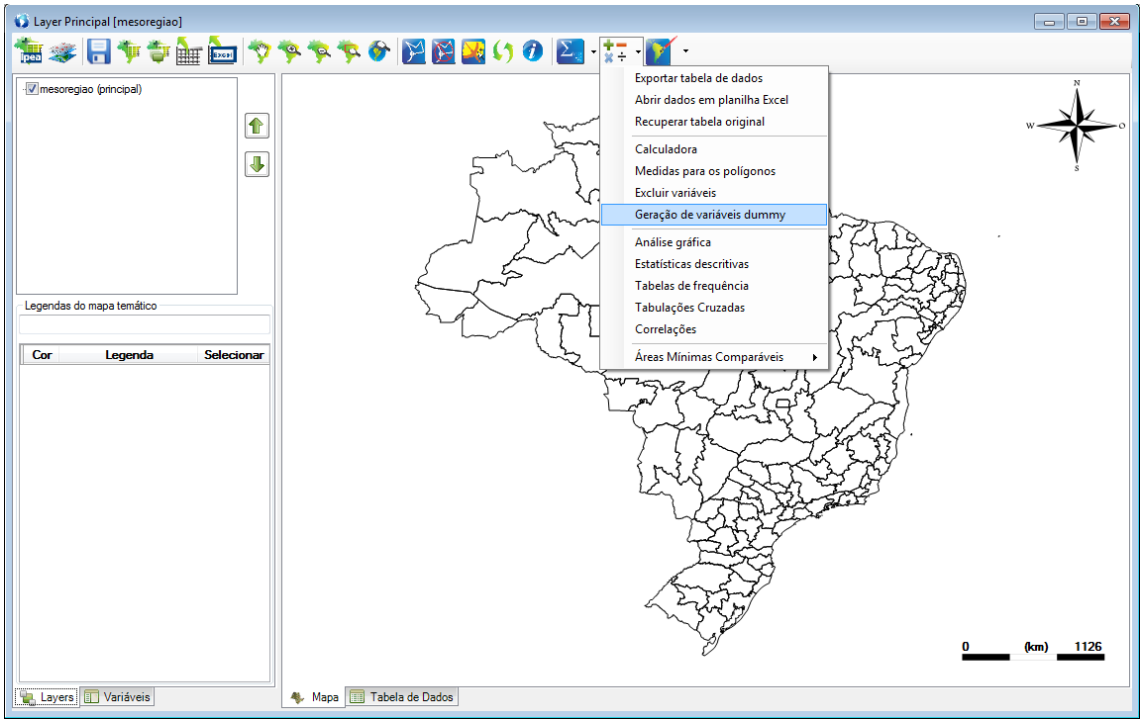
Geração de Variáveis *Dummy*

Visão Geral

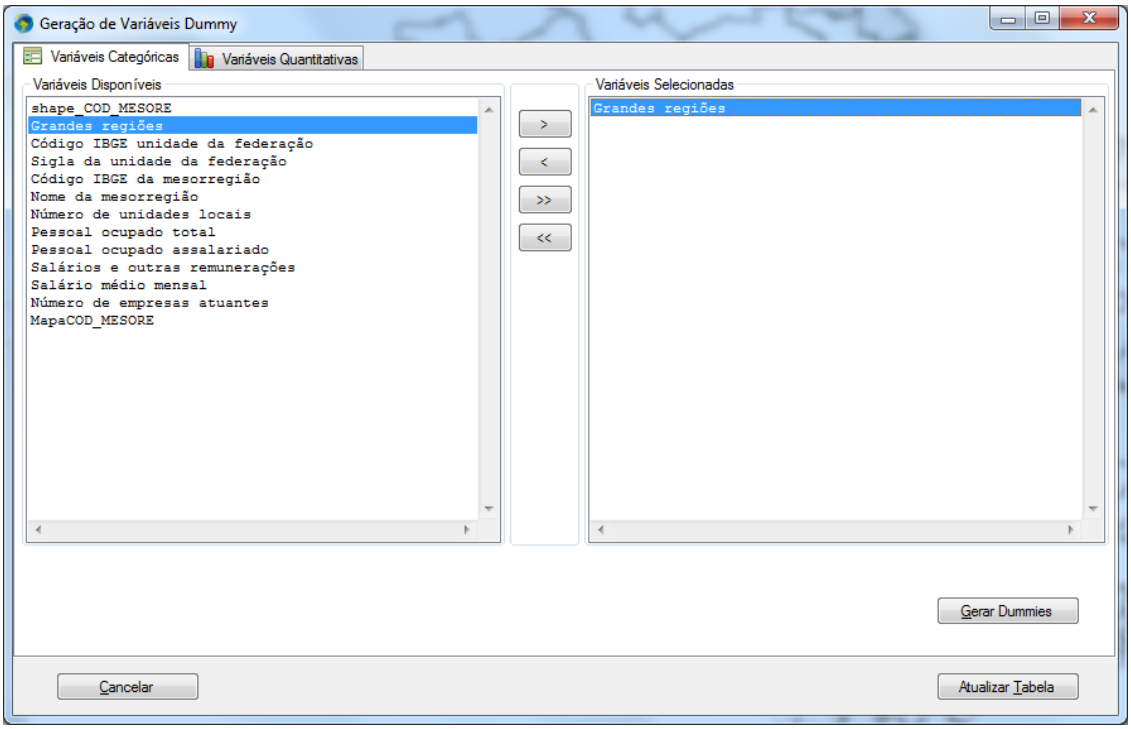
O IpeaGEO permite a geração de variáveis *dummies* para posterior utilização em qualquer outra técnica implementada no software.

Exemplo

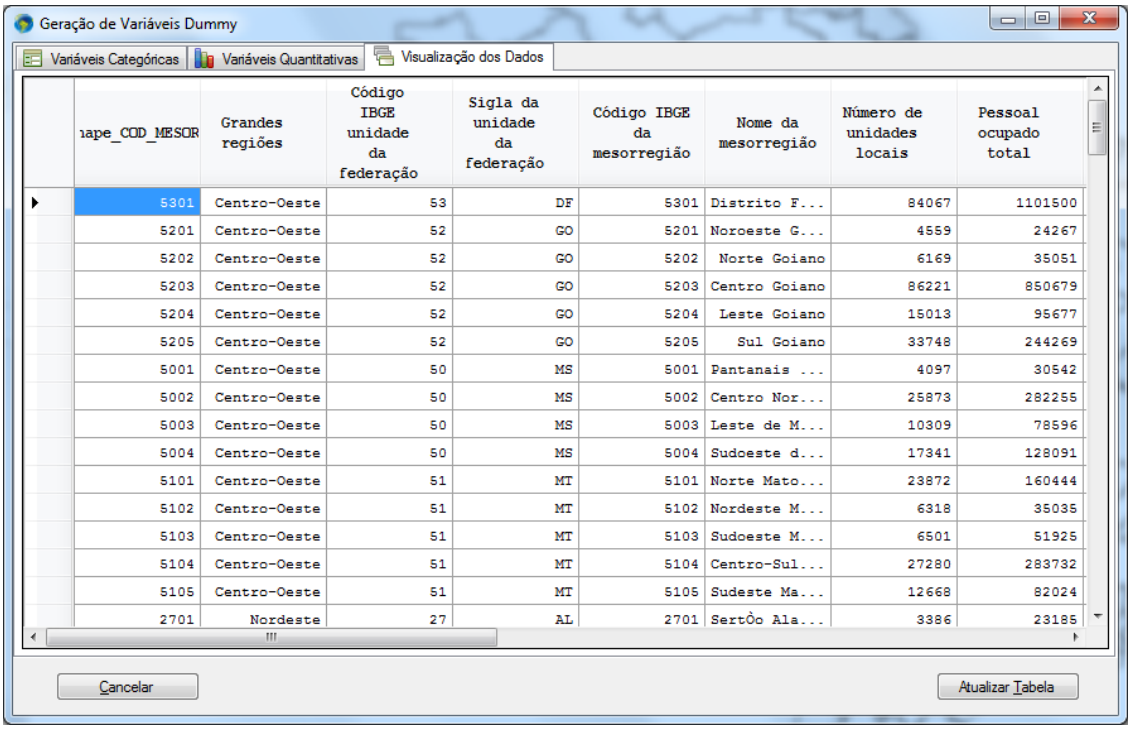
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique na opção “Análise Exploratória e manipulação de dados” e em seguida na opção “Geração de Variáveis *Dummy*”.



O seguinte formulário irá se abrir contendo as abas "Variáveis Categóricas" e "Variáveis Quantitativas":

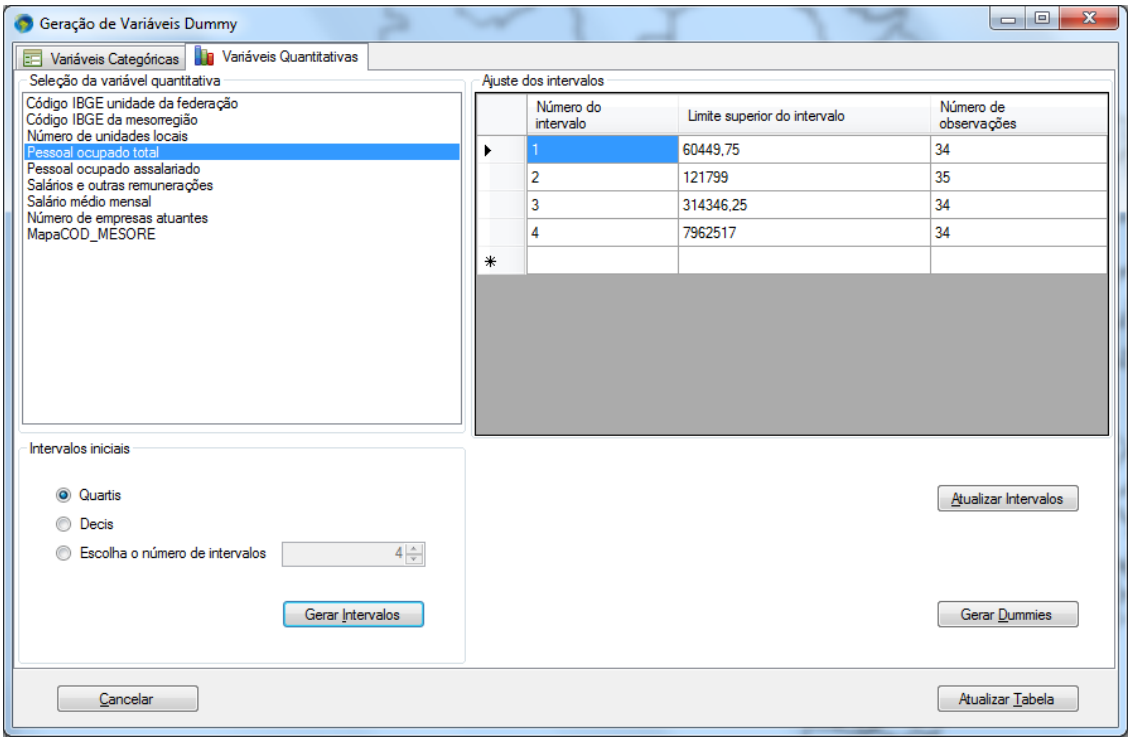


Para gerar *dummies* de variáveis categóricas basta selecionar as variáveis de interesse e clicar em "Gerar *Dummies*". Uma aba denominada "Visualização dos dados" irá aparecer contendo as *dummies* geradas nas ultimas colunas. Clique no botão "Atualizar Tabela" para que as *dummies* geradas sejam salvas definitivamente na tabela de dados. Caso a variável categórica escolhida possua mais do que 40 categorias distintas não é possível gerar *dummies*.



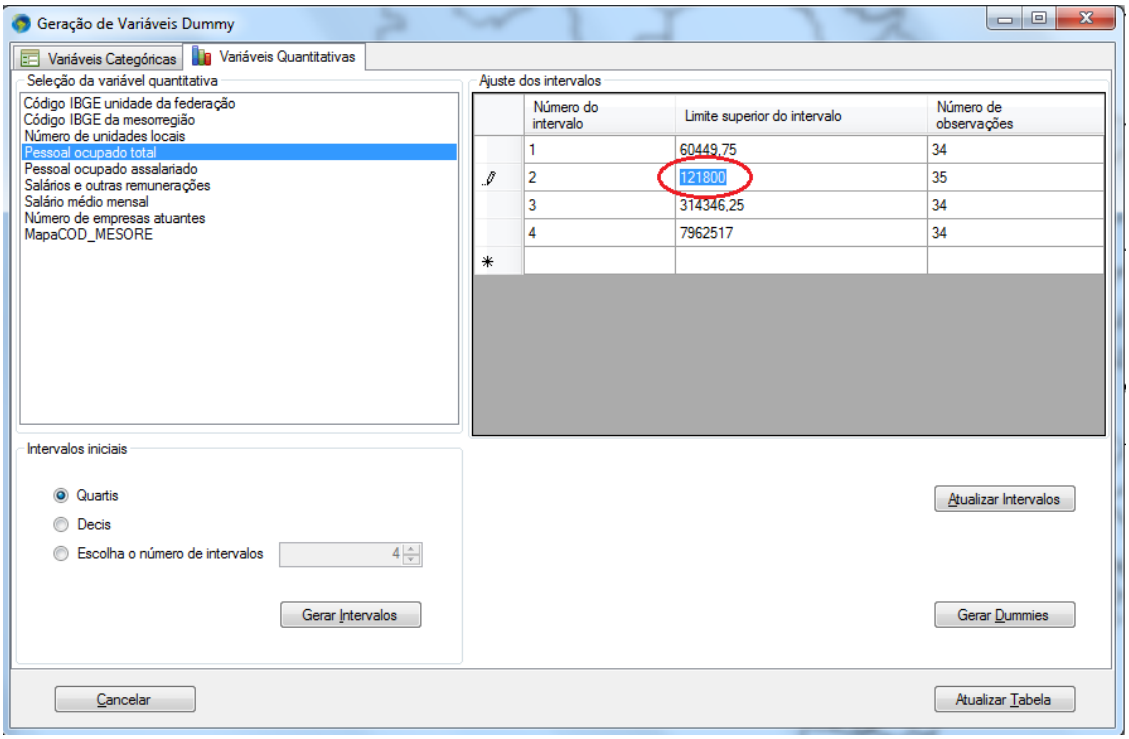
ıape_COD_MESOR	Grandes regiões	Código IBGE unidade da federação	Sigla da unidade da federação	Código IBGE da mesorregião	Nome da mesorregião	Número de unidades locais	Pessoal ocupado total
5301	Centro-Oeste	53	DF	5301	Distrito F...	84067	1101500
5201	Centro-Oeste	52	GO	5201	Noroeste G...	4559	24267
5202	Centro-Oeste	52	GO	5202	Norte Goiano	6169	35051
5203	Centro-Oeste	52	GO	5203	Centro Goiano	86221	850679
5204	Centro-Oeste	52	GO	5204	Leste Goiano	15013	95677
5205	Centro-Oeste	52	GO	5205	Sul Goiano	33748	244269
5001	Centro-Oeste	50	MS	5001	Pantanais ...	4097	30542
5002	Centro-Oeste	50	MS	5002	Centro Nor...	25873	282255
5003	Centro-Oeste	50	MS	5003	Leste de M...	10309	78596
5004	Centro-Oeste	50	MS	5004	Sudoeste d...	17341	128091
5101	Centro-Oeste	51	MT	5101	Norte Mato...	23872	160444
5102	Centro-Oeste	51	MT	5102	Nordeste M...	6318	35035
5103	Centro-Oeste	51	MT	5103	Sudoeste M...	6501	51925
5104	Centro-Oeste	51	MT	5104	Centro-Sul...	27280	283732
5105	Centro-Oeste	51	MT	5105	Sudeste Ma...	12668	82024
2701	Nordeste	27	AL	2701	Sertão Ala...	3386	23185

Para gerar *dummies* de variáveis quantitativas deve-se selecionar a variável de interesse, em seguida selecionar o método de geração das *dummies* (Quartis, Decis ou escolha o número de intervalos que deseja) e clicar em "Gerar Intervalos".



Caso os intervalos gerados sejam satisfatórios, clique em "Gerar *Dummies*". A aba denominada "Visualização dos dados" irá abrir contendo as *dummies* geradas nas últimas colunas. Clique no botão "Atualizar Tabela" para que as *dummies* geradas sejam salvas definitivamente na tabela de dados.

Caso os intervalos gerados não sejam satisfatórios, será necessário fazer a alteração manual, clicando sobre o intervalo que deseja alterar e em seguida em "Atualizar Intervalos".



Depois de encontrado o resultado satisfatório, clique em "Gerar *Dummies*", a aba "Visualização dos dados" exibirá, nas últimas colunas da tabela, as variáveis *dummies* geradas. Clique no botão "Atualizar Tabela" para que as *dummies* geradas sejam salvas definitivamente na tabela de dados.

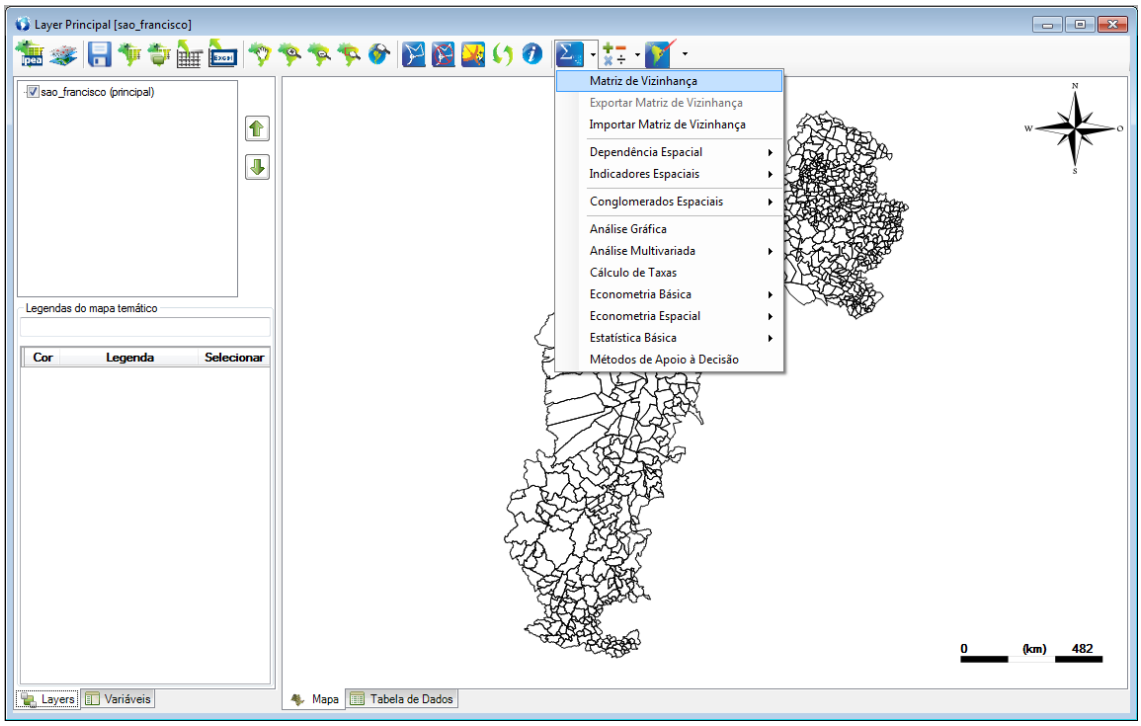
Matriz de Vizinhança

Visão Geral

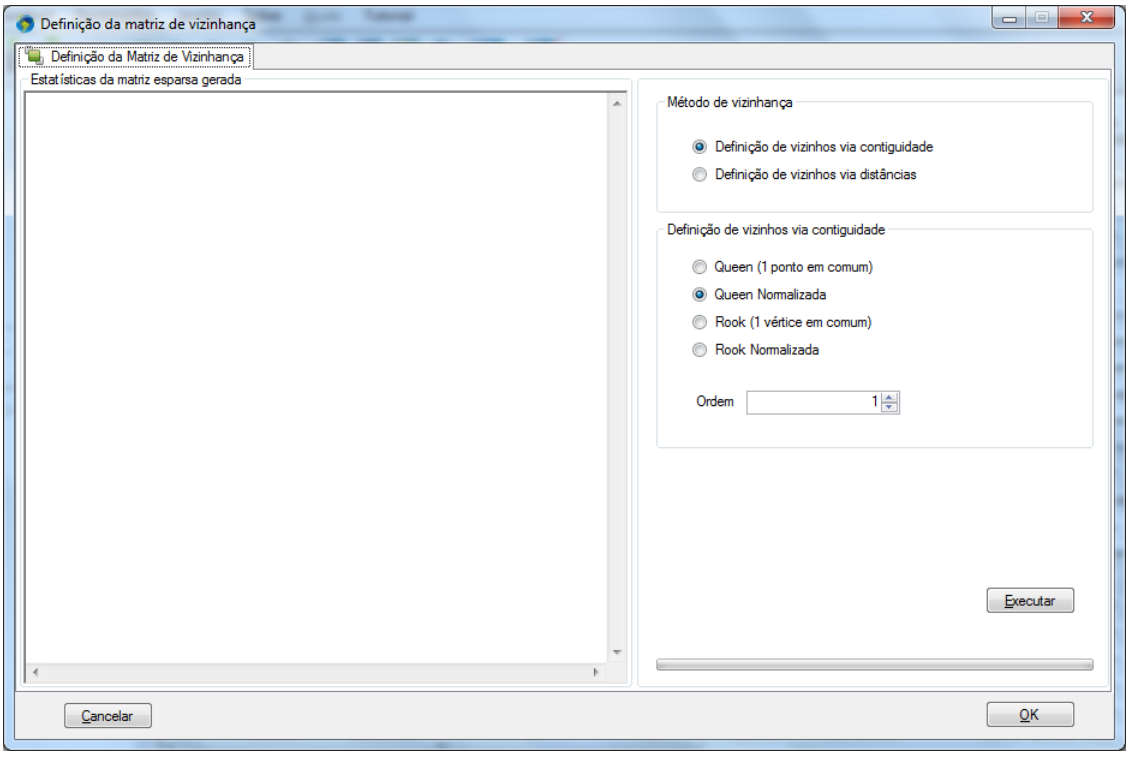
No IpeaGEO é possível gerar uma matriz de vizinhança ou ainda importar a matriz já calculada diretamente do seu computador.

Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique na opção “Matriz de Vizinhança” localizada no botão "Análise Exploratória e Manipulação de Dados" .



O seguinte formulário irá se abrir:



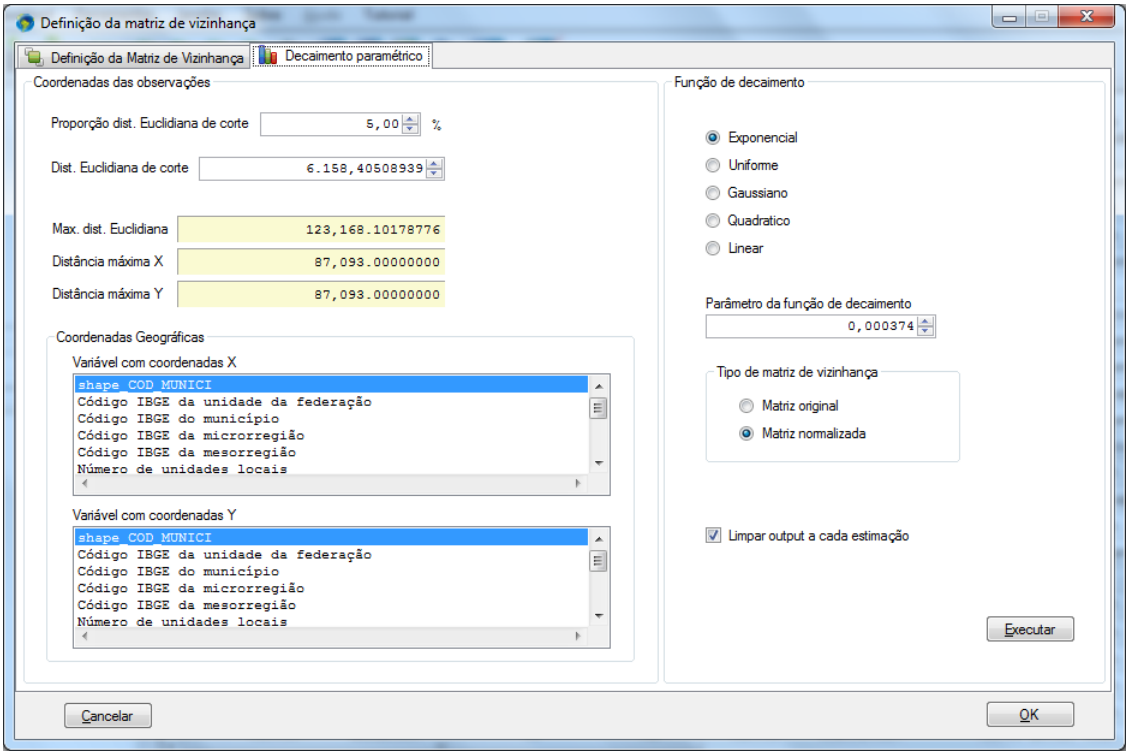
Na aba “Definição da Matriz de Vizinhança” selecione o Método de vizinhança que deseja aplicar. Caso seja escolhido o método “Definição de vizinhos via contiguidade” selecione uma das seguintes definições de vizinhos:

- Queen (1 ponto em comum);

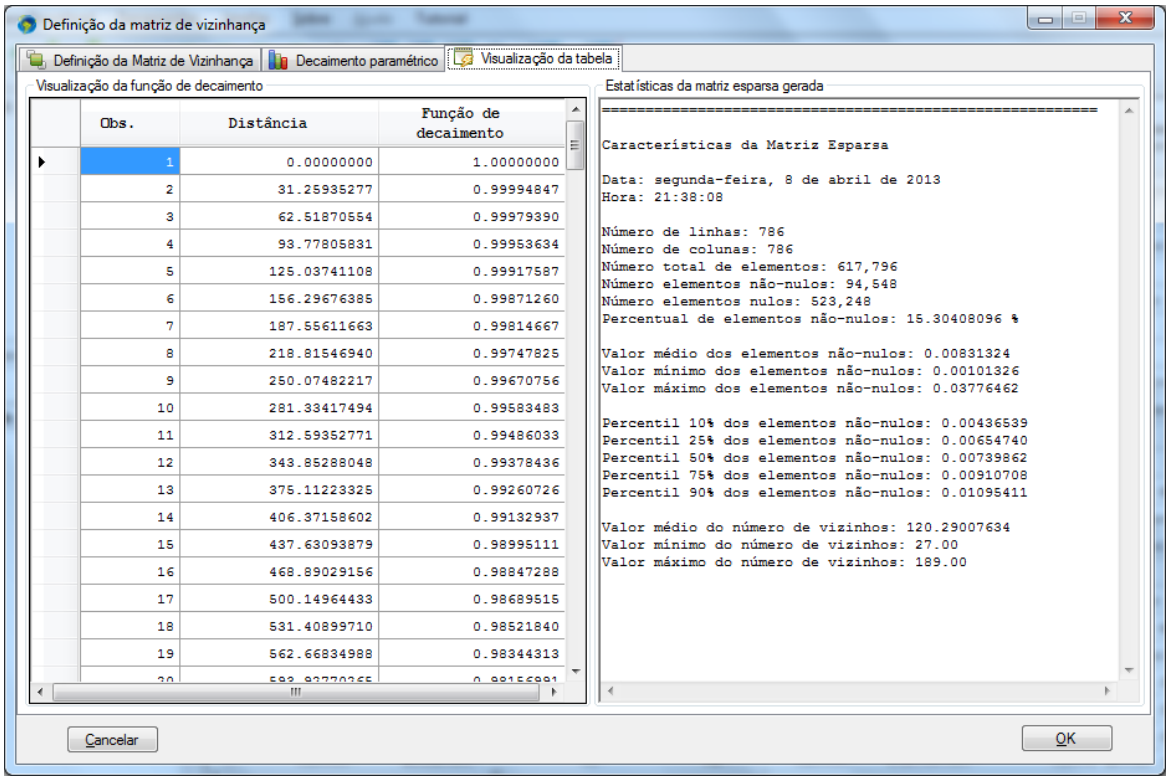
- Queen Normalizada;
- Rook (1 vértice em comum);
- Rook Normalizada.

Selecionar também a ordem desejada.

Caso seja escolhido “Definição de vizinhos via distâncias” uma nova aba denominada “Decaimento Paramétrico” irá se abrir:



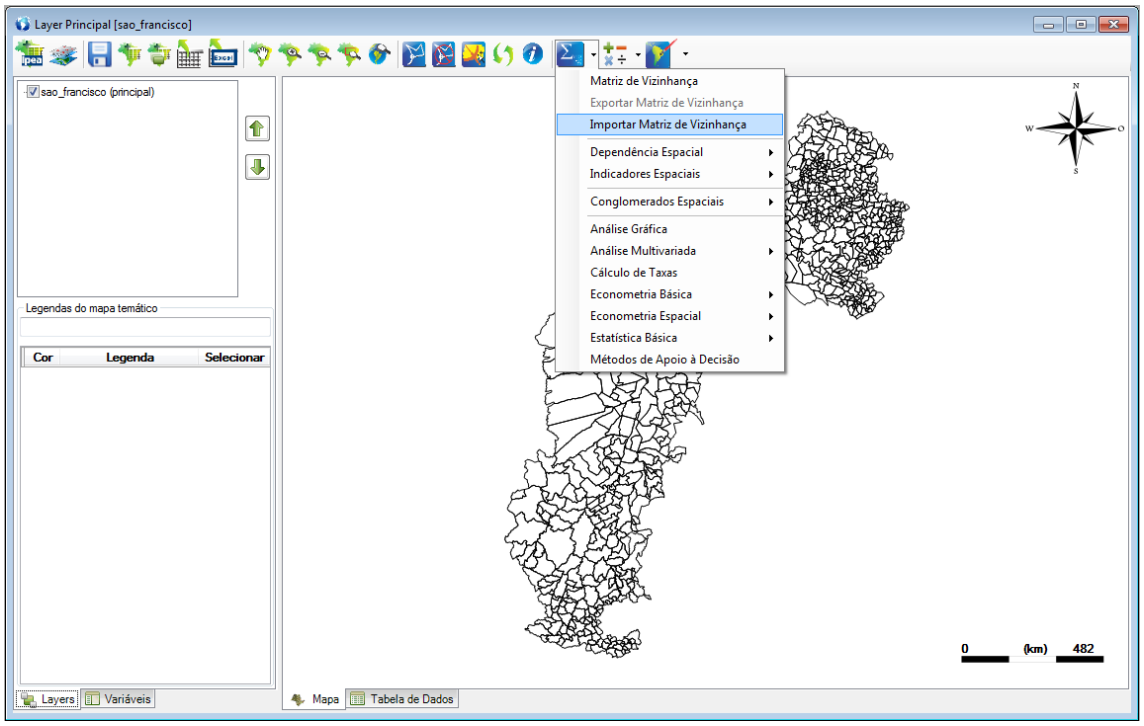
Depois de assinaladas as opções desejadas, clicar no botão “Executar”. Na parte a direita do formulário aparecerá as estatísticas da matriz esparsa gerada.



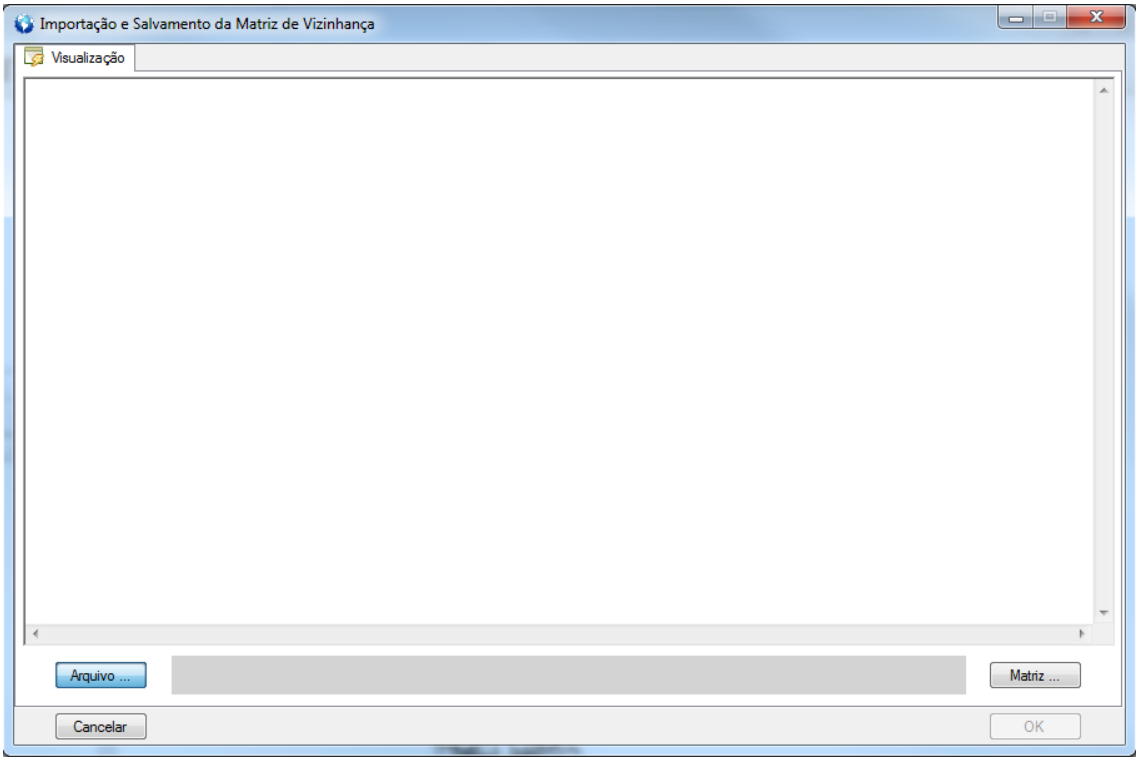
Por último, clicando no botão "Ok" a matriz de vizinhança é atualizada, e os novos cálculos serão feitos com base na nova matriz.

- Importar matriz de vizinhança

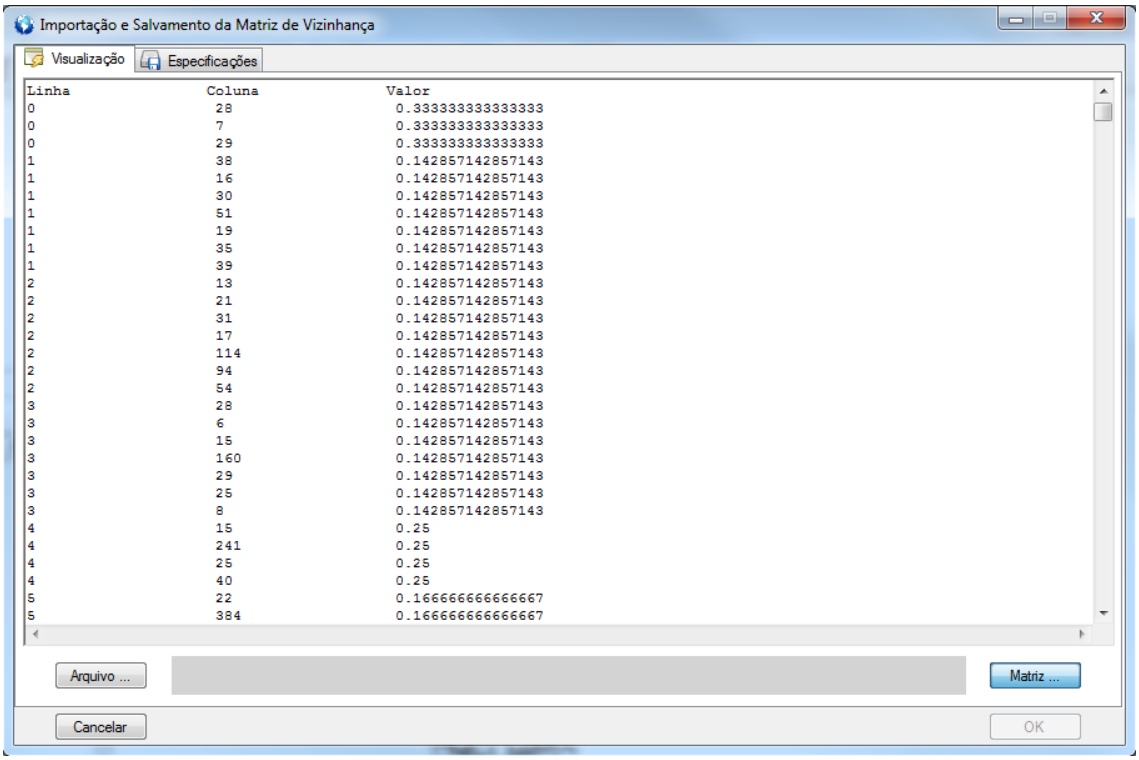
Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital com sua tabela de dados (ver tópico de concatenação), clique na opção “Importar Matriz de Vizinhança” localizada no ícone “Ferramentas” .



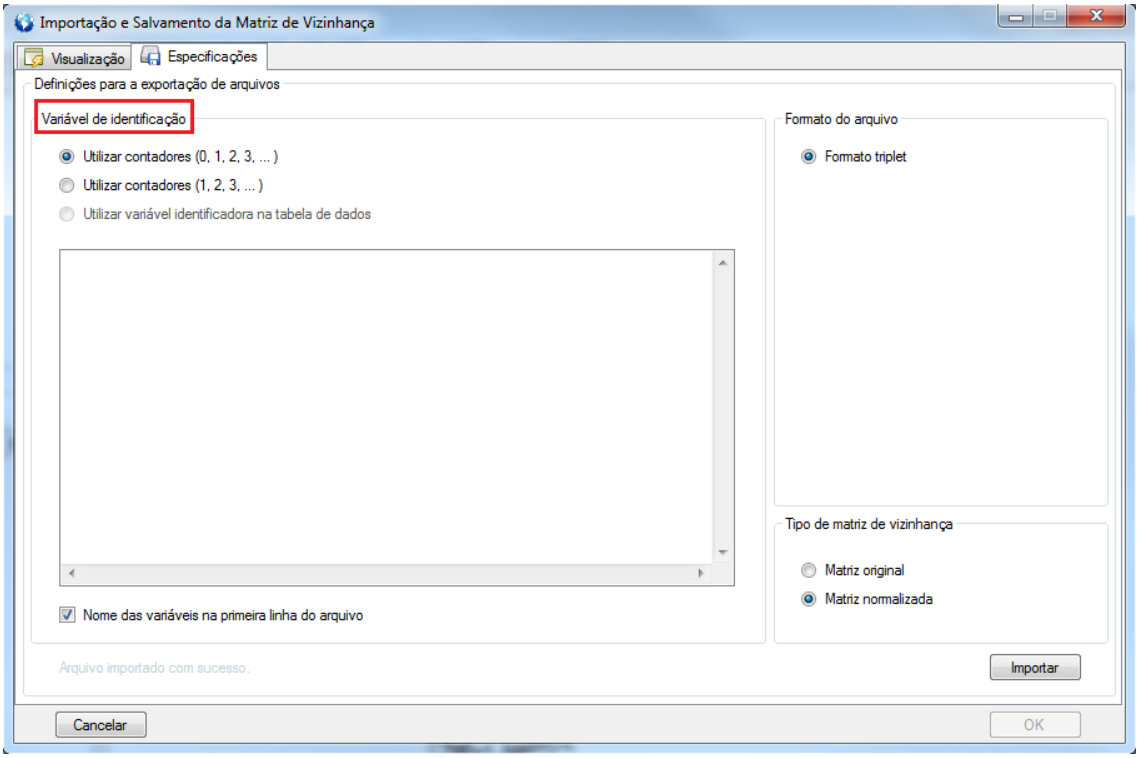
O seguinte formulário irá se abrir. Ao clicar no botão “Arquivo” selecione a matriz que deseja importar para o IpeaGEO.



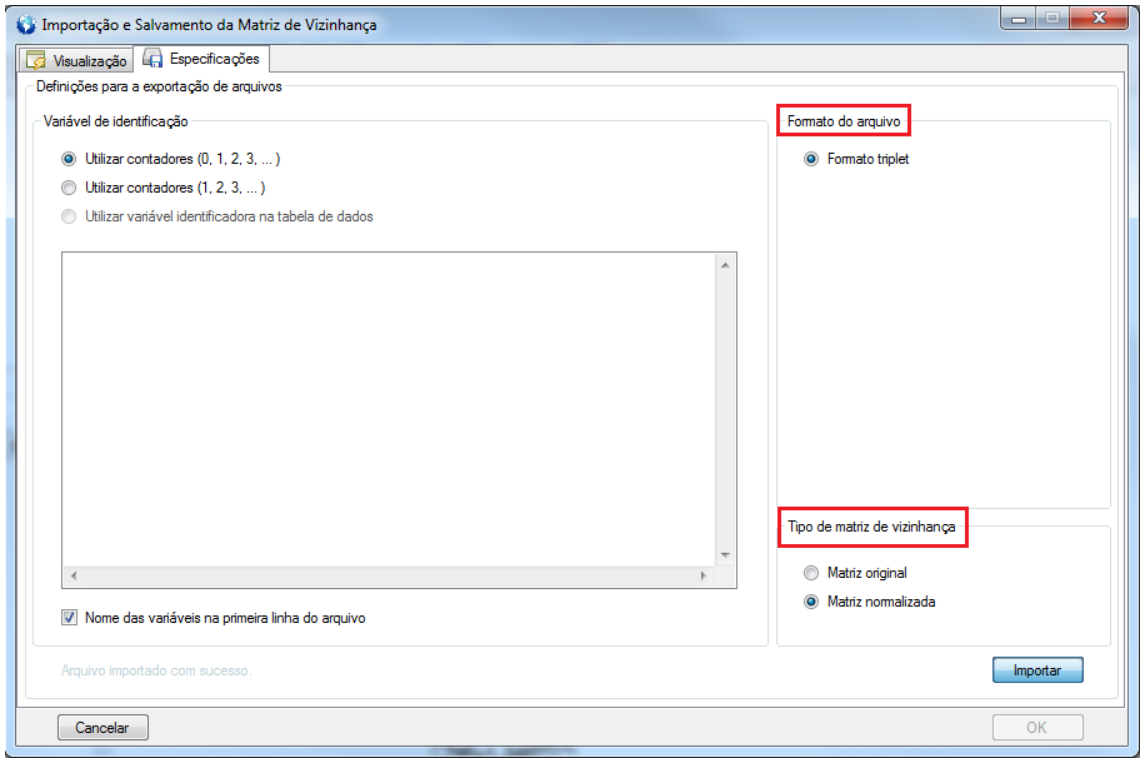
Depois de selecionada a matriz, aparecerá na aba “Visualização” os dados referentes ao arquivo importado. Em seguida, clique no botão “Matriz...”.



Na nova aba “Especificações”, marque a opção desejada para “Variável de identificação”.

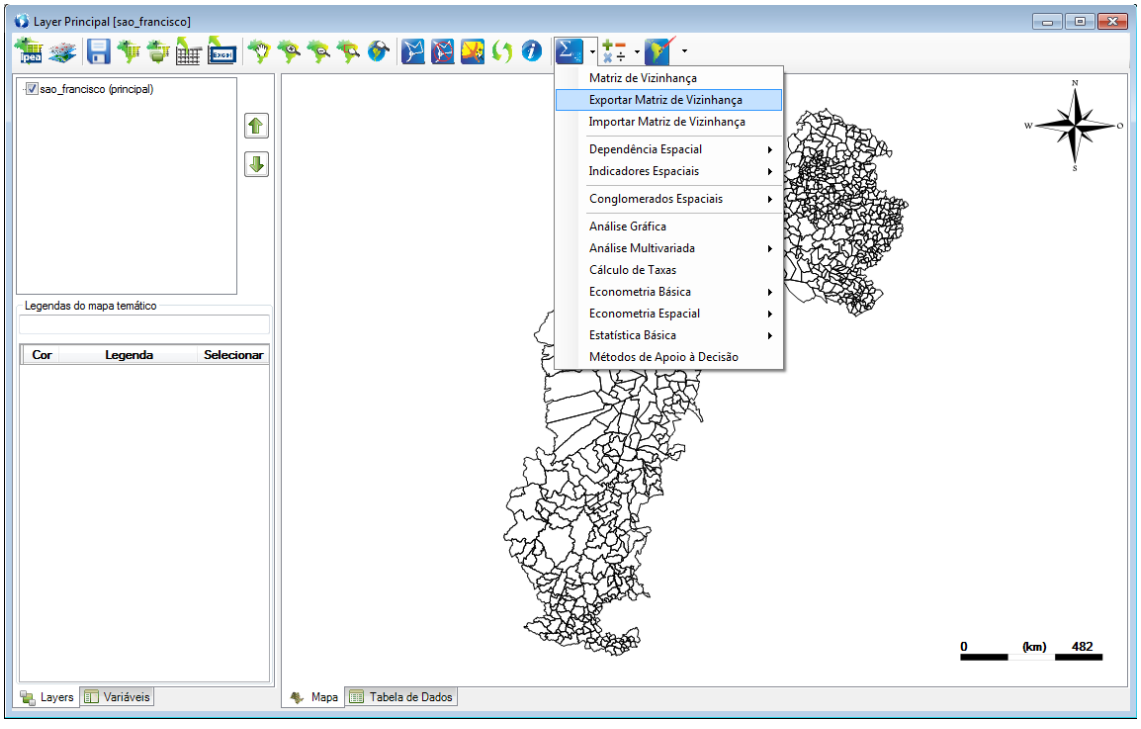


Estão disponíveis também as opções para “Formato do arquivo” e “Tipo de matriz de vizinhança”. Por fim, clique no botão “Importar” para utilizar as ferramentas do IpeaGEO com base na nova matriz de vizinhança.

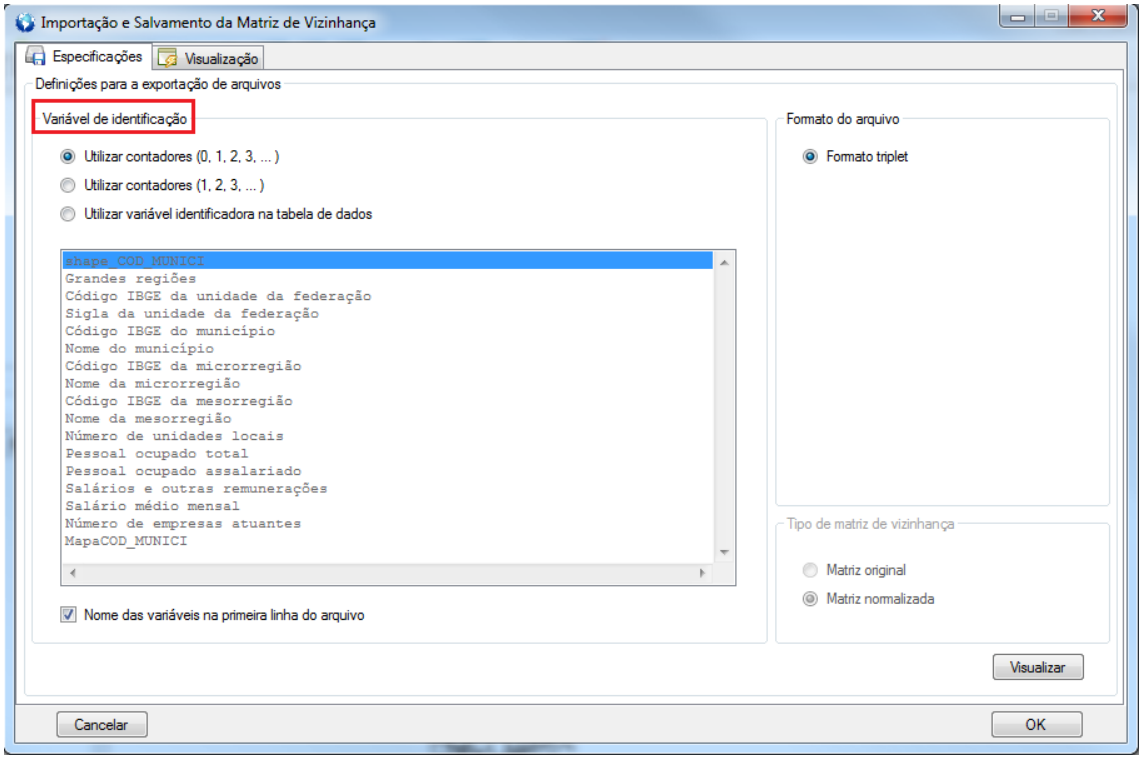


- Exportar matriz de vizinhança

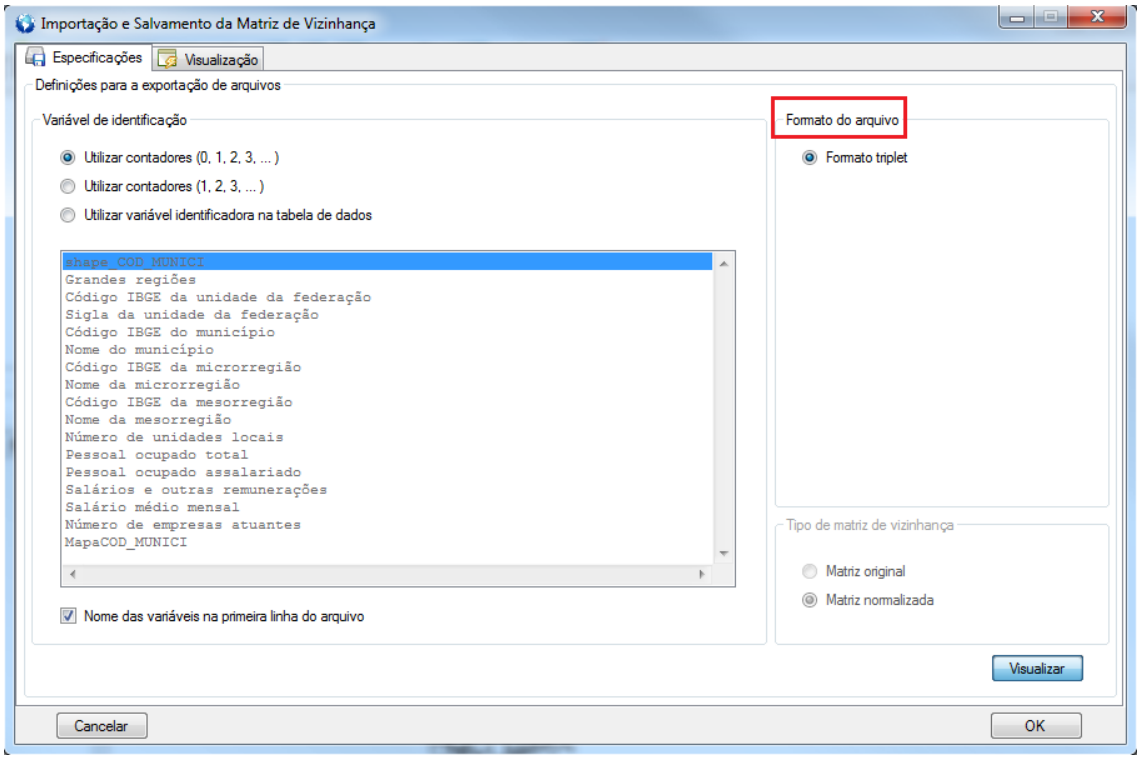
Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital com sua tabela de dados (ver tópico de concatenação), clique na opção “Exportar Matriz de Vizinhança” fixada no ícone “Ferramentas” . A opção “Exportar Matriz de Vizinhança” fica disponível após a matriz de vizinhança ser gerada, como explicado nos tópicos anteriores deste tutorial.



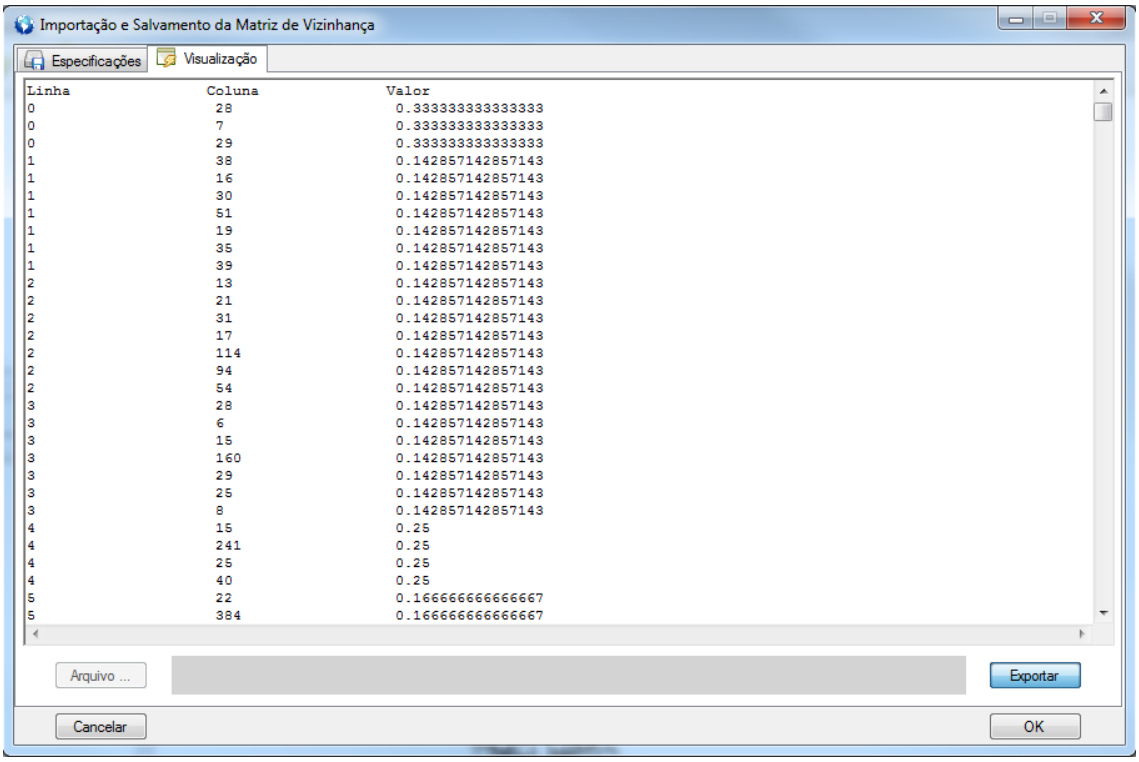
O seguinte formulário irá se abrir. Na aba “Especificações” escolha a forma como deseja realizar a identificação da variável.



Neste formulário as opções para “Tipo de matriz de vizinhança” não ficam disponíveis. Para alterá-la, gere uma nova matriz de vizinhança. A matriz será exportada no formato triplet, como indicado na seção “Formato do arquivo” abaixo. Em seguida, clique no botão “Visualizar”.



Na aba “Visualização” estão contidas as informações da matriz a ser exportada. Ao clicar no botão “Exportar” aparecerá o local no qual se deseja salvar o arquivo. Por fim, clique no botão “Ok”.



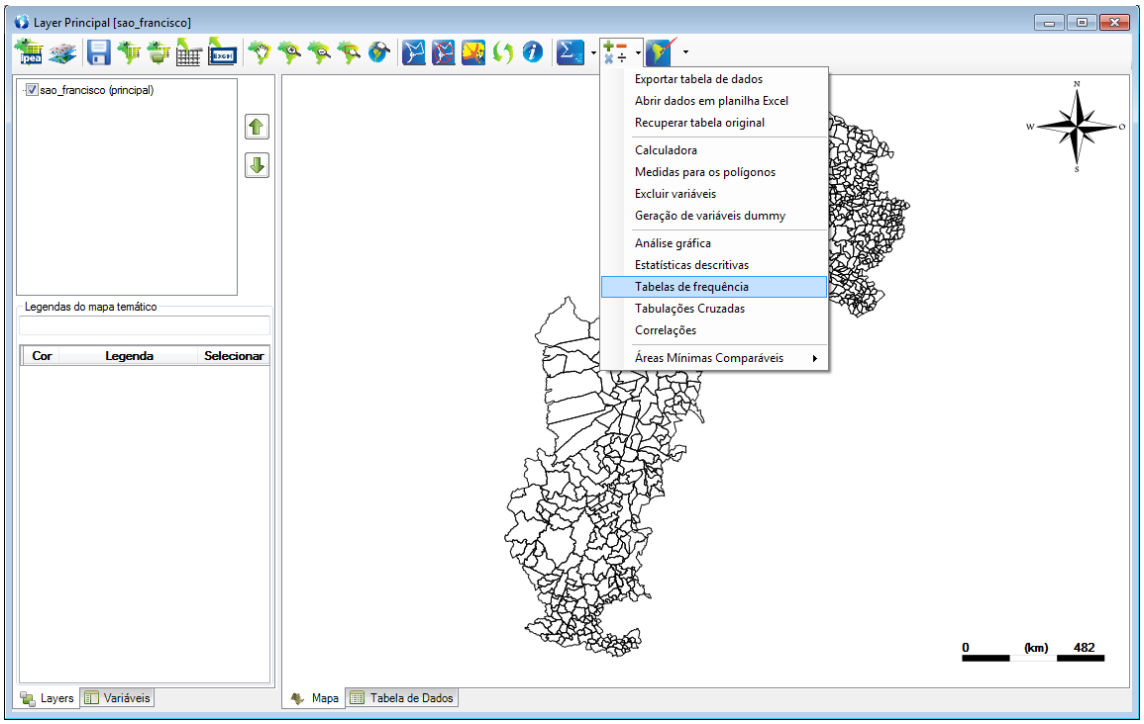
Tabelas de Frequência

Visão Geral

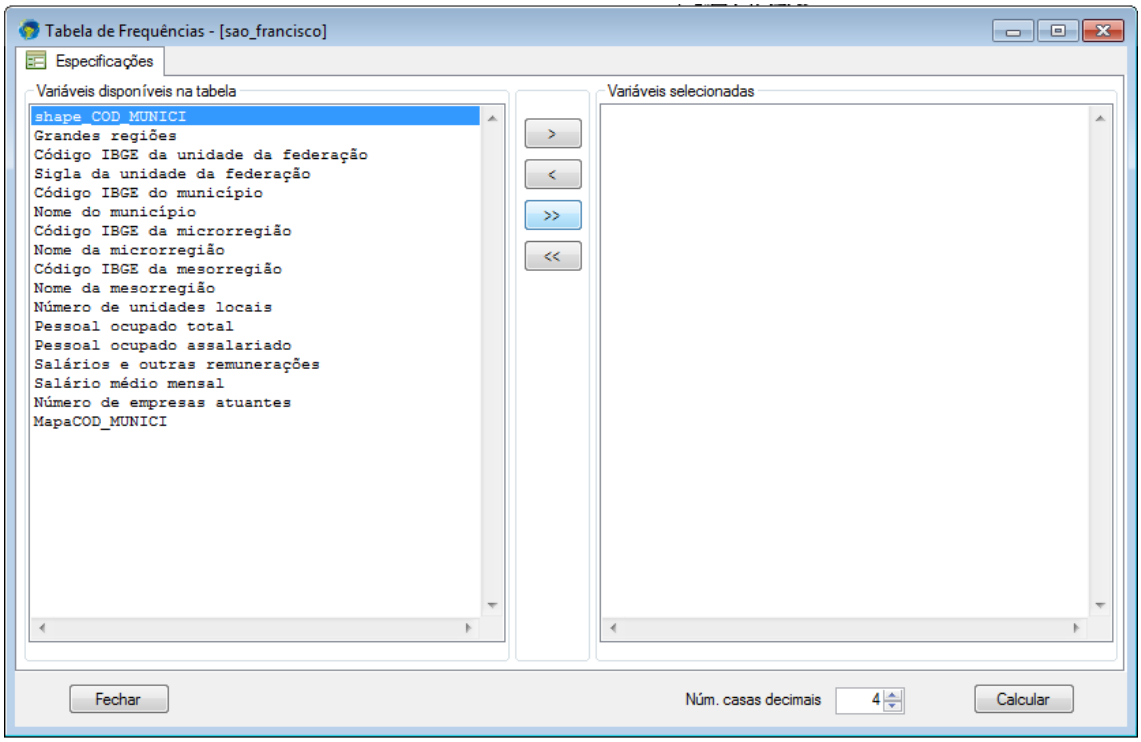
As tabelas de frequência permitem ao usuário verificar, em seu conjunto de dados, a existência de valores *missings*, *outliers*, isto é, a distribuição dos dados em que está trabalhando.

Exemplo

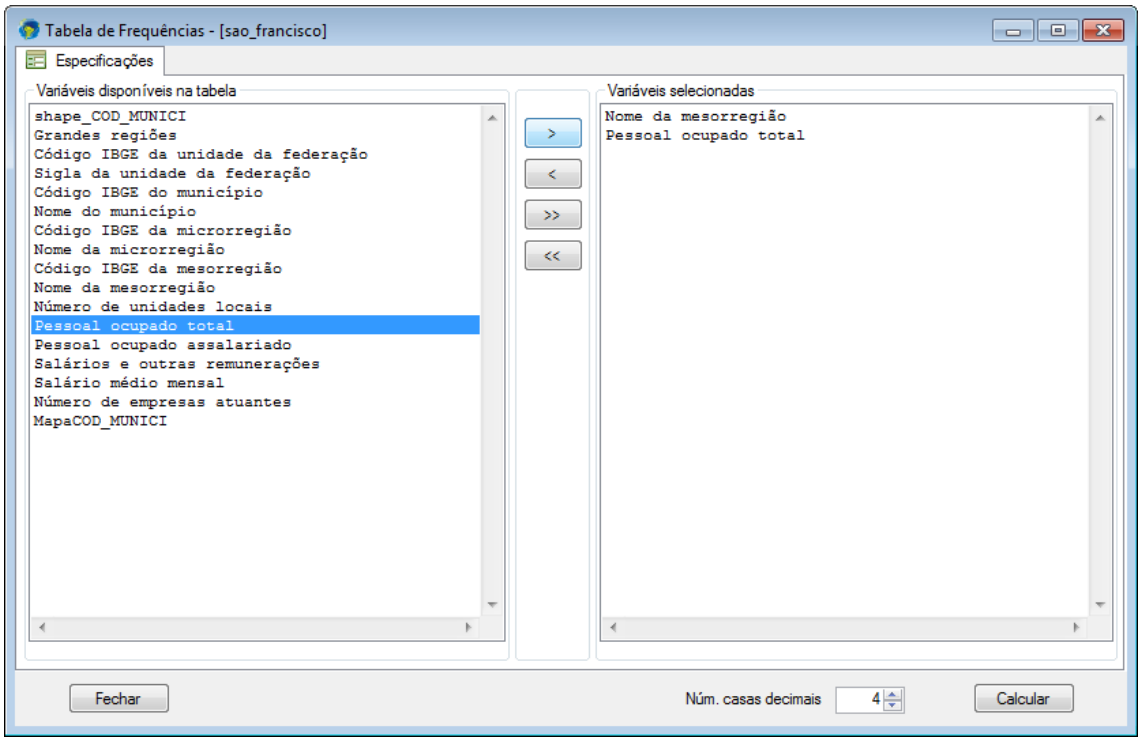
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clicar na opção “Tabelas de Frequência”, presente dentro do ícone "Análise Exploratória e Manipulação de Dados" .



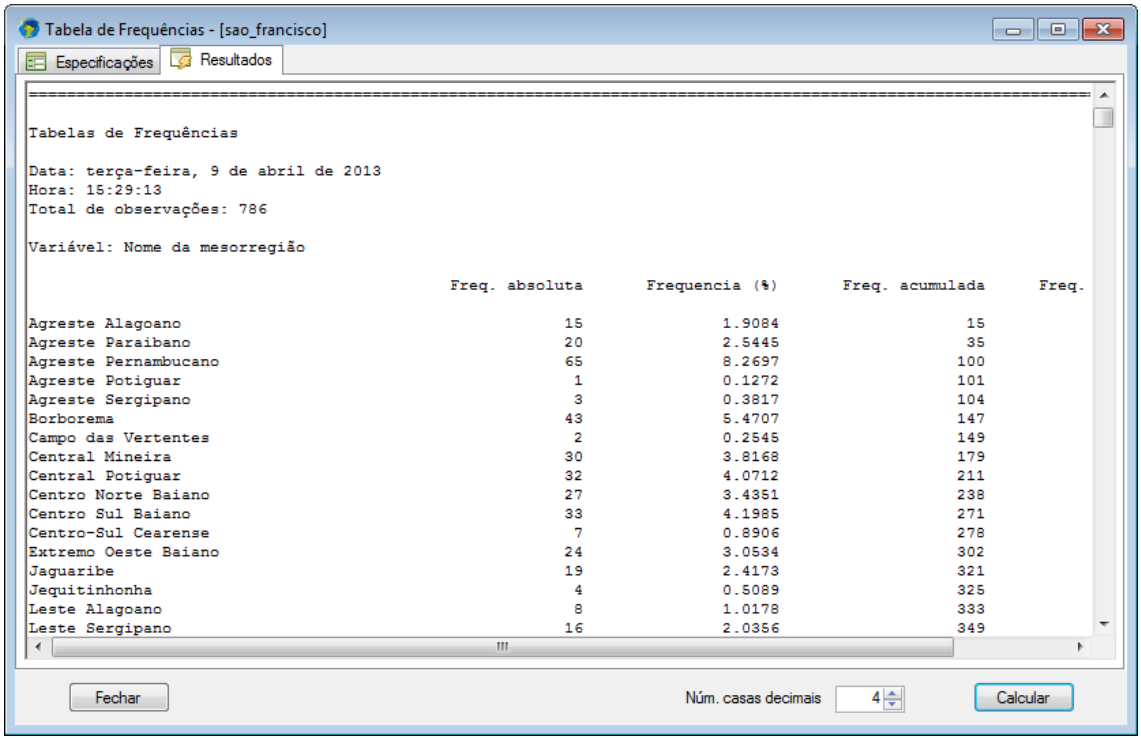
O seguinte formulário irá se abrir:



Na janela aberta, selecione as variáveis que se deseja saber as frequências, clicando sobre elas e em seguida clicando em [>] ou [>>] (caso deseje usar todas as variáveis). É possível escolher a quantidade de casas decimais para a frequência relativa (percentuais).



Após serem feitas todas as escolhas, clique em “Calcular”. Os resultados serão exibidos em uma aba com o título “Resultados”.



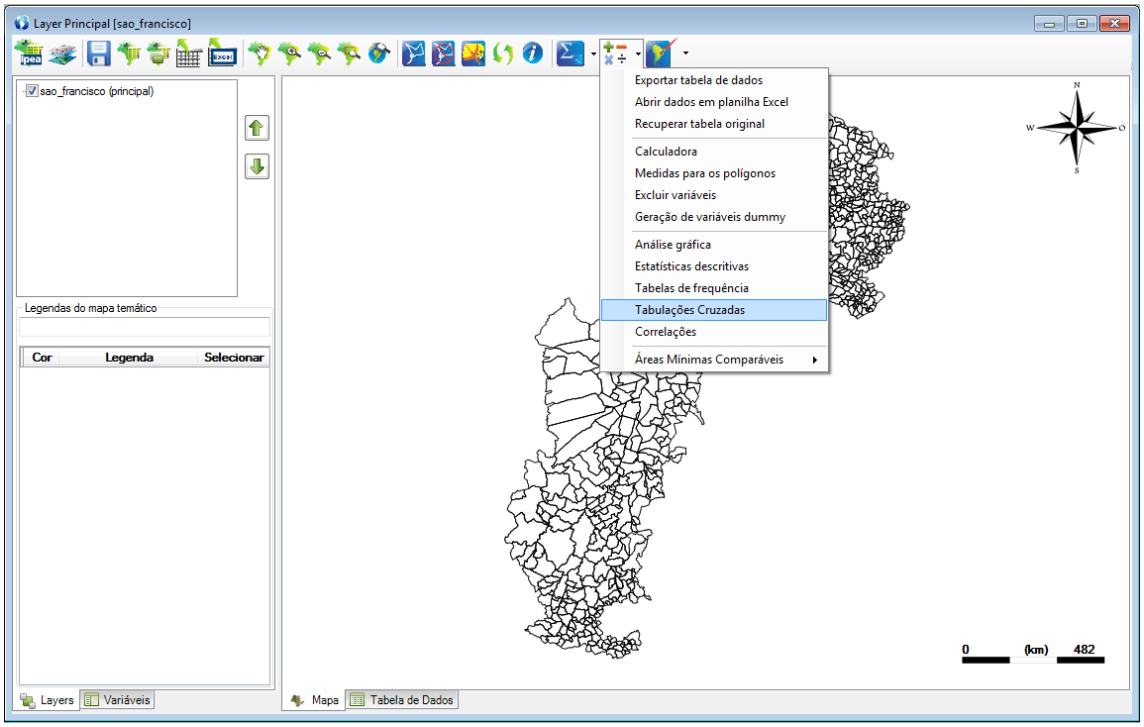
Tabulações Cruzadas

Visão Geral

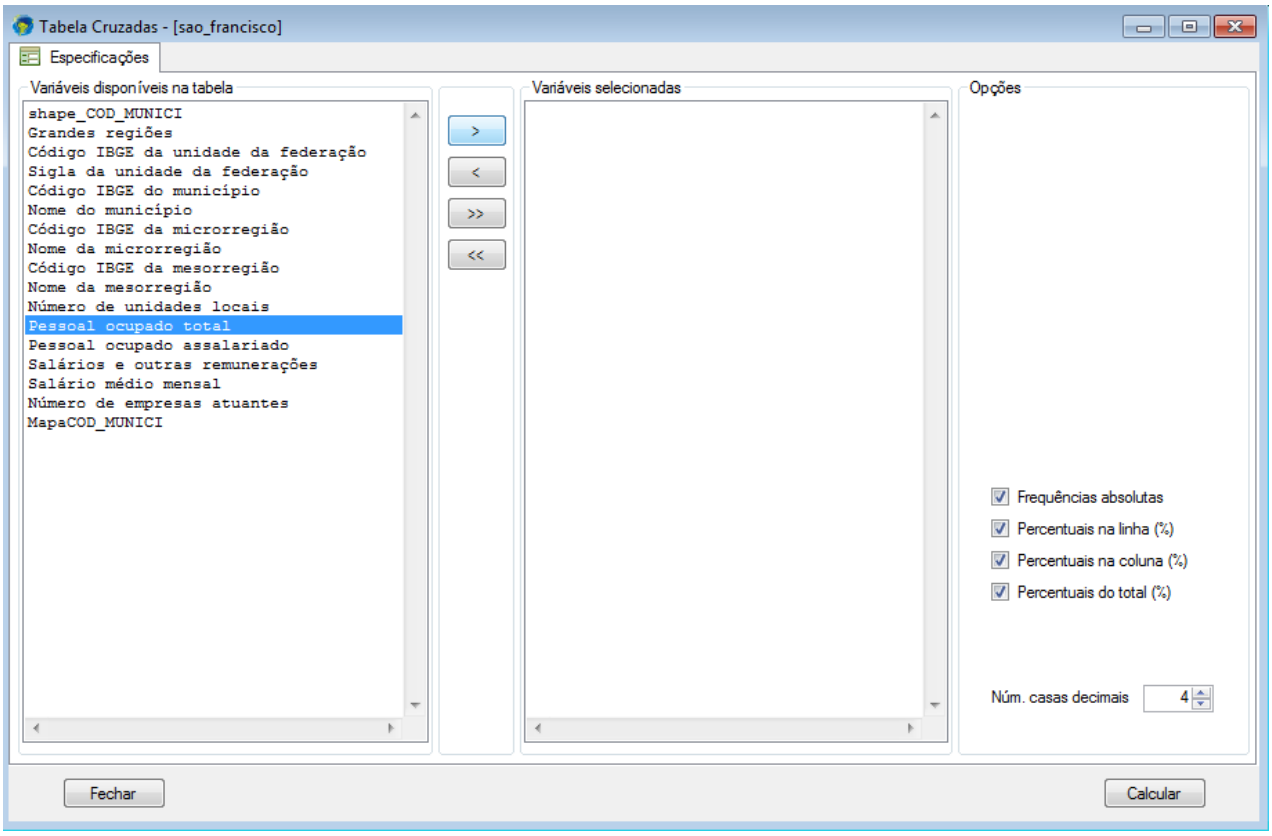
As tabulações cruzadas permitem ao usuário verificar a distribuição de frequências conjunta de duas variáveis do seu conjunto de dados.

Exemplo

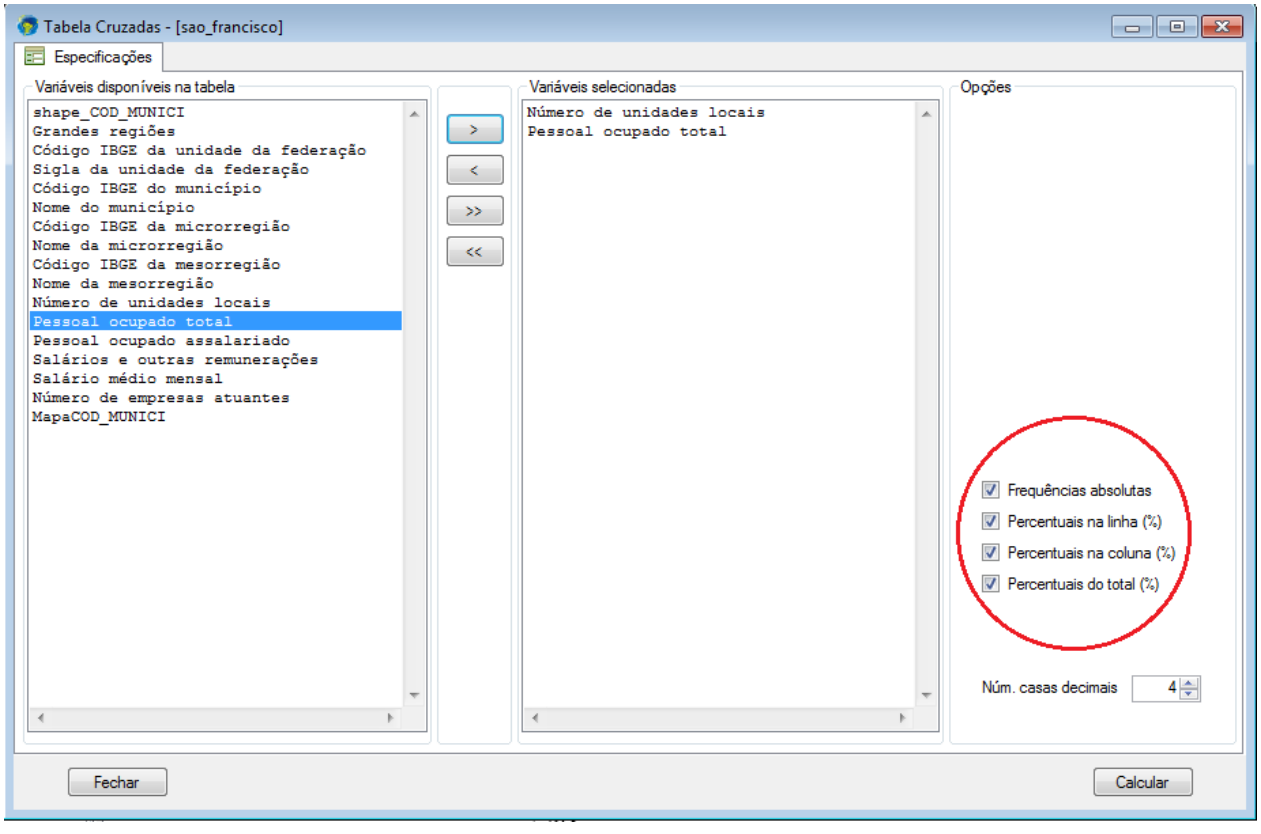
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique na opção “Tabulações Cruzadas” localizada na caixa “Análise Exploratória e Manipulação de Dados” .



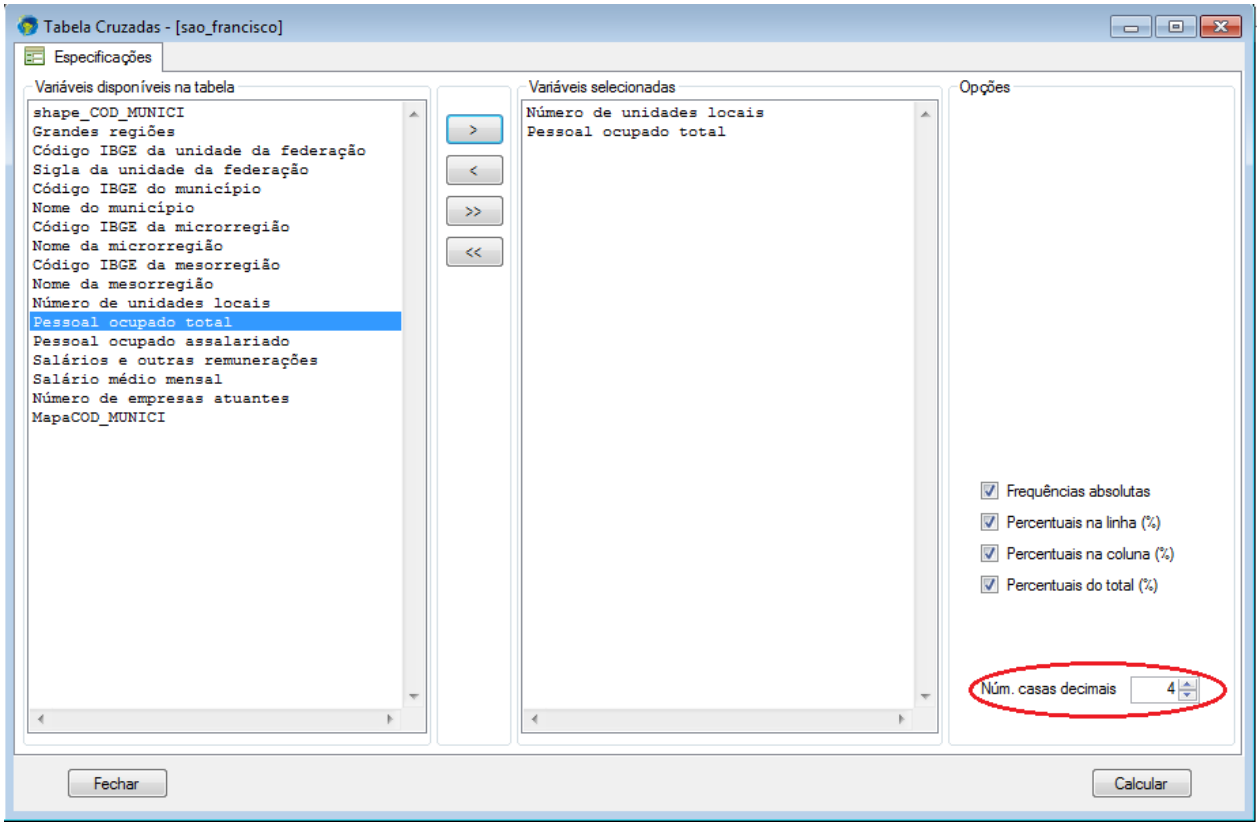
Usando o botão > é possível selecionar uma variável. O botão >> seleciona todas as variáveis disponíveis.



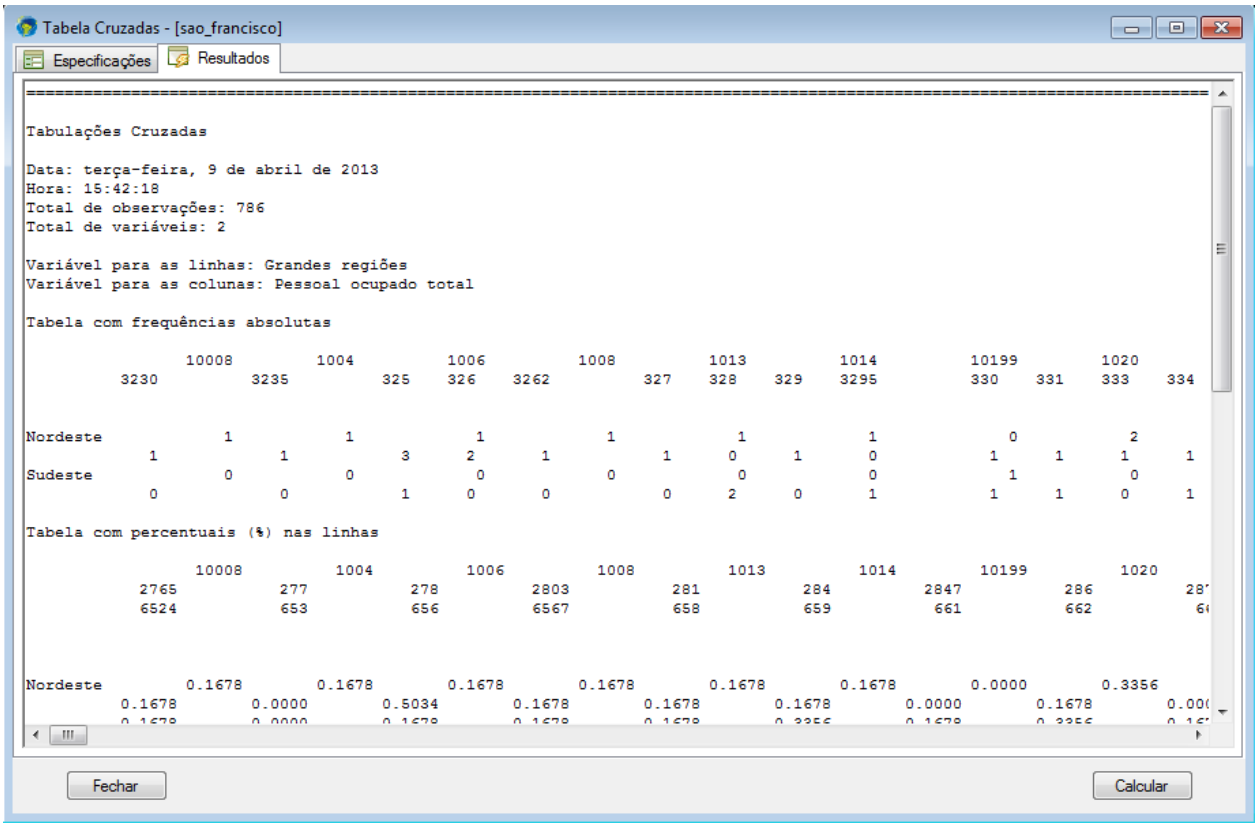
Para efetuar o cálculo é preciso selecionar duas variáveis. No lado inferior direito estão localizadas as opções que permitem acrescentar ao resultado as frequências absolutas e os percentuais nas linhas, colunas e do total.



Também é possível escolher a quantidade de números decimais para formatar o resultado.



Para gerar o resultado, clique no botão “Calcular”.



Analise Multivariada

Analise de Clusters

Visão Geral

A análise de clusters ou de agrupamento é uma técnica estatística multivariada que tem por objetivo agrupar elementos amostrais de acordo com alguma medida de similaridade definida em função de um conjunto de variáveis. Tal agrupamento é realizado de modo que a variabilidade dentro dos grupos seja a menor possível, enquanto que a variabilidade entre os grupos seja maximizada.

Existem duas famílias de métodos para realizar a análise de clusters, os métodos hierárquicos e o não hierárquicos.

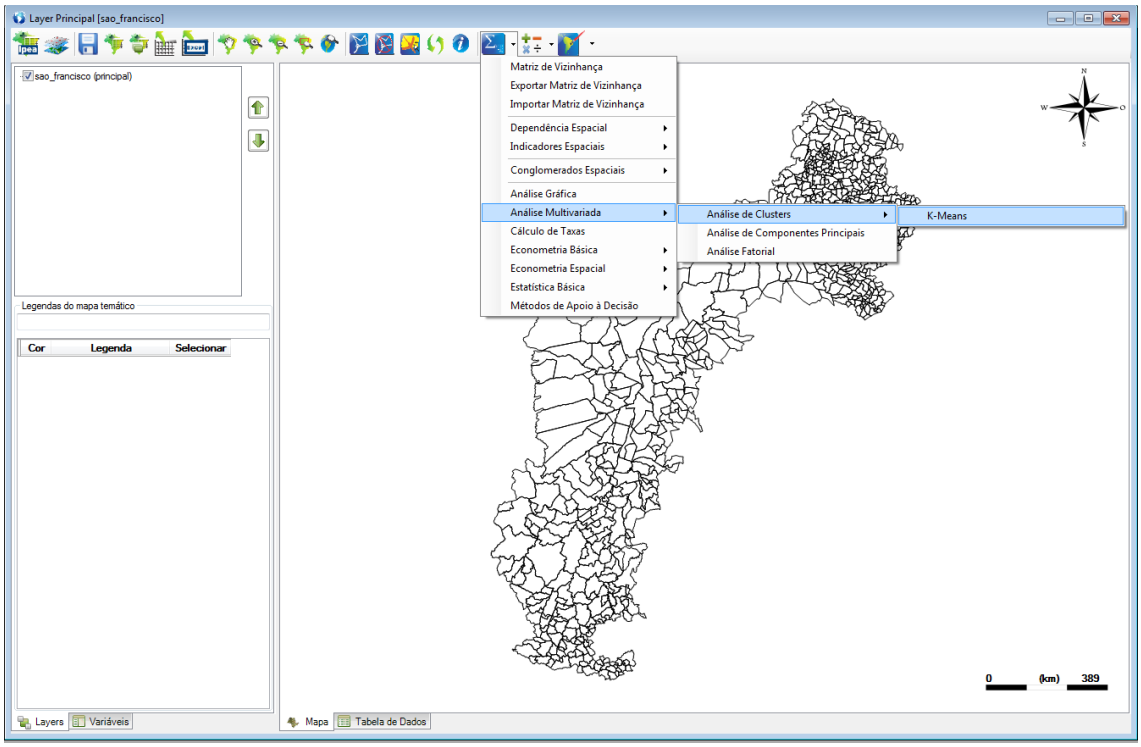
No agrupamento hierárquico é calculada uma matriz de distâncias (similaridades) para cada par de elementos amostrais; o processamento se inicia com o agrupamento do par de elementos com a menor distância (maior similaridade). A cada passo a matriz de distâncias é recalculada considerando os grupos formados nos passos anteriores. O processamento termina quando houver apenas um grupo que concentra toda a amostra.

No agrupamento não hierárquico os elementos amostrais são distribuídos aleatoriamente em um número fixo de grupos. A cada passo o centroide de cada grupo é calculado, assim como a distância dos elementos amostrais com relação a cada centroide dos grupos. Se houver algum elemento alocado em um grupo, mas que tenha uma distância menor para o centroide de outro grupo, tal elemento é deslocado para o grupo de menor distância. Neste ponto os centroides dos grupos envolvidos com o deslocamento do elemento amostral são recalculados. O processamento termina quando todos os elementos amostrais estiverem nos grupos cujas distâncias para os centroides sejam mínimas.

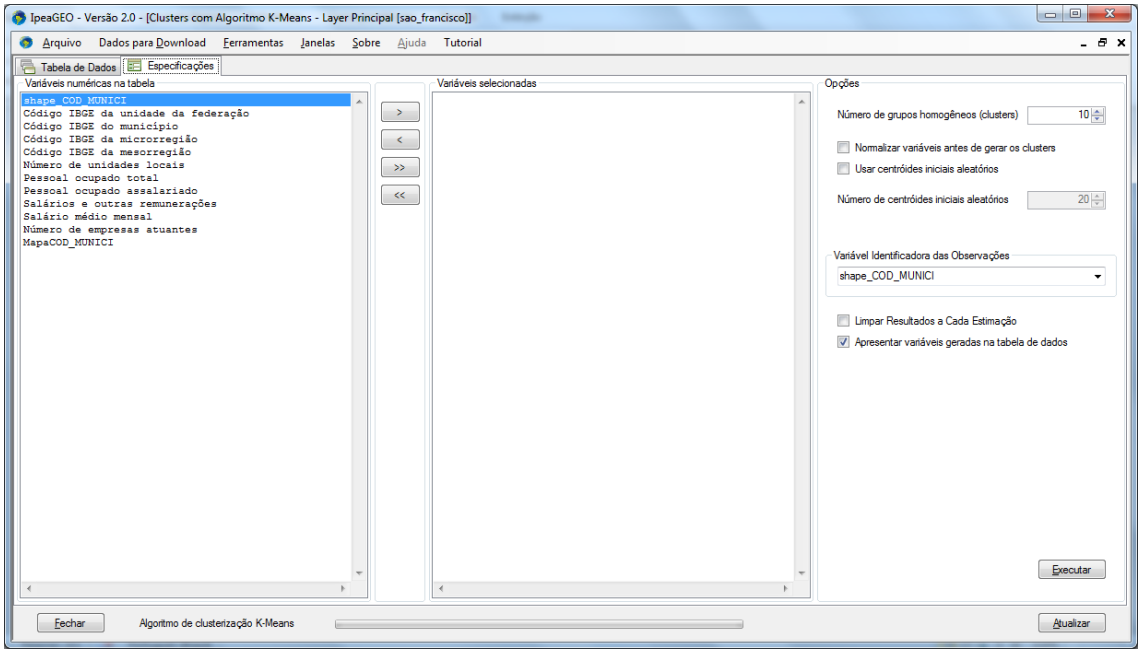
As principais fundamentações dessa metodologia podem ser vistas em **Johnson e Wichern (2007)** e **Mingoti (2005)**.

Exemplo

Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com sua tabela de dados (ver tópico de concatenação) clique na opção “Ferramentas” → “Análise Multivariada” → “Análise de Clusters” → “K-Means”.

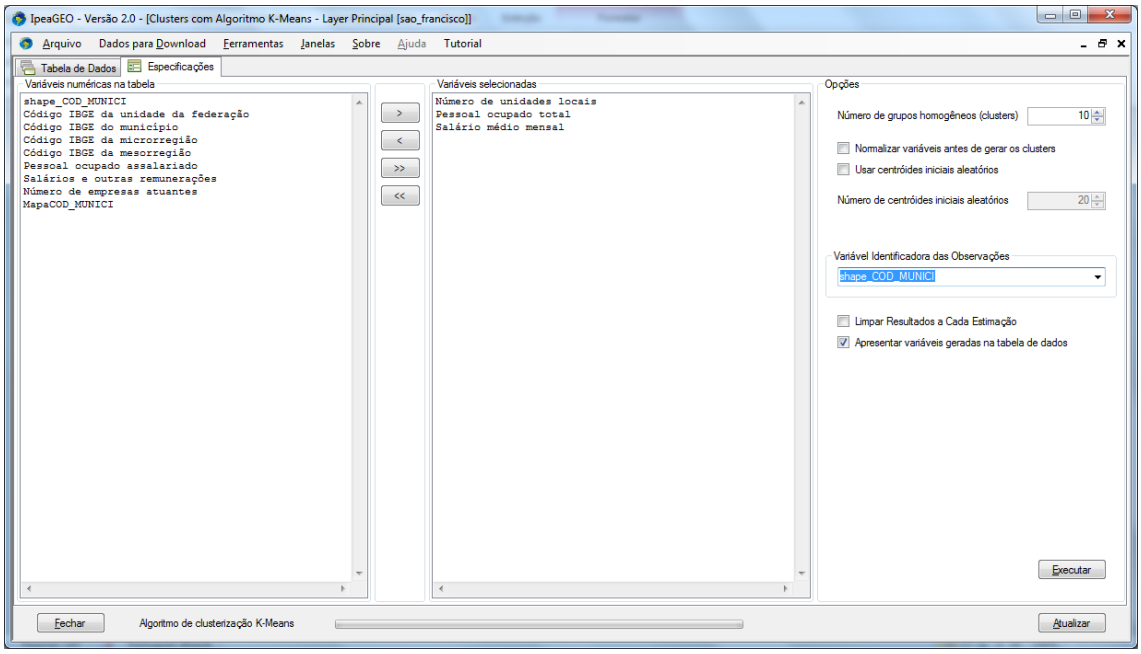


O seguinte formulário irá se abrir:



A opção “Apresentar variáveis geradas na tabela de dados”, presente no lado direito da tela, já está selecionada. Tal opção pode ser desmarcada caso não seja de interesse armazenar na tabela de dados a variável gerada no processo de agrupamento.

Na aba “Especificações”, selecione, dentre as variáveis numéricas disponíveis, aquelas que deseja utilizar para a análise se agrupamentos e clique no botão [>]. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis selecionadas.



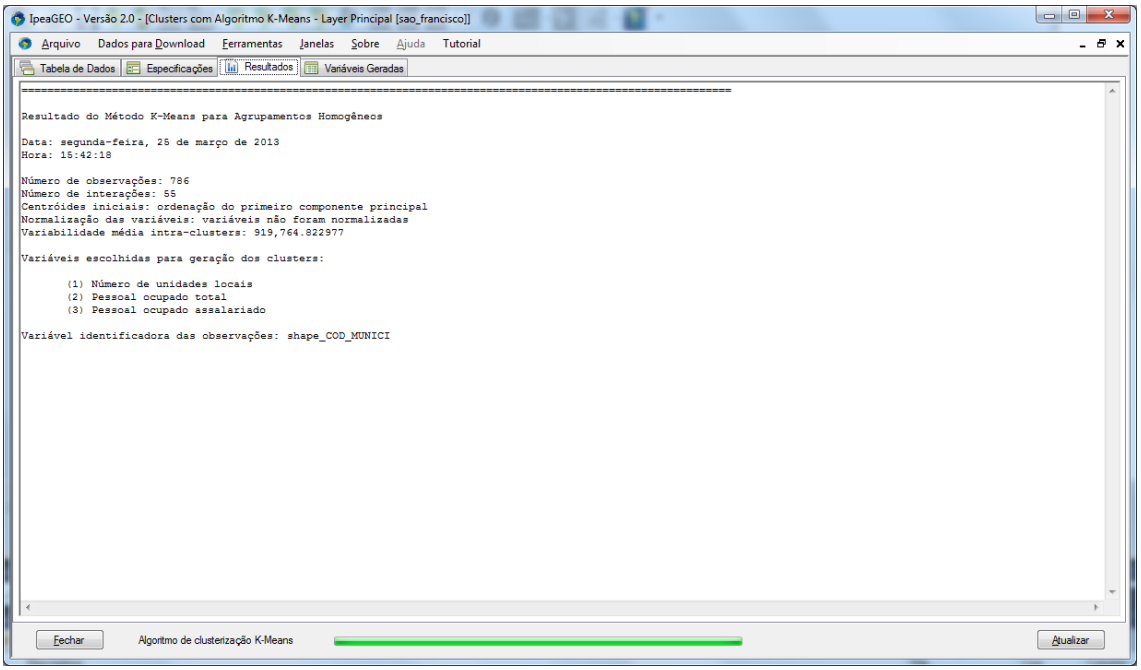
Observe que na aba “Especificações” existem as seguintes opções:

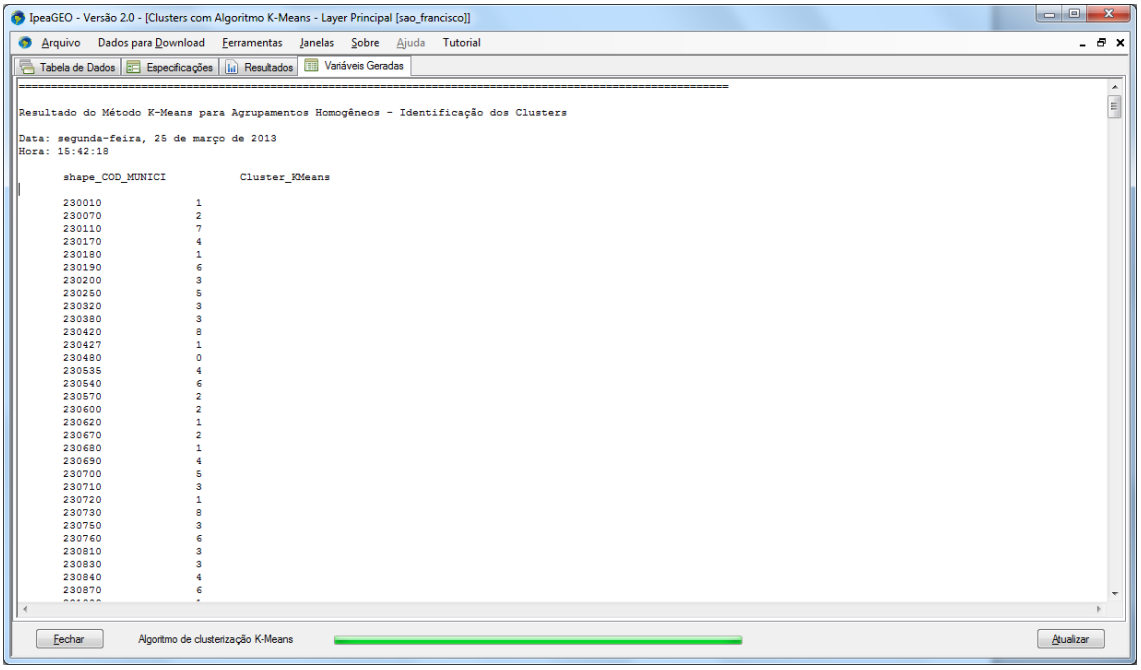
- Número de grupos homogêneos (clusters) – permite especificar o número de clusters a serem gerados

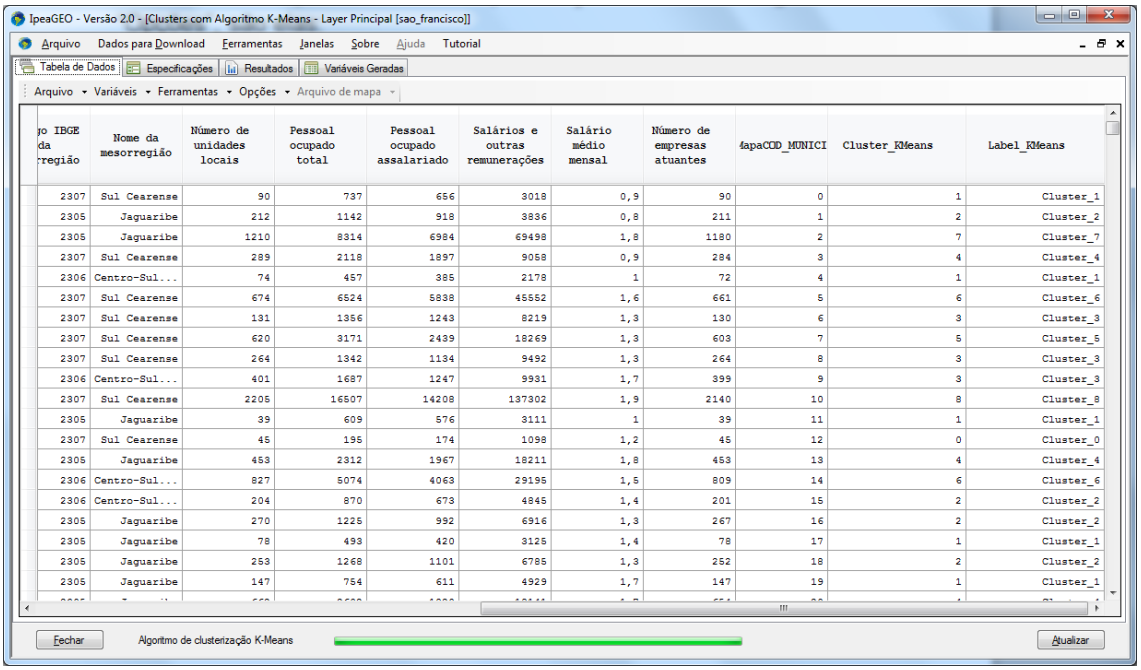
- Normalizar as variáveis antes de gerar os clusters – permite que as variáveis sejam normalizadas (diferença entre os valores e as respectivas médias, dividida pelo desvio padrão) para evitar que variáveis com magnitudes diferentes influenciem no cálculo da matriz de distâncias (similaridades).

- Usar centroides iniciais aleatórios – permite partir de um cenário inicial em que os elementos amostrais são agrupados aleatoriamente. O padrão do IpeaGEO é usar as observações iniciais do banco de dados como centroide.

Depois de selecionadas as opções desejadas, clique no botão “Executar” para realizar a Análise de Clusters. Duas novas abas denominadas “Resultados” e “Variáveis Geradas” apresentam os resultados da análise. Note, também, que são acrescidas duas novas colunas na tabela de dados, contendo a identificação e o *label* do cluster em que cada observação foi alocada.


IpeaGEO - Versão 2.0 - [Clusters com Algoritmo K-Means - Layer Principal [sao_francisco]]
Arquivo
Dados para Download
Ferramentas
Janelas
Sobre
Ajuda
Tutorial
Tabela de Dados
Especificações
Resultados
Variáveis Geradas
Resultado do Método K-Means para Agrupamentos Homogêneos
Data: segunda-feira, 25 de março de 2013
Hora: 15:42:18
Número de observações: 786
Número de interações: 55
Centróides iniciais: ordenação do primeiro componente principal
Normalização das variáveis: variáveis não foram normalizadas
Variabilidade média intra-clusters: 919,764.822977
Variáveis escolhidas para geração dos clusters:
(1) Número de unidades locais
(2) Pessoal ocupado total
(3) Pessoal ocupado assalariado
Variável identificadora das observações: shape_COD_MUNICI
Fechar
Algoritmo de clusterização K-Means
Atualizar


IpeaGEO - Versão 2.0 - [Clusters com Algoritmo K-Means - Layer Principal [sao_francisco]]
Arquivo
Dados para Download
Ferramentas
Janelas
Sobre
Ajuda
Tutorial
Tabela de Dados
Especificações
Resultados
Variáveis Geradas
Resultado do Método K-Means para Agrupamentos Homogêneos - Identificação dos Clusters
Data: segunda-feira, 25 de março de 2013
Hora: 15:42:18
shape_COD_MUNICI Cluster_KMeans
230010 1
230070 2
230110 7
230170 4
230180 1
230190 6
230200 3
230250 5
230320 3
230380 3
230420 8
230427 1
230480 0
230535 4
230540 6
230570 2
230600 2
230620 1
230670 2
230680 1
230690 4
230700 5
230710 3
230720 1
230730 8
230750 3
230760 6
230810 3
230830 3
230840 4
230870 6
Fechar
Algoritmo de clusterização K-Means
Atualizar



...go IBGE da ...região	Nome da mesorregião	Número de unidades locais	Pessoal ocupado total	Pessoal ocupado assalariado	Salários e outras remunerações	Salário médio mensal	Número de empresas atuantes	...mapaCOD_MUNICI	Cluster_KMeans	Label_KMeans
2307	Sul Cearense	90	737	656	3018	0,9	90	0	1	Cluster_1
2305	Jaguaribe	212	1142	918	3836	0,8	211	1	2	Cluster_2
2305	Jaguaribe	1210	8314	6984	69498	1,8	1180	2	7	Cluster_7
2307	Sul Cearense	289	2118	1897	9058	0,9	284	3	4	Cluster_4
2306	Centro-Sul...	74	457	385	2178	1	72	4	1	Cluster_1
2307	Sul Cearense	674	6524	5838	45552	1,6	661	5	6	Cluster_6
2307	Sul Cearense	131	1356	1243	8219	1,3	130	6	3	Cluster_3
2307	Sul Cearense	620	3171	2439	18269	1,3	603	7	5	Cluster_5
2307	Sul Cearense	264	1342	1134	9492	1,3	264	8	3	Cluster_3
2306	Centro-Sul...	401	1687	1247	9931	1,7	399	9	3	Cluster_3
2307	Sul Cearense	2205	16507	14208	137302	1,9	2140	10	8	Cluster_8
2305	Jaguaribe	39	609	576	3111	1	39	11	1	Cluster_1
2307	Sul Cearense	45	195	174	1098	1,2	45	12	0	Cluster_0
2305	Jaguaribe	453	2312	1967	18211	1,8	453	13	4	Cluster_4
2306	Centro-Sul...	827	5074	4063	29195	1,5	809	14	6	Cluster_6
2306	Centro-Sul...	204	870	673	4845	1,4	201	15	2	Cluster_2
2305	Jaguaribe	270	1225	992	6916	1,3	267	16	2	Cluster_2
2305	Jaguaribe	78	498	420	3125	1,4	78	17	1	Cluster_1
2305	Jaguaribe	253	1268	1101	6785	1,3	252	18	2	Cluster_2
2305	Jaguaribe	147	754	611	4929	1,7	147	19	1	Cluster_1

Analise de Componentes Principais

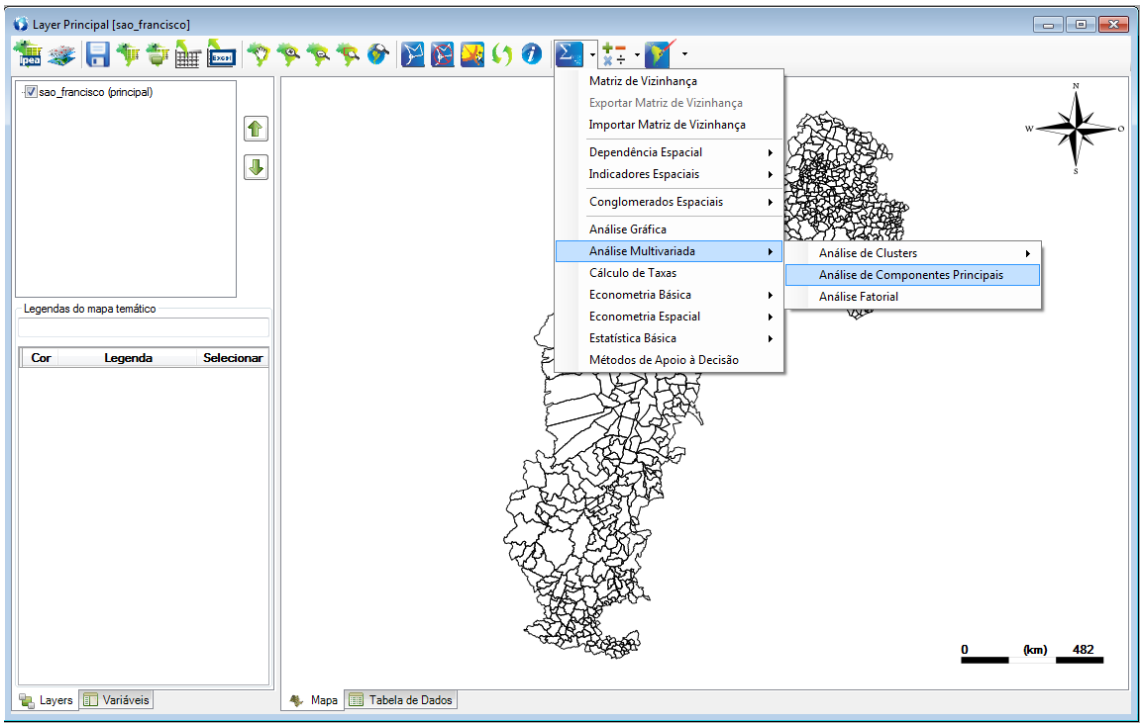
Visão Geral

A análise de componentes principais é uma técnica multivariada não paramétrica que permite representar os dados como uma coleção de combinações lineares das variáveis originais (chamados de componentes principais) cuja dimensão é a mesma dos dados originais. O interessante da técnica é que tais componentes são ortogonais, ou seja, são não correlacionados. Isto permite selecionar um subconjunto destes componentes com o objetivo de reduzir a dimensão do conjunto original de dados. É possível avaliar o peso de cada componente na variância total dos dados, o que permite selecionar aquelas que explicam a maior parcela de variância. Deste modo, o conjunto de dados com dimensão reduzida é selecionado de modo que não ocorram perdas significativas quanto à explicação da variabilidade original dos dados. A técnica de componentes principais é um método de estimação usado para se fazer análise fatorial. Além de ser uma forma não paramétrica para a redução da dimensão de um conjunto de dados ainda permite identificar relações entre variáveis que não são visíveis apenas observando as correlações entre elas.

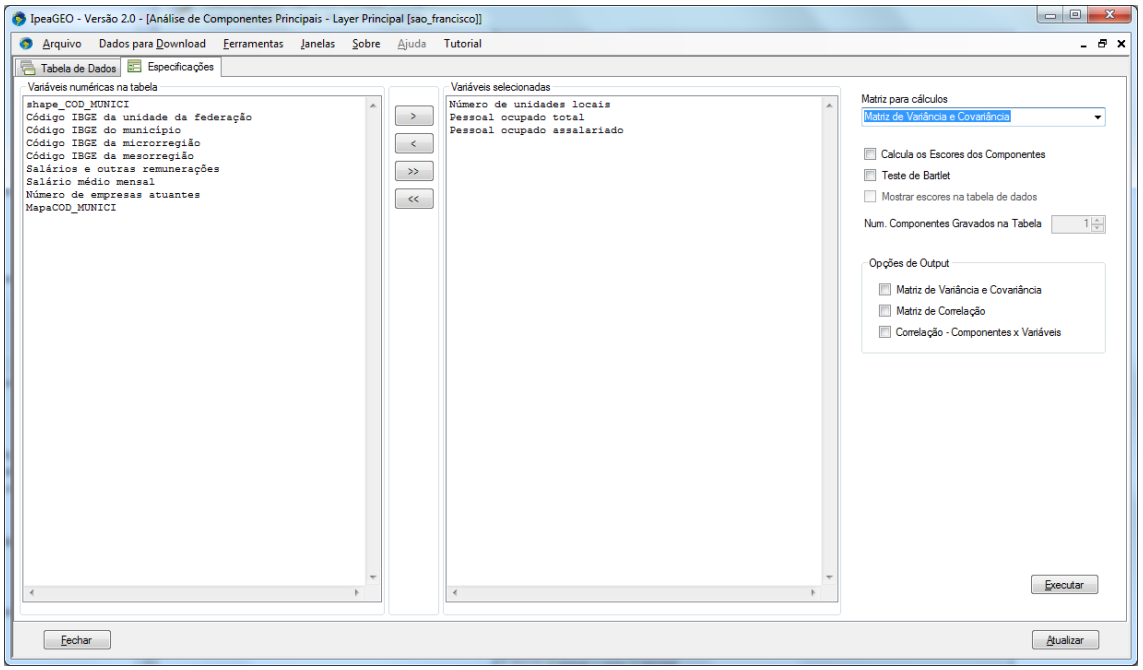
As principais fundamentações dessa metodologia podem ser vistas em **Johnson e Wichern (2007)** e **Mingoti (2005).**

Exemplo

Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital com sua tabela de dados (ver tópico de concatenação) clique na opção “Ferramentas” → “Análise Multivariada” → “Análise de Componentes Principais”:



O seguinte formulário irá se abrir:



Na aba “Especificações”, selecione as variáveis numéricas desejadas para a análise com o botão [>]. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis selecionadas.

É necessário definir como a dependência entre as variáveis será avaliada: via “Matriz de Variância e Covariância” ou “Matriz de Correlações”. A matriz de variância e covariância somente deve ser usada se as escalas das variáveis são semelhantes, ou seja, se o padrão de variabilidade é semelhante para as diversas variáveis. Se houver discrepância com relação à variância das variáveis envolvidas é aconselhável usar a

matriz de correlação.

As seguintes opções também estão disponíveis aos usuários:

- Calcular os Escores dos Componentes – calcula as combinações lineares de cada variável do conjunto de dados original, para cada componente selecionado.

- Realizar o Teste de Bartlet – é um teste que verifica se as variáveis originais são correlacionadas. Se as variáveis originais forem não correlacionadas, as componentes principais serão as próprias variáveis.

- Mostrar os Escores na tabela de dados – permite armazenar os escores calculados para cada componente principal na base de dados.

Pode-se definir a quantidade de componentes a serem gravados na tabela de dados, lembrando que o máximo de componentes gerados é igual ao número de variáveis utilizadas para o cálculo.

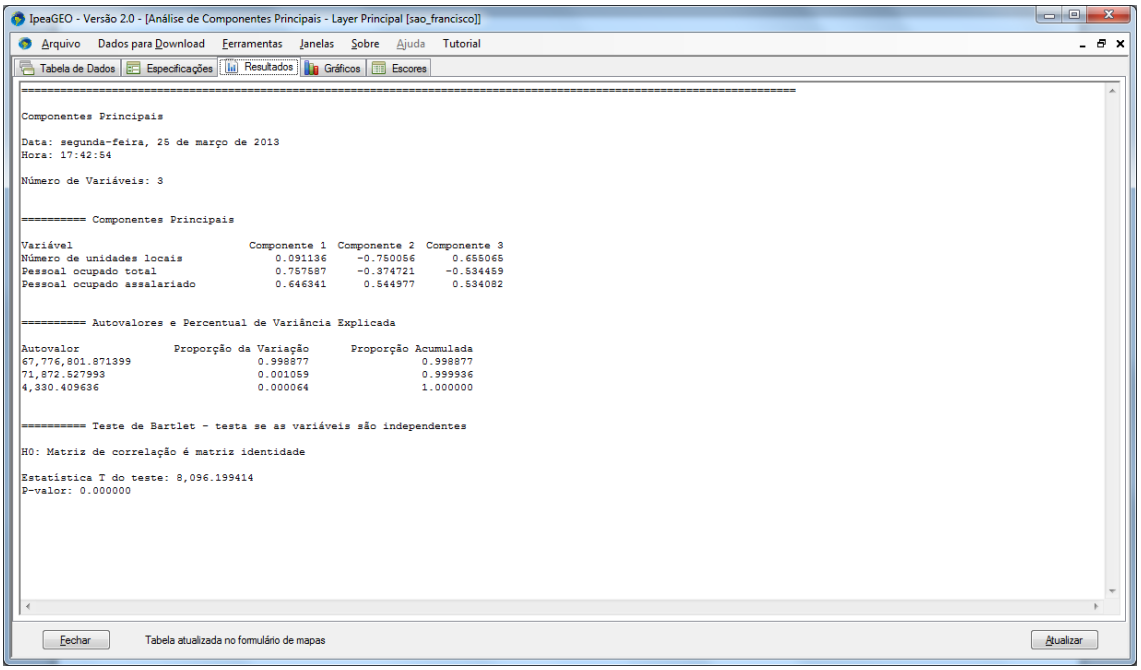
As seguintes opções também estão disponíveis para exibição de resultados:

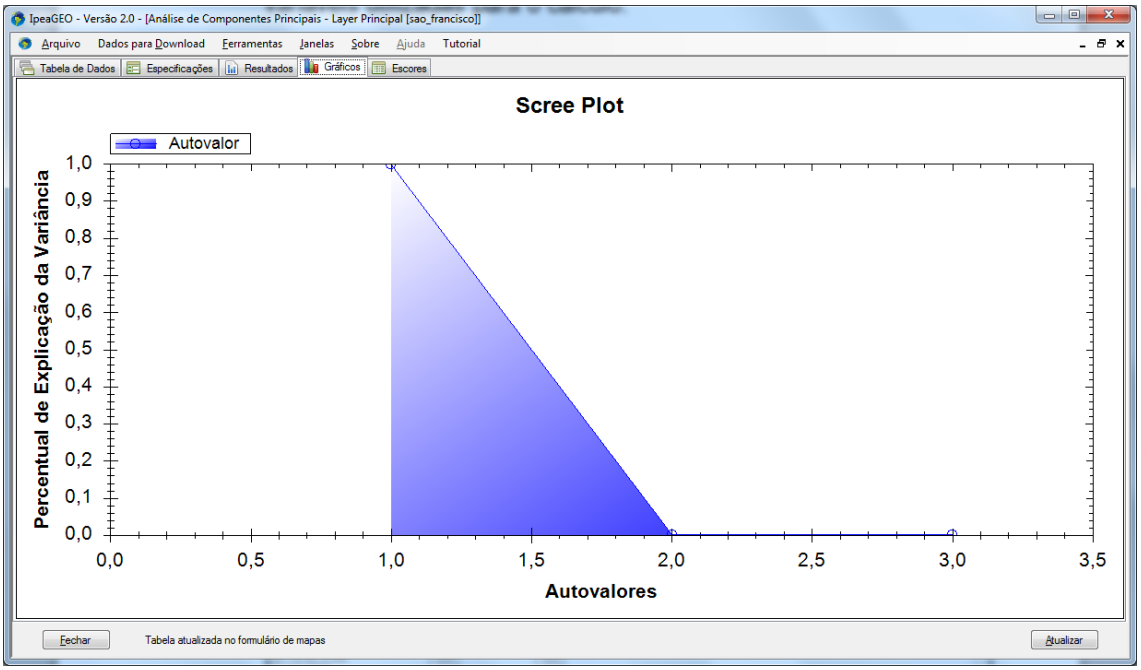
- Matriz de Variância e Covariância;

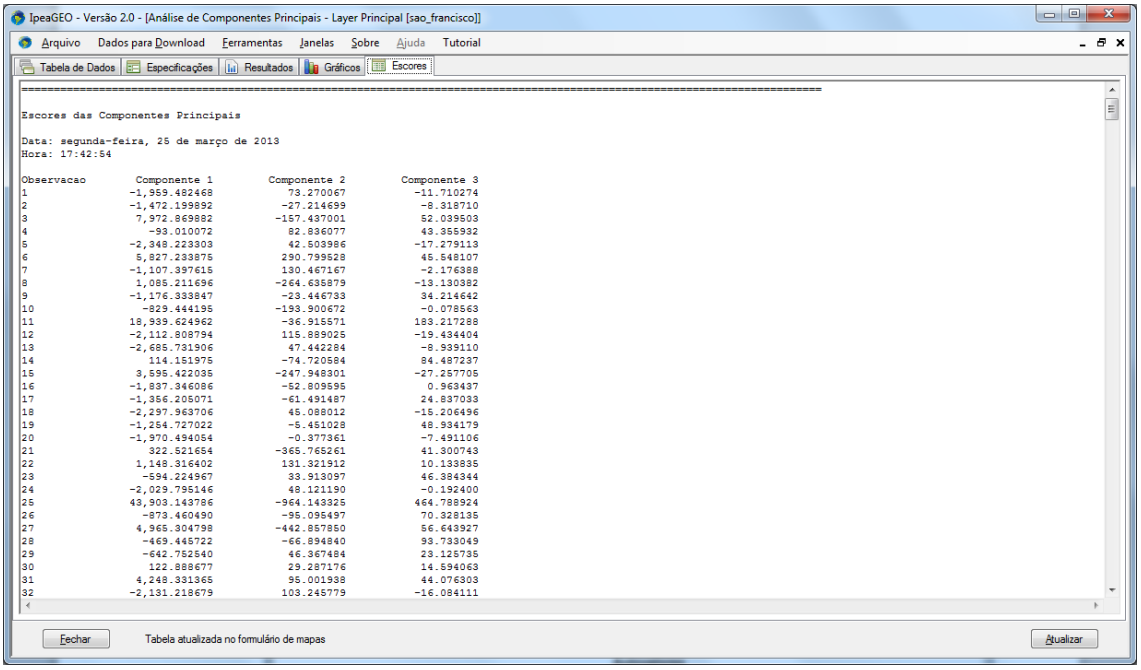
- Matriz de Correlações;

- Correlação entre as Componentes e as Variáveis.

Depois de selecionadas as opções desejadas, clique no botão “Executar” para realizar a Análise de Componentes Principais. Três novas abas denominadas “Resultados”, “Gráficos” e “Escores” irão aparecer contendo os resultados da análise.







Caso seja interesse, é possível atualizar a tabela de dados com as componentes geradas clicando em "Atualizar". Estes resultados podem ser usados para análises posteriores como, por exemplo, gerar um mapa temático.

Analise Fatorial

Visão Geral

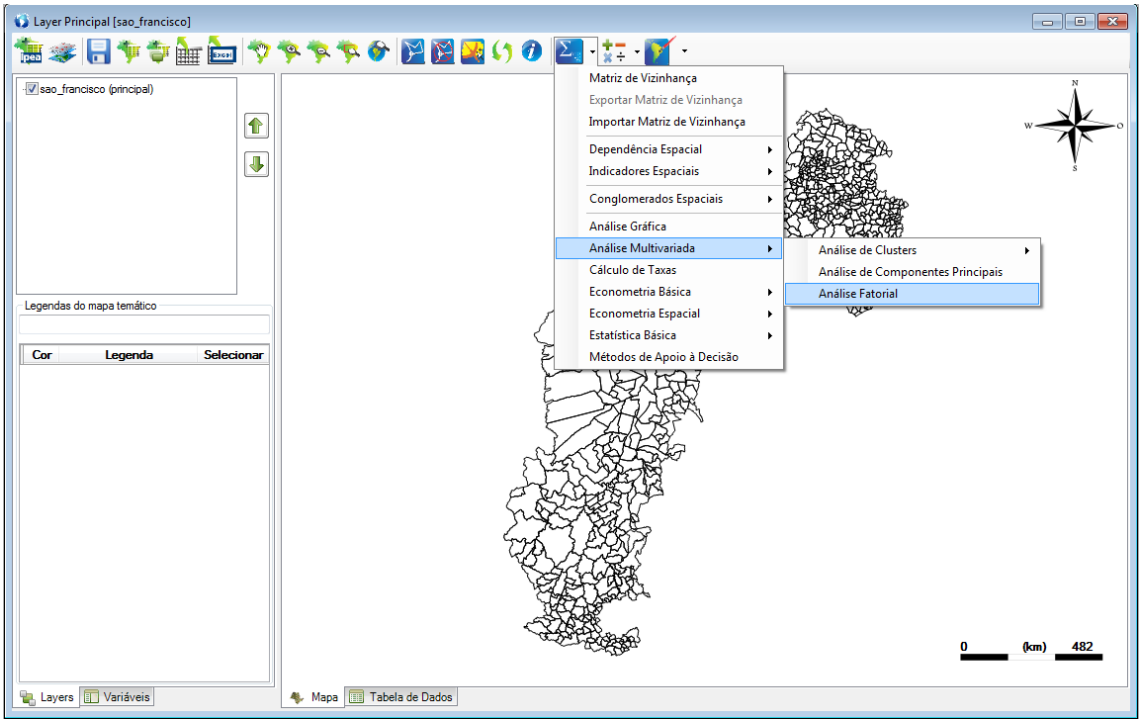
A análise fatorial é uma técnica multivariada paramétrica que permite expressar a matriz de variância-covariância ou a matriz de correlação dos dados originais como a

soma de dois componentes: uma matriz que depende apenas da combinação linear das variáveis originais e uma matriz que representa a variabilidade residual. Desta forma, é possível criar variáveis latentes, chamadas de fatores, que são combinações lineares das variáveis originais – semelhante aos modelos de regressão linear – de modo que a matriz de variabilidade associada a uma quantidade menor de fatores que das variáveis originais represente uma parcela significativa da variabilidade total. Esta técnica permite reduzir a dimensão do conjunto de dados, bem como identificar variáveis latentes que podem ter significados práticos. Observe que a semelhança com o método dos componentes principais não é por acaso, tendo em vista que os componentes principais é um dos métodos usados para se extrair os fatores.

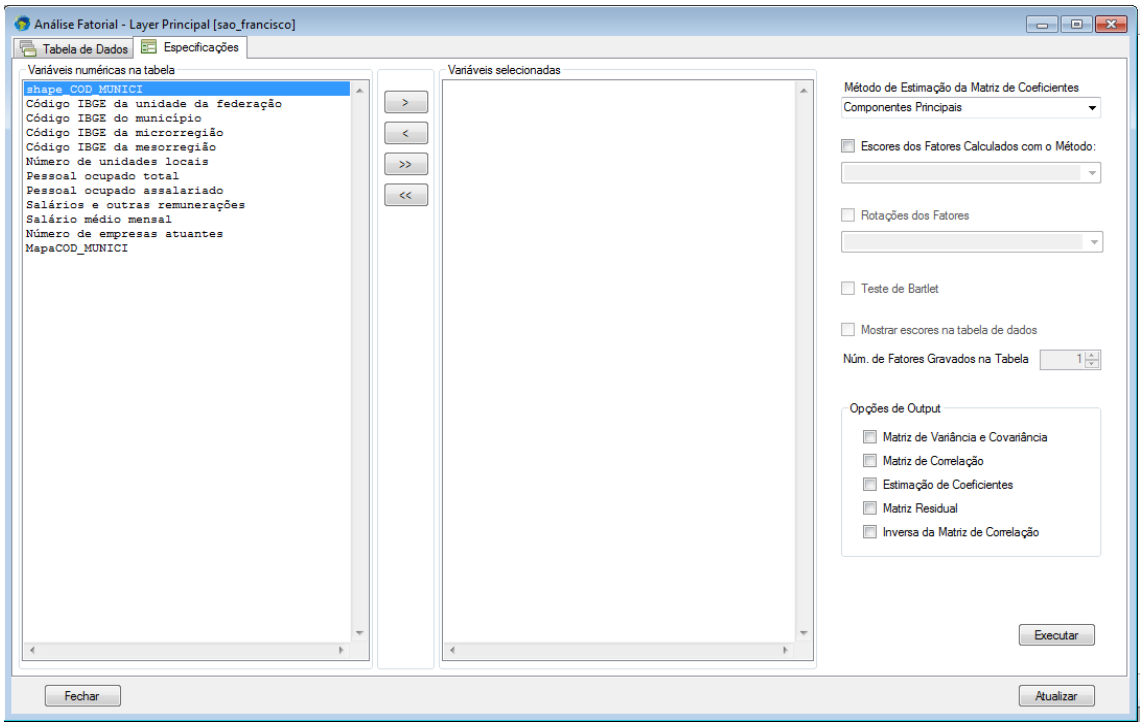
O IpeaGEO permite a estimação da matriz de coeficientes (chamada de matriz de cargas fatoriais) pelos métodos dos componentes principais, fatores principais e pela máxima verossimilhança. Os escores podem ser calculados por mínimos quadrados ponderados ou por regressão. Além disso, é possível aplicar as seguintes rotações aos dados: quartimax, varimax, entropia mínima, quatimin, bi-quartimin e covarimin. Na apresentação dos resultados é possível exibir o screeplot, os autovalores e as comunalidades, as matrizes de variância-covariância e de correlação, os valores estimados para os coeficientes (cargas fatoriais), a matriz residual e a inversa da matriz de correlação. Existe também a possibilidade de se realizar o teste de Bartlett para verificar se as variáveis originais são correlacionadas; caso sejam não correlacionadas tais variáveis formam uma base ortogonal e serão idênticas aos fatores.

Exemplo

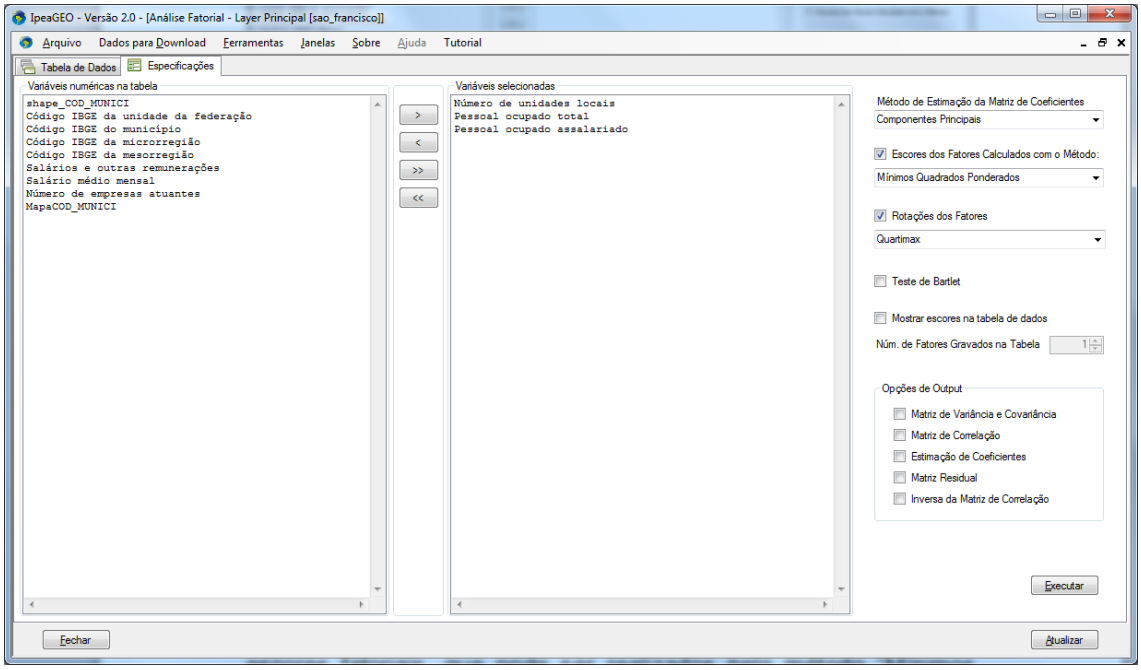
Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital (shape) com sua tabela de dados (ver tópico de concatenação) clicando na opção “Ferramentas” → “Análise Multivariada” → “Análise Fatorial”.



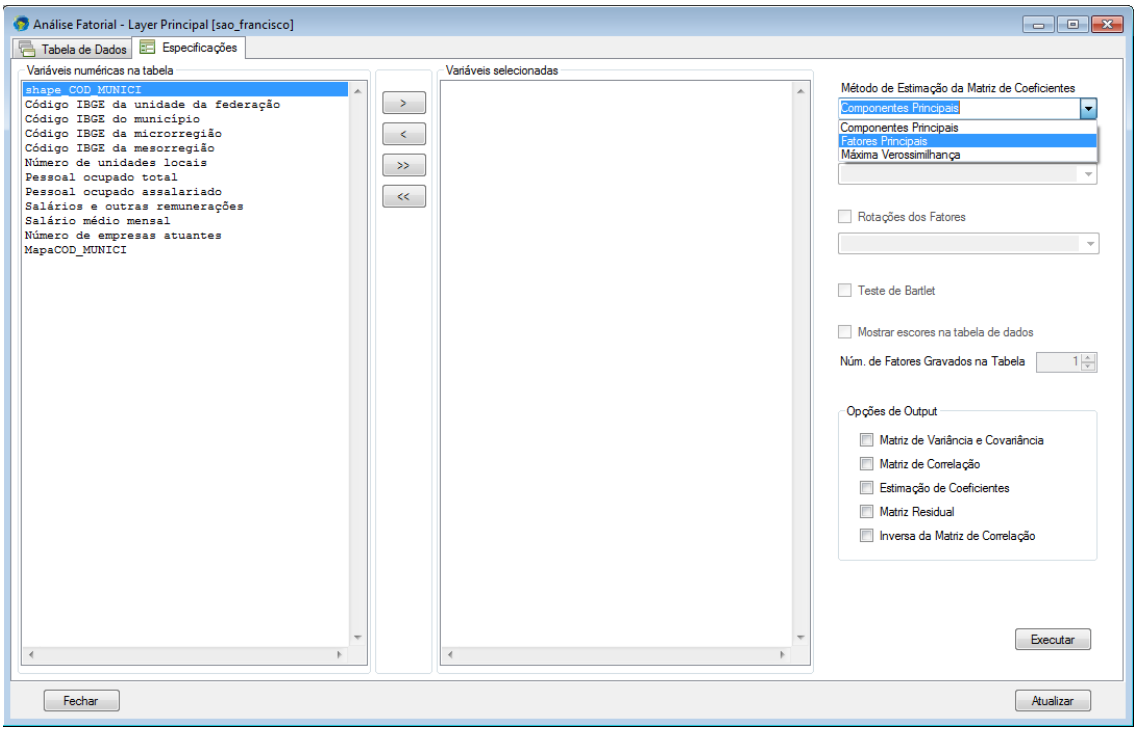
O seguinte formulário irá se abrir:



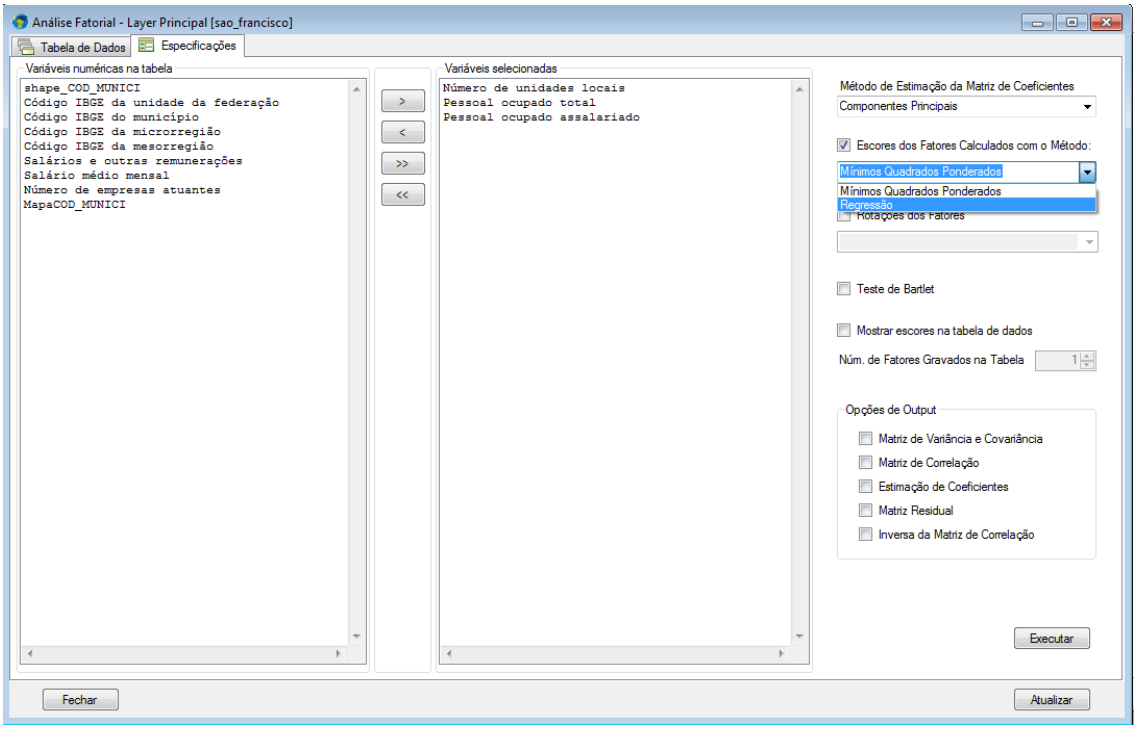
Na aba “Especificações”, selecione, dentre as variáveis numéricas disponíveis, aquelas que deseja utilizar para a análise e clique no botão [>]. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis selecionadas.



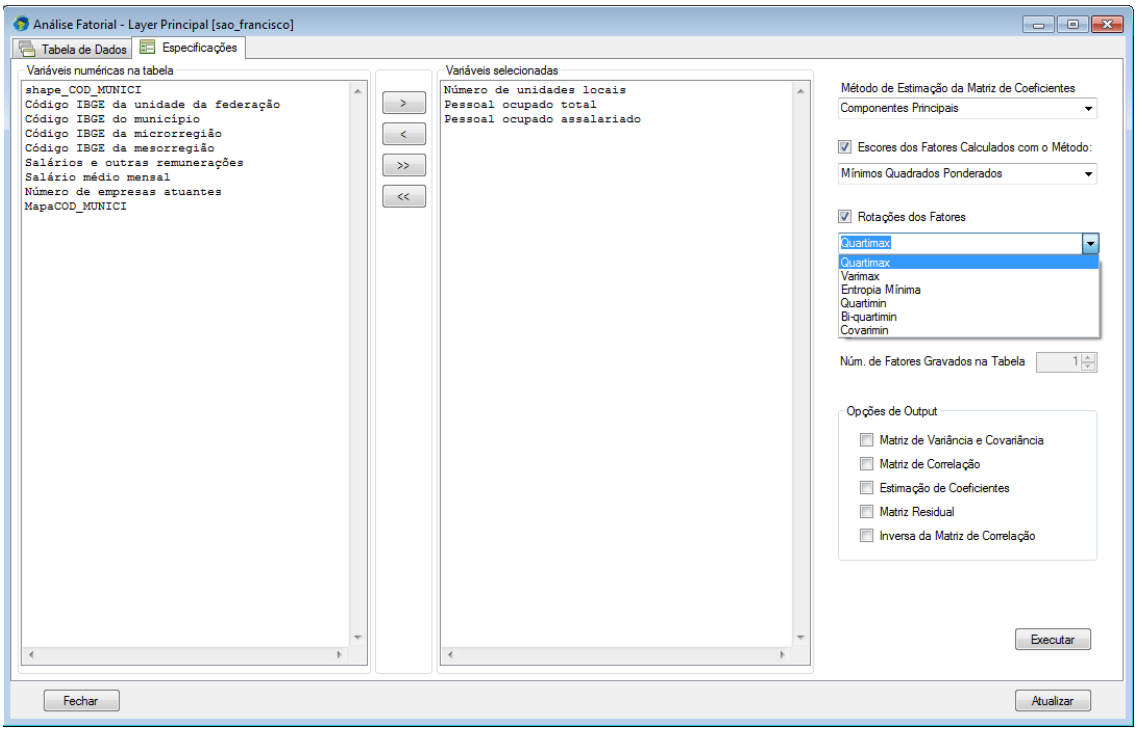
É necessário definir o método de estimação da Matriz de Coeficientes, que pode ser realizado via “Componentes Principais”, “Fatores Principais” ou “Máxima Verossimilhança”.



É possível definir o método de cálculo dos escores fatoriais, que pode “Mínimos Quadrados Ponderados” ou “Regressão”.



Caso seja de interesse, os fatores podem ser rotacionados pelos métodos: “Quatimax”, “Varimax”, “Entropia Mínima”, “Quartimin”, “Bi-quartimin” ou “Covarimin”.

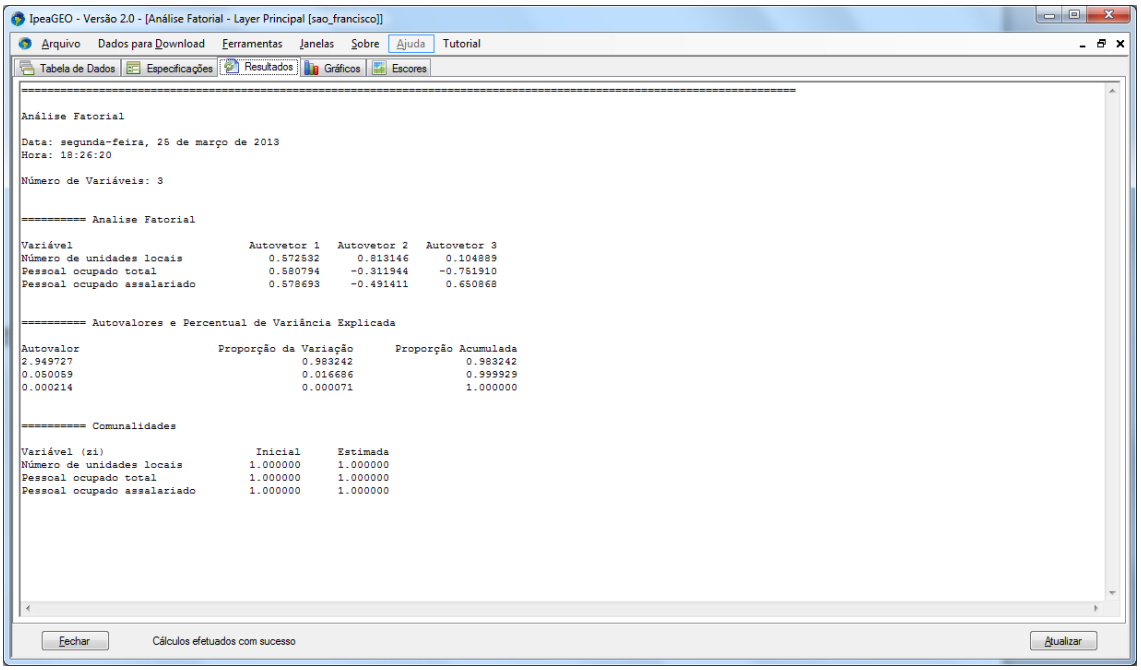


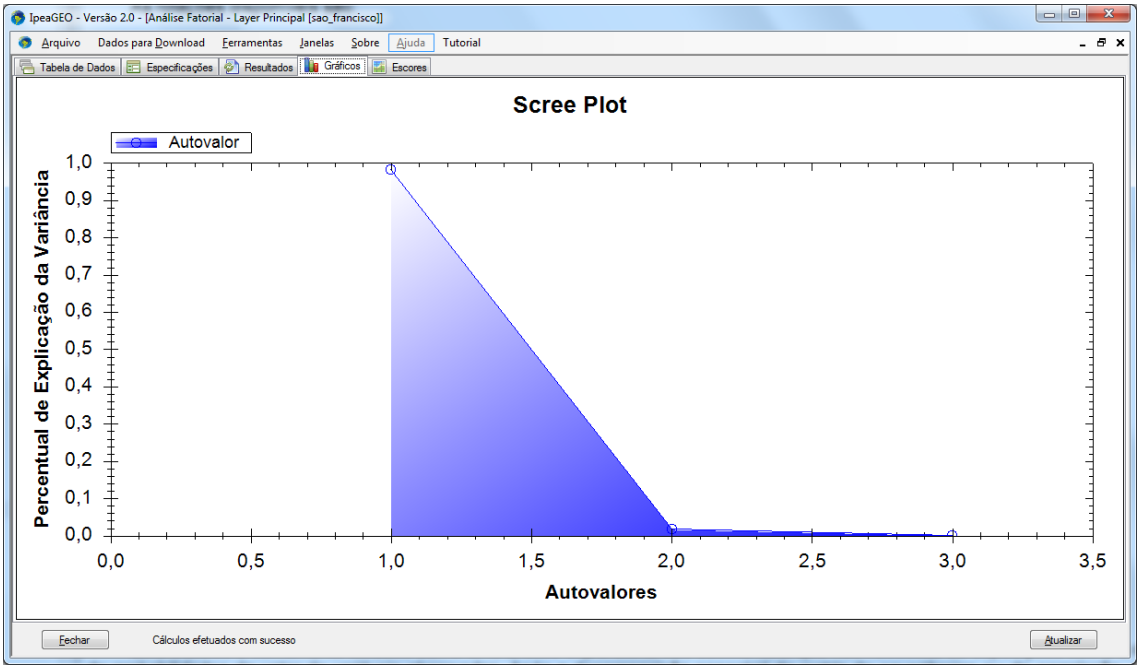
As seguintes opções para a saída (output) estão disponíveis aos usuários:

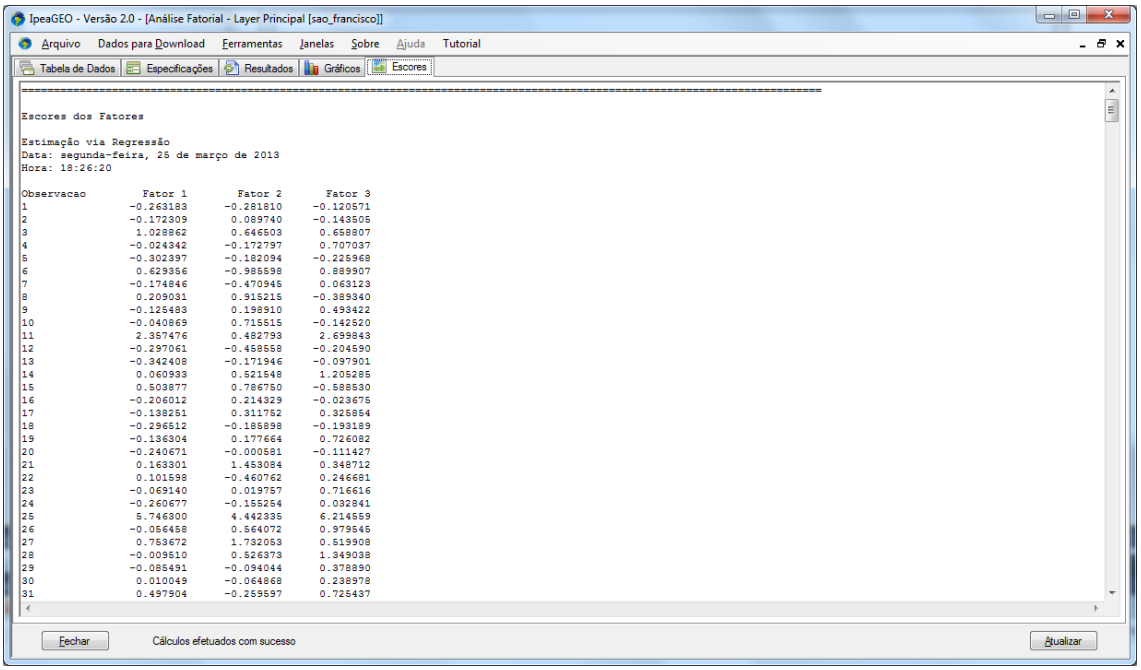
- Scree Plot;
- Autovalores e Comunalidades;
- Matriz de Variância e Covariância;
- Matriz de Correlação;
- Valor dos Coeficientes Estimados;
- Matriz Residual;
- Inversa da Matriz de Correlação.

Pode-se definir a quantidade de fatores a serem gravados na tabela de dados, lembrando que a quantidade máxima é igual ao número de variáveis utilizadas para o cálculo.

Depois de selecionadas as opções desejadas, clique no botão “Executar” para realizar a Análise Fatorial. Três novas abas denominadas “Resultados”, “Gráficos” e “Escores” irão aparecer contendo os resultados da análise.


IpeaGEO - Versão 2.0 - [Análise Fatorial - Layer Principal [sao_francisco]]
Arquivo
Dados para Download
Ferramentas
Janelas
Sobre
Ajuda
Tutorial
Tabela de Dados
Especificações
Resultados
Gráficos
Escores
Análise Fatorial
Data: segunda-feira, 25 de março de 2013
Hora: 18:26:20
Número de Variáveis: 3
========== Analise Fatorial
Variável Autovetor 1 Autovetor 2 Autovetor 3
Número de unidades locais 0.572532 0.813146 0.104889
Pessoal ocupado total 0.580794 -0.311944 -0.751910
Pessoal ocupado assalariado 0.578693 -0.491411 0.650868
========== Autovalores e Percentual de Variância Explicada
Autovalor Proporção da Variação Proporção Acumulada
2.949727 0.983242 0.983242
0.050059 0.016686 0.999929
0.000214 0.000071 1.000000
========== Comunalidades
Variável (zi) Inicial Estimada
Número de unidades locais 1.000000 1.000000
Pessoal ocupado total 1.000000 1.000000
Pessoal ocupado assalariado 1.000000 1.000000
Fechar
Cálculos efetuados com sucesso
Atualizar


IpeaGEO - Versão 2.0 - [Análise Fatorial - Layer Principal [sao_francisco]]
Arquivo
Dados para Download
Ferramentas
Janelas
Sobre
Ajuda
Tutorial
Tabela de Dados
Especificações
Resultados
Gráficos
Escores
Scree Plot
Autovalor
Percentual de Explicação da Variância
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
0,0
0,5
1,0
1,5
2,0
2,5
3,0
3,5
Autovalores
Fechar
Cálculos efetuados com sucesso
Atualizar



Para que os resultados da análise sejam salvos na tabela de dados principal, contida na aba “Tabela de Dados”, é necessário que a opção "Mostrar novas variáveis na tabela de dados" esteja marcada e em seguida clicar no botão “Atualizar”, localizado no canto inferior direito do formulário. Note que as variáveis “Observacao”, “Y_observado” e “Prob_predita_” foram incluídas na tabela.

Cálculo de Taxas

Visão Geral

O mapeamento da taxa de ocorrência de eventos é uma importante ferramenta para o direcionamento de políticas públicas. O cálculo de taxas brutas pode levar a distorções como, por exemplo, para de municípios muito pequenos. Nesses casos a estimação de taxas brutas é pouco viável, pois o pequeno número de observações pode levar a distorções nas estimativas das taxas, inclusive com a inflação na variabilidade estimada.

Clayton e Kaldor (1987) solucionaram tal problema por meio da suavização Bayesiana. Além desse trabalho, Marshall (1991) propôs uma estimativa de taxas de alta estabilidade, que inclui a localização geográfica do município como informação adicional. Tais metodologias são chamadas de empíricas uma vez que, em ambos os casos, os dados auxiliares são obtidos da própria amostra. Pringle (1996) faz uma revisão de técnicas Bayesianas empíricas tendo como base estudos de caso de câncer e mortalidade neonatal.

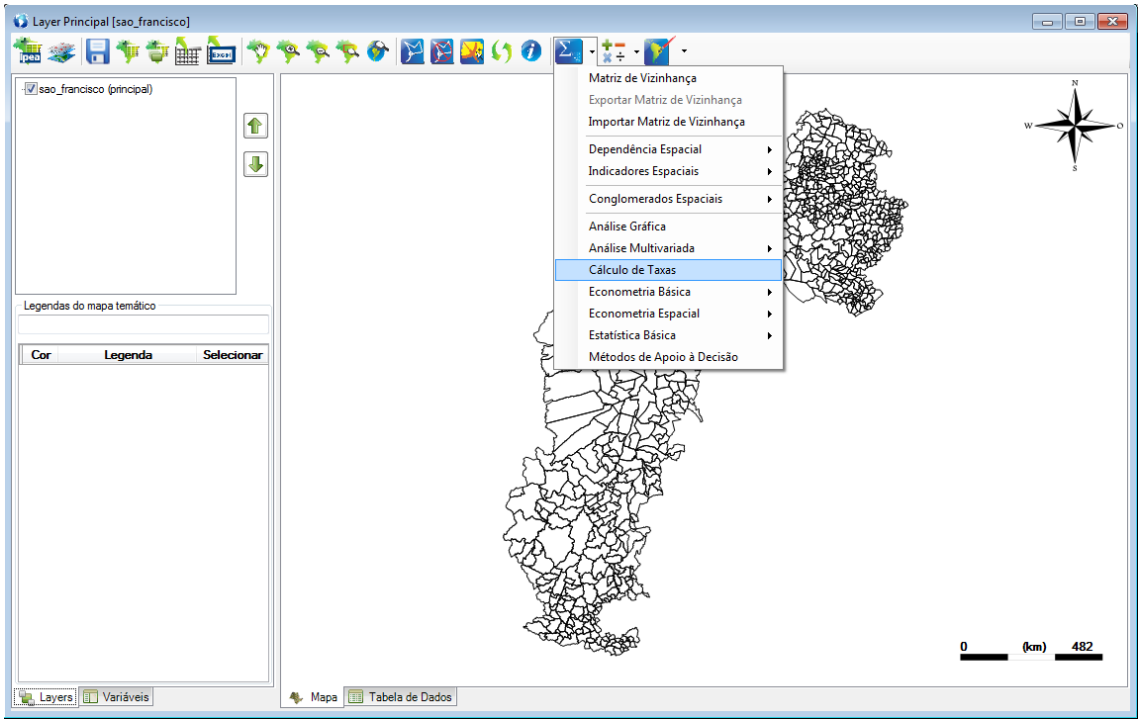
Nesse módulo, são apresentadas rotinas para cálculo de taxas nas quais a Estatística Bayesiana e a Espacial são as ferramentas metodológicas básicas. Com isso, têm-se estimativas com menor variabilidade e, por isso, mais indicadas.

O formulário de Cálculo de Taxas do IpeaGEO contém as seguintes opções de metodologia (veja detalhes abaixo):

- Taxa bruta;
- Taxa espacial;
- Taxa Bayesiana empírica;
- Taxa Bayesiana empírica espacial;
- Risco relativo;
- Risco relativo Bayesiano (Clayton e Kaldor);
- Taxa Bayesiana (Clayton e Kaldor);
    - Priori Gamma;
    - Priori Log-Normal;
    - Priori Gamma Espacial;
    - Priori Log-Normal Espacial;

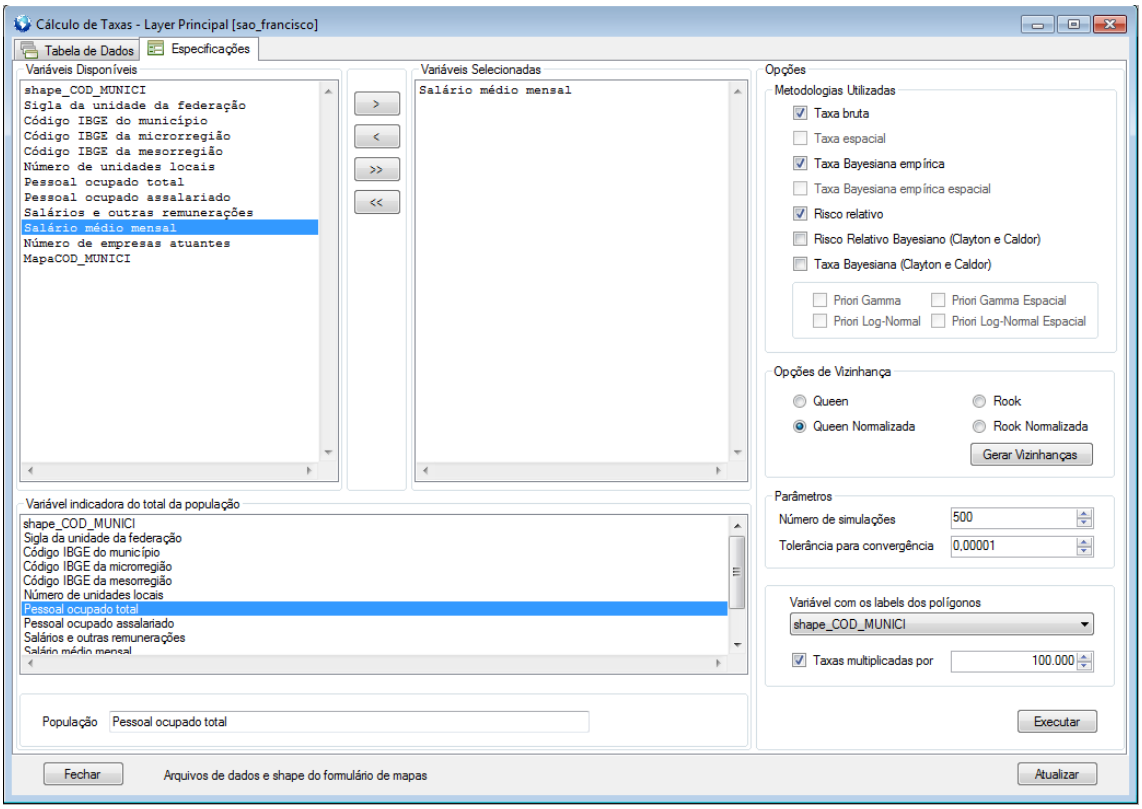
Exemplo

O formulário de cálculo de Taxas está disponível no seguinte local: Após concatenar um arquivo de malha digital (shape) com os dados, clicar em “Ferramentas” → Cálculo de Taxas.

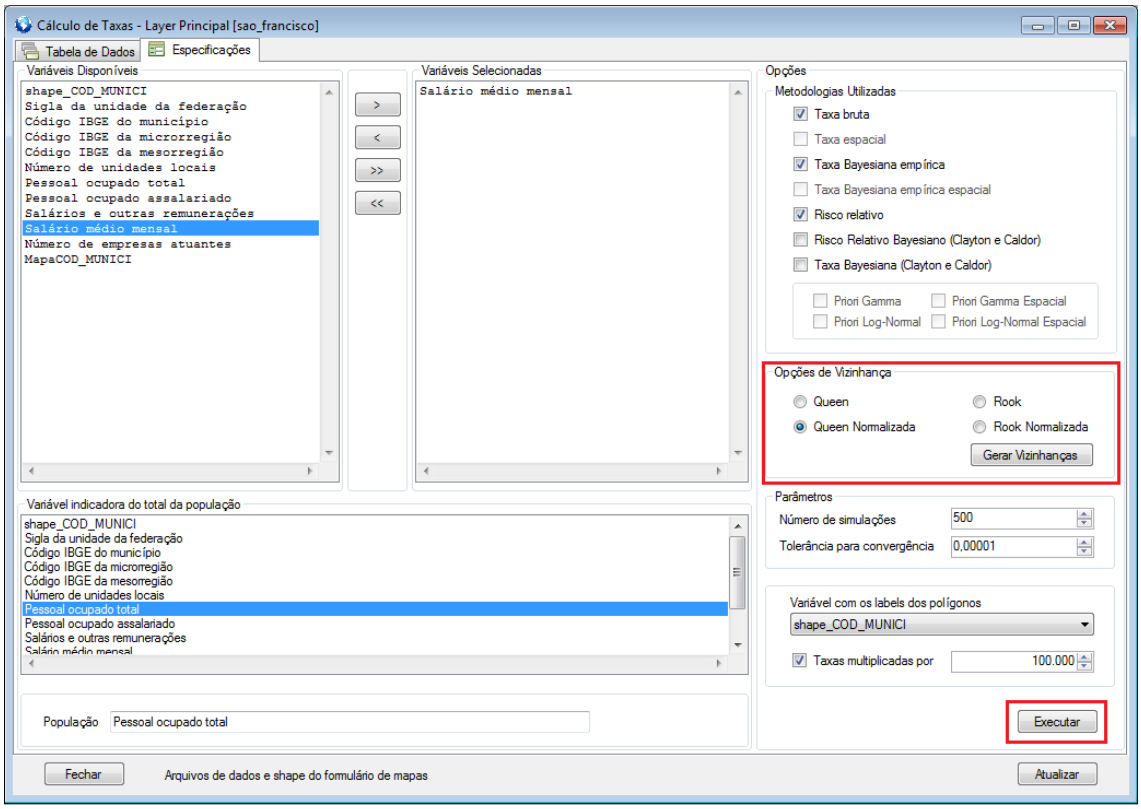


Seleciona-se uma variável indicadora do total da população (variável selecionada

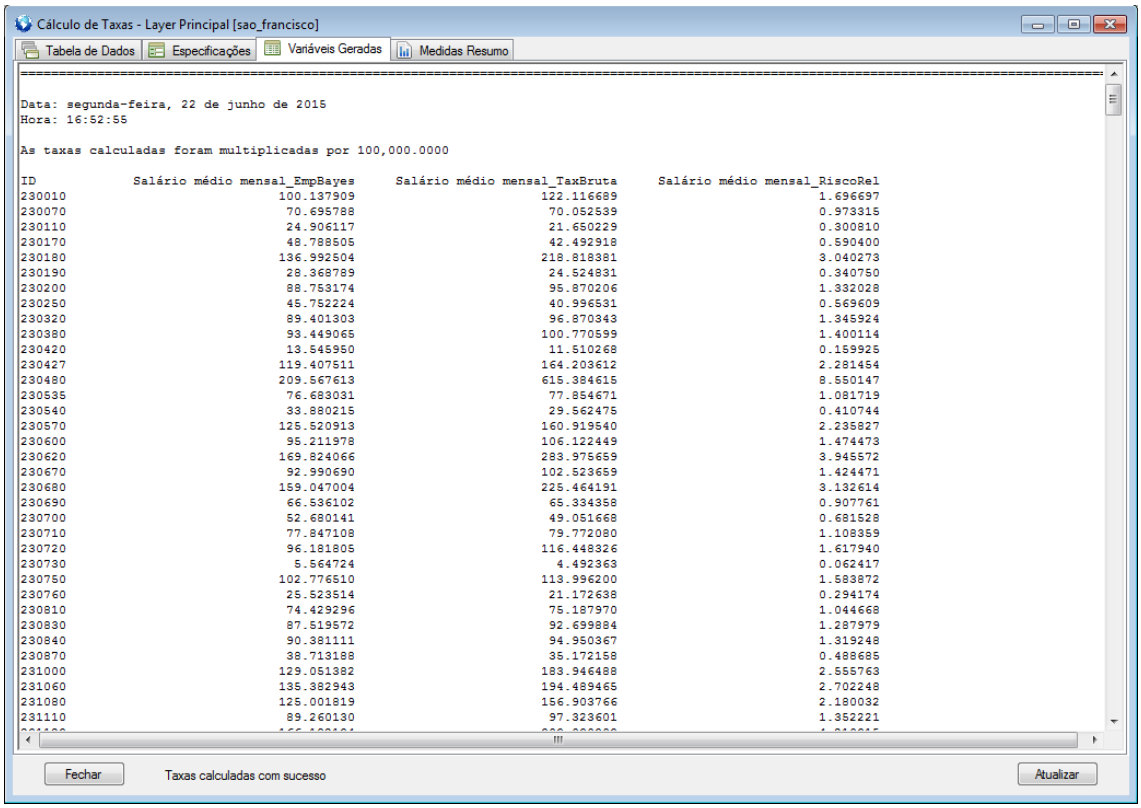
em azul) e uma variável de interesse de estudo (selecionada no quadro da direita), os tipos de taxas a serem calculadas e os parâmetros a serem utilizados para a simulação (número de réplicas e tolerância nas estimativas).

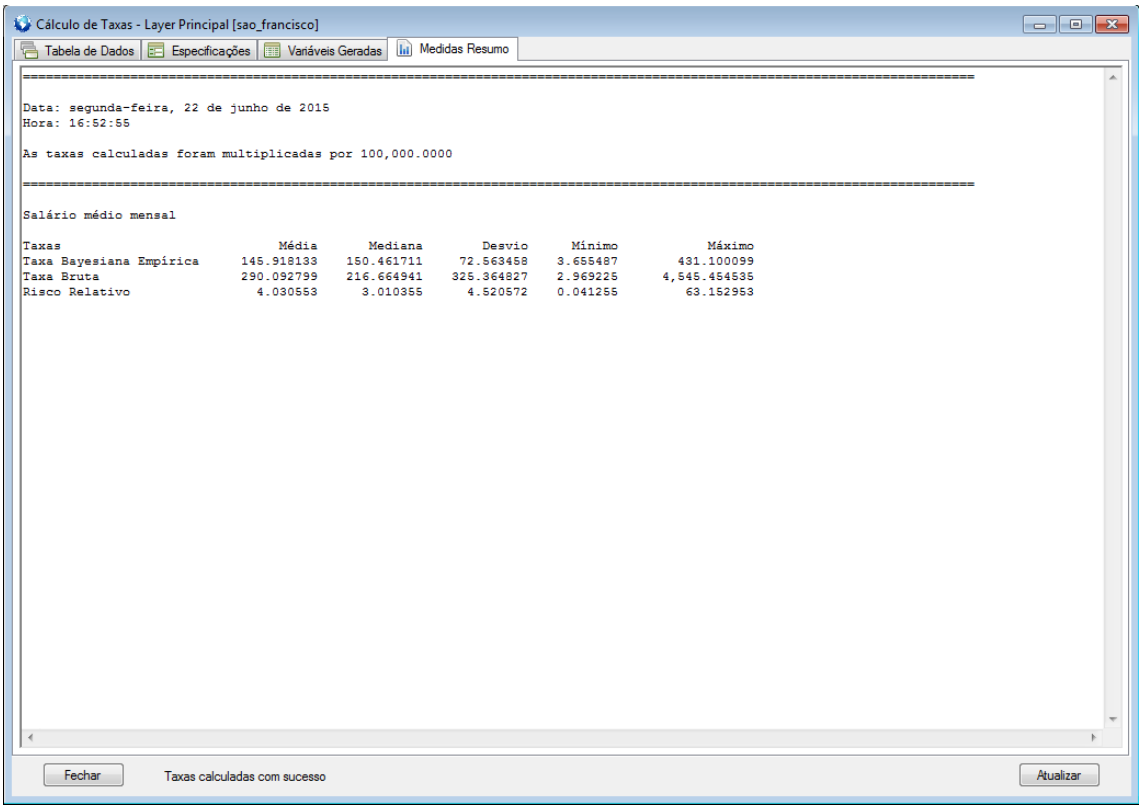


Posteriormente, seleciona-se o tipo de vizinhança a ser considerada no cálculo e gera-se a matriz de vizinhanças clicando em “Gerar Vizinhança”. Para concluir, clique em “Executar” para calcular as taxas correspondentes.



Duas novas abas denominadas “ Variáveis Geradas” e “Medidas Resumo” irão aparecer contendo os resultados da análise.





Econometria Básica

Propensity Score Matching

Visão Geral

O *propensity score matching* (PSM) é uma técnica estatística utilizada para a construção de grupos com o objetivo de fazer comparações estatísticas entre grupos. Esta técnica é baseada em um modelo de probabilidade da participação de um individuo em um determinado grupo ou tratamento, usando características observadas.

Os participantes são então associados (*matched*) com base na probabilidade estimada de participação em cada grupo, ou com base no escore de propensão (*propensity score*).

O efeito médio do tratamento (average treatment effect - ATE) de um programa é então calculado como a diferença média entre os resultados (*outcomes*) para os dois grupos estudados: participantes e não participantes do programa.

A validade do PSM depende de duas condições:

- Independência condicional: fatores não observados não afetam a participação do individuo no programa em questão.

- Existência de uma interseção (com base no escore de propensão) entre os participantes e os não participantes. Essa condição é necessária para que o efeito do programa possa ser comparado.

Diferentes abordagens são usadas para associar (*match*) os participantes e os não participantes com base no escore de propensão. Entre os métodos podemos citar o vizinho mais próximo (nearest-neighbor – NN), caliper, associação por raio (radius matching), estratificação e associação por intervalos (interval matching), além da associação por núcleos (kernel matching).

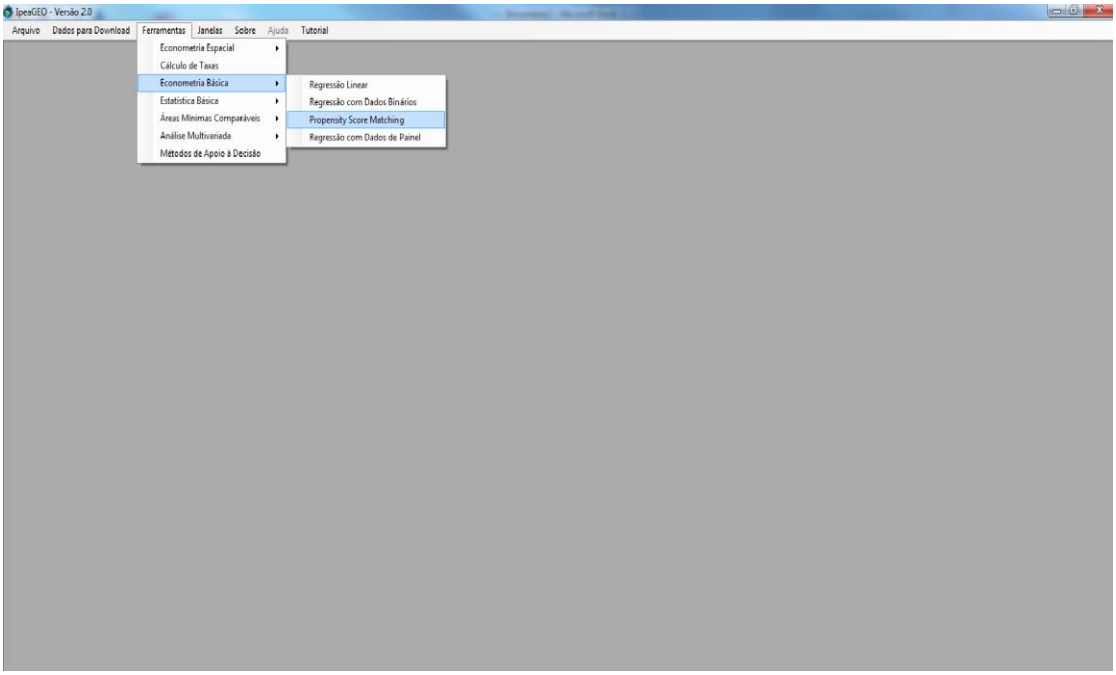
No IpeaGEO os métodos implementados são o vizinho mais próximo (nearest neighbor) e os núcleos Gaussiano, Epanechinikov, Biweight, Triangular e Retangular.

Os parâmetros dos núcleos (*kernels*) e do método do vizinho mais próximo (*nearest neighbor*) podem ser definidos automaticamente pelo IpeaGEO ou então informados manualmente.

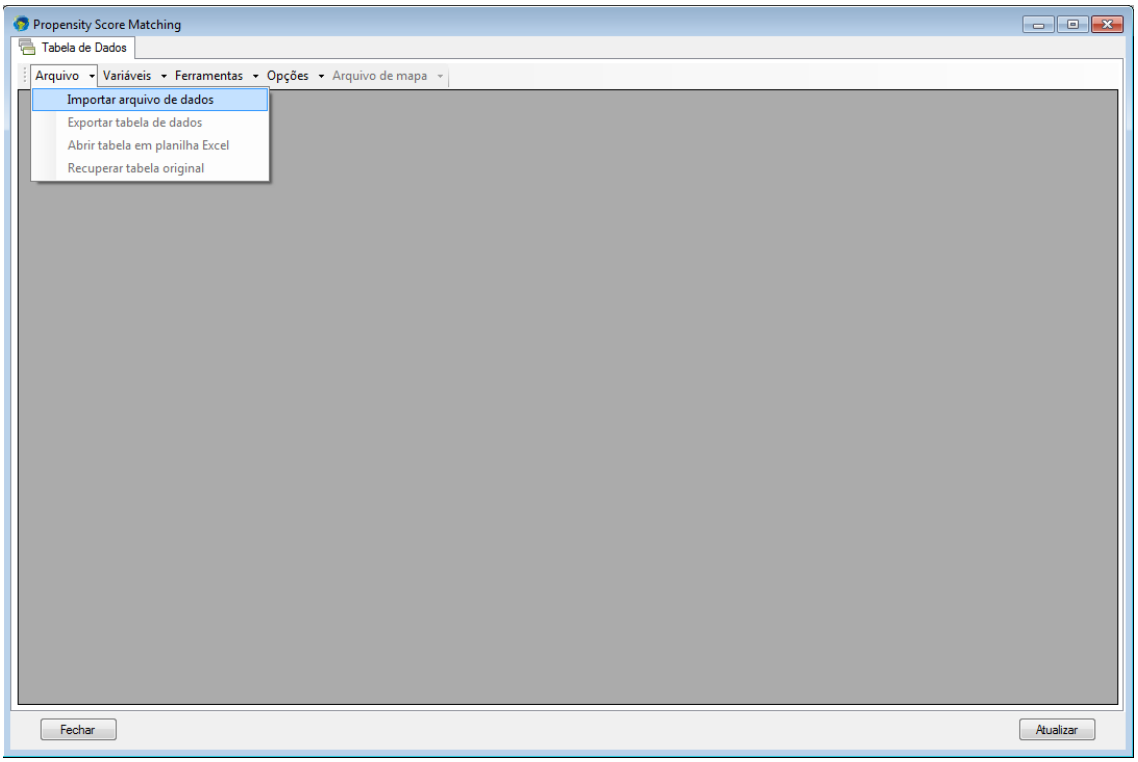
O PSM é uma abordagem útil quando somente as características observadas são críveis como fatores que afetam a participação em determinado grupo ou tratamento, por exemplo, a renda de uma família é um dos fatores explicativos para determinar se uma família vem a ser detentora do programa Bolsa Família ou não.

Exemplo

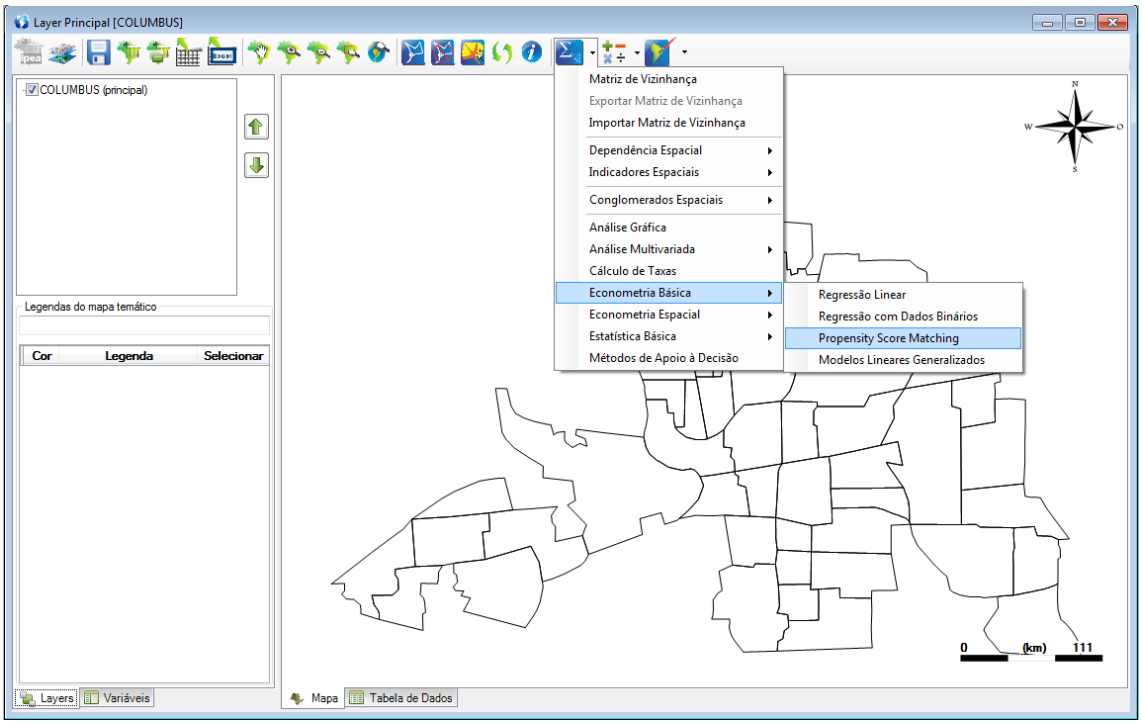
No botão “Ferramentas”**,** clique na opção “Econometria Básica” e, em seguida, na opção “*Propensity Score Matching*”.



No formulário que irá se abrir, selecione a opção “Importar arquivo de dados” em “Arquivo” para escolher a base de dados a ser utilizada na análise.

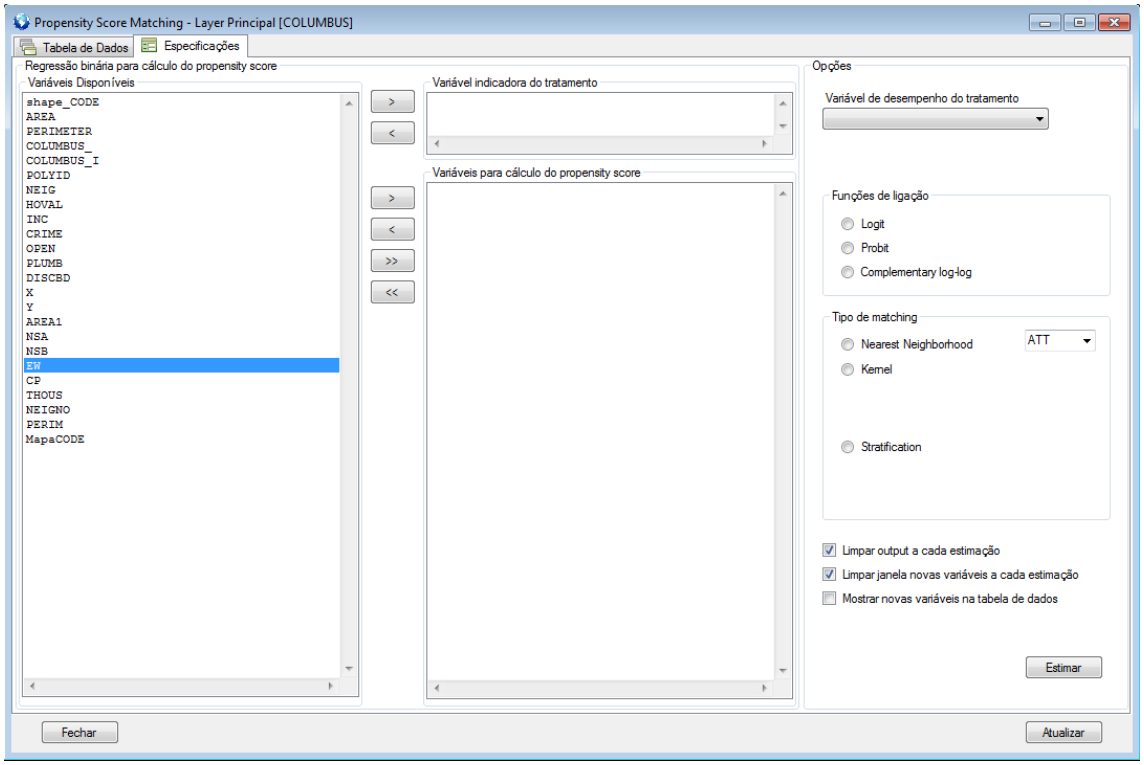


Outra forma de chegar ao formulário do Propensity Score Matching é clicar no ícone “Ferramentas” , opção “Econometria Básica” e em seguida em “Propensity Score Matching” após concatenar o arquivo shape com a tabela de dados (ver tópico de concatenação):

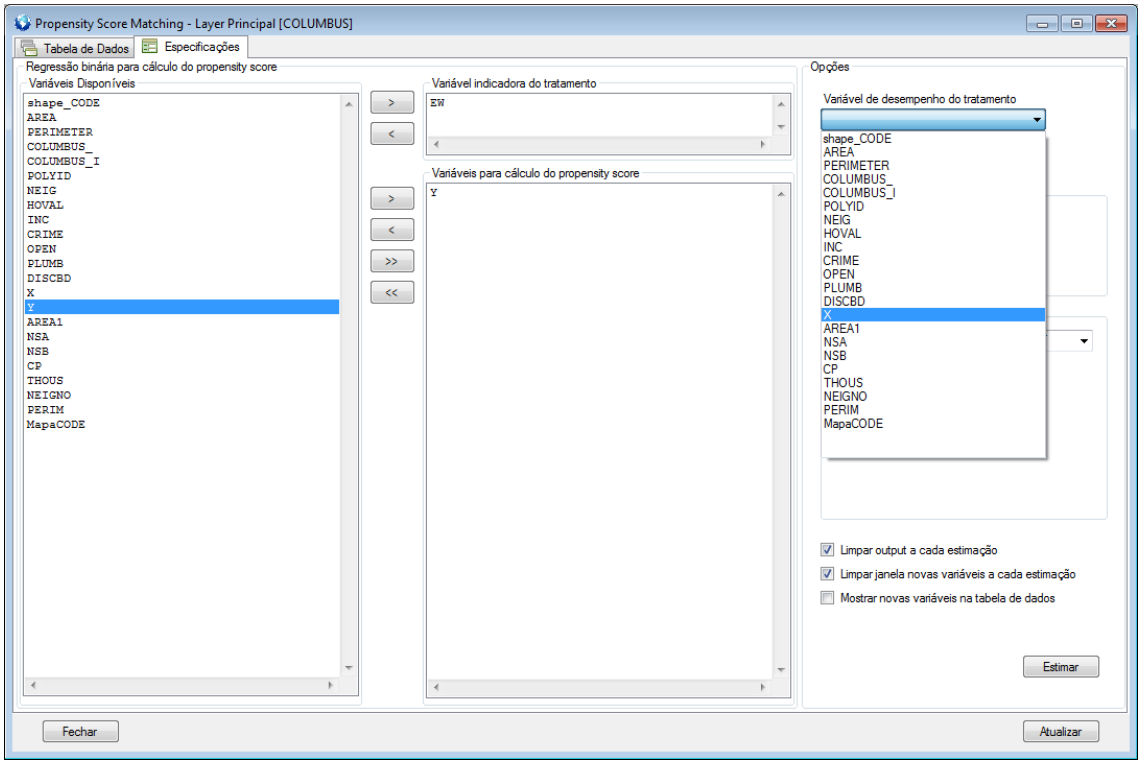


Após importar a tabela de dados (ver tópico Importação de dados), o seguinte

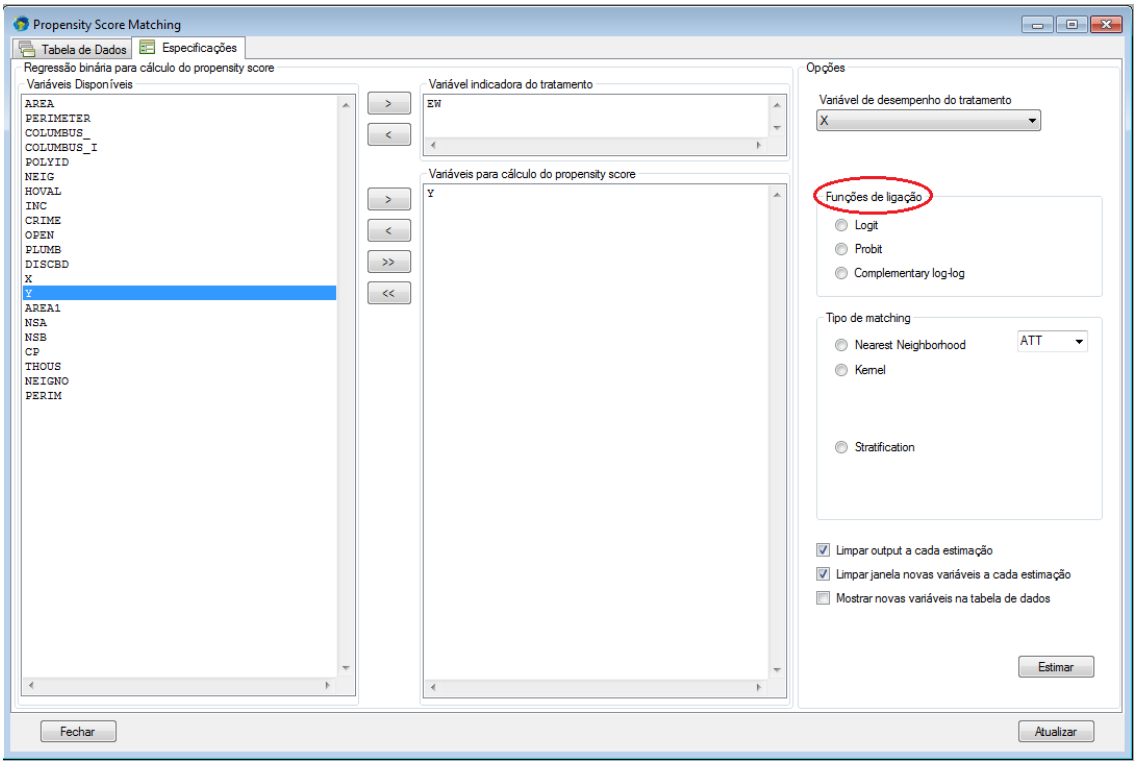
formulário estará à disposição:



O usuário deve selecionar a "Variável indicadora de tratamento" que deve ser binária, as "Variáveis para cálculo do propensity score" e a “Variável de desempenho do tratamento”, como indicado abaixo:

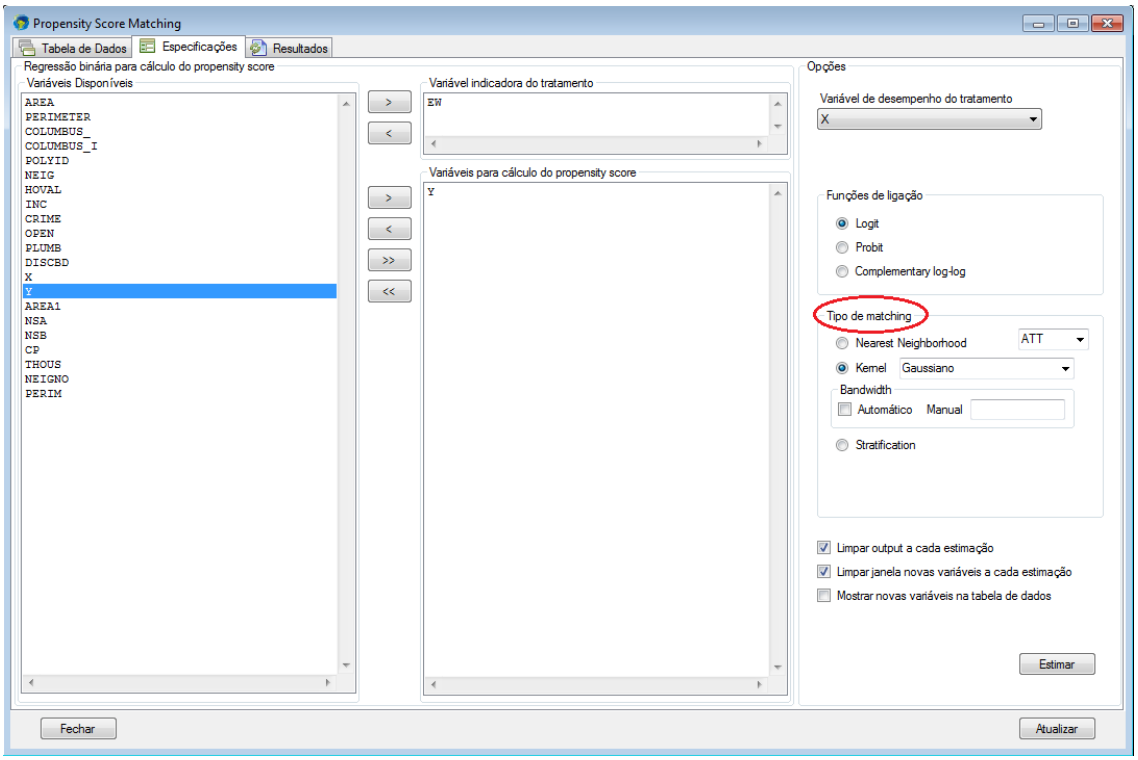


O usuário dispõe das seguintes funções de ligação para realização da análise: Logit, Probit e complemento log-log (Complementary log-log).



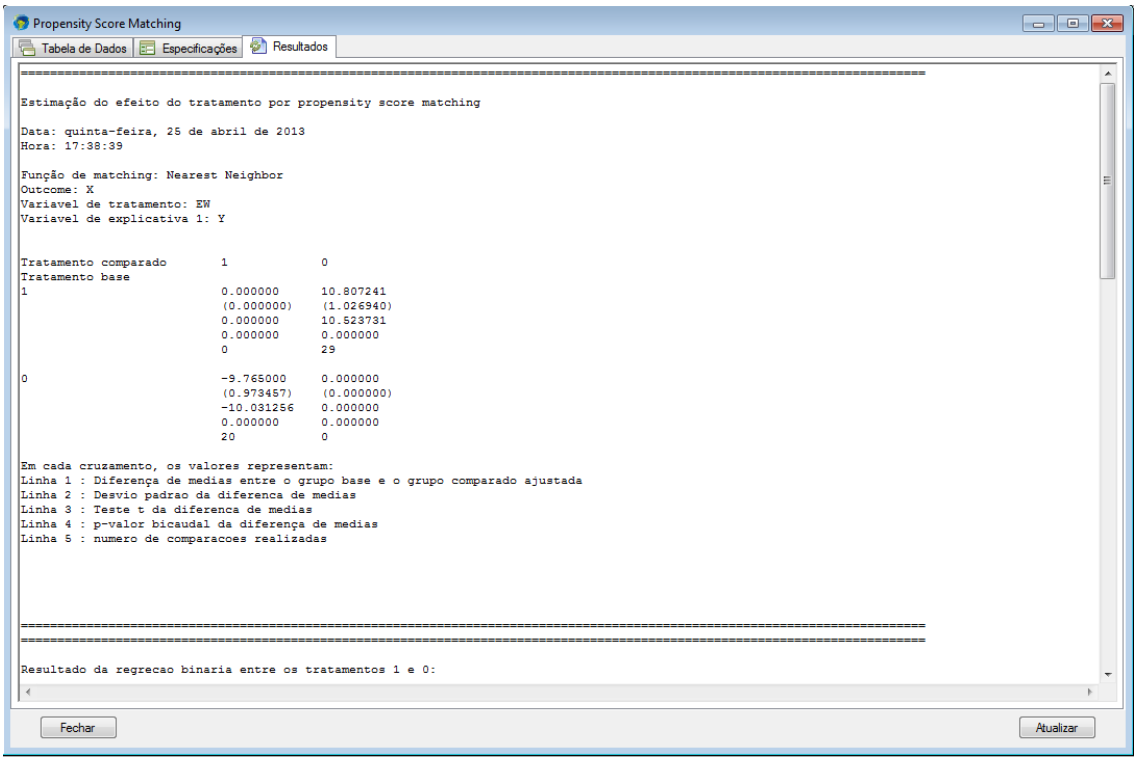
Escolha também o tipo de pareamento (matching) a ser aplicado na analise. As opções disponíveis são: vizinho mais próximo (Nearest Neighbohood), ATT (Average Treatment Effect on Treated – Efeito Médio do Tratamento Sobre os Tratados) e ATE (Average Treatment Effect – Efeito Médio do Tratamento).

Os núcleos (Kernels) de distribuição disponíveis no IpeaGEO são Gaussiano, Epanechnikov, Biweight, Triangular, Retangular, largura de banda (Bandwidth) com valores obtidos manualmente ou automaticamente, e Stratification.



Ainda existem as opções de limpar o resultado da análise (output) a cada estimação, limpar a janela de novas variáveis a cada estimação e mostrar novas variáveis na tabela de dados.

Depois de selecionadas todas as opções desejadas, clique no botão “Estimar”. Uma nova aba chamada “Resultados” irá aparecer contendo as informações da análise.



Regressão com Dados Binários

Visão Geral

Dados binários ou dicotômicos são os que admitem apenas dois resultados para a variável resposta, como por exemplo: i) concessão de crédito de um banco (aprovado ou não aprovado); ii) resultado de exame laboratorial (positivo ou negativo); iii) inspeção de uma peça (defeituosa ou não-defeituosa). Existe também a possibilidade de transformar variáveis contínuas em binárias, classificando em duas faixas de valores (acima e abaixo de um valor): é o que ocorre, por exemplo, quando se classifica se um aluno foi aprovado ou reprovado com base em uma nota mínima de aprovação. De tal forma, os problemas com dados binários podem ser tratados como casos de sucesso (1) e fracasso (0), sendo comumente designado ao sucesso o resultado mais importante da resposta ou o que se deseja relacionar com as variáveis explicativas.

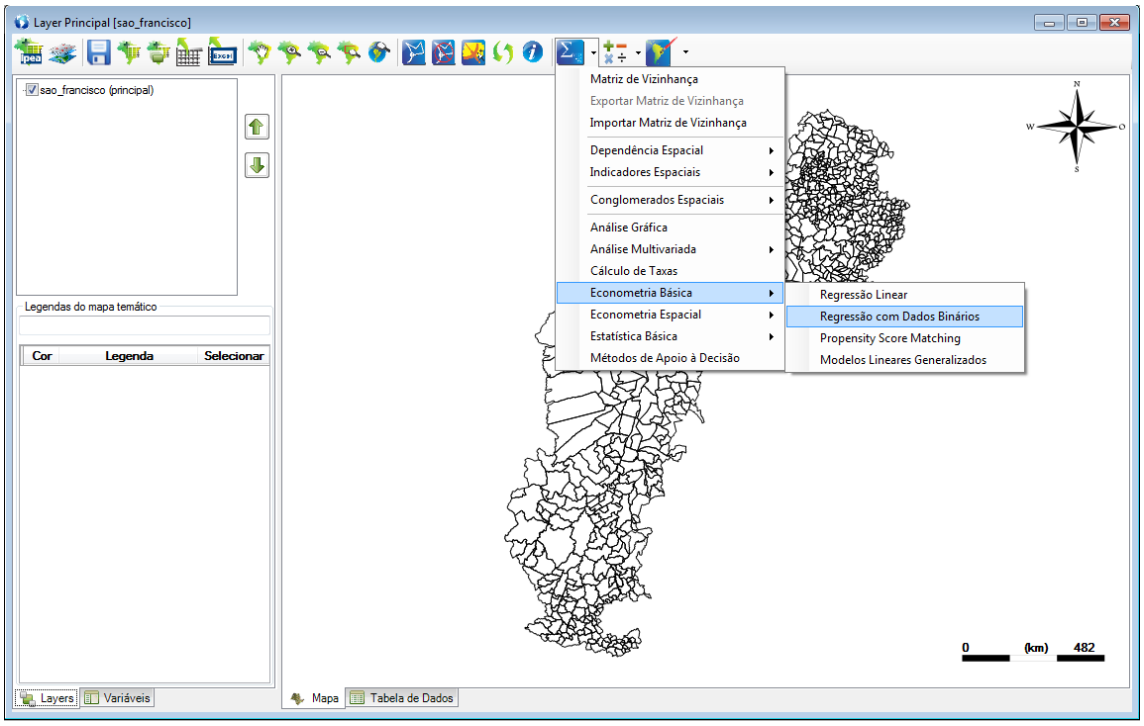
Para dados binários pode-se supor que a distribuição de Bernoulli caracterize bem a variável resposta, aplicando a abordagem dos modelos lineares generalizados (MLG) proposta por Nelder e Wedderburn (1972). É possível escolher uma função de ligação entre o preditor linear do modelo e as covariáveis de modo que se garanta que para qualquer valor estipulado para os parâmetros lineares do modelo, a proporção estimada de sucesso seja um valor no intervalo [0,1].

O IpeaGEO possibilita a estimação da reta de regressão com ou sem intercepto, além de permitir apresentar a matriz de covariância dos coeficientes e realizar a análise de multicolinearidade. As funções de ligação que podem ser utilizadas são a logito (Logit), probito (Probit) e a complementar log-log (Complementary log-log).

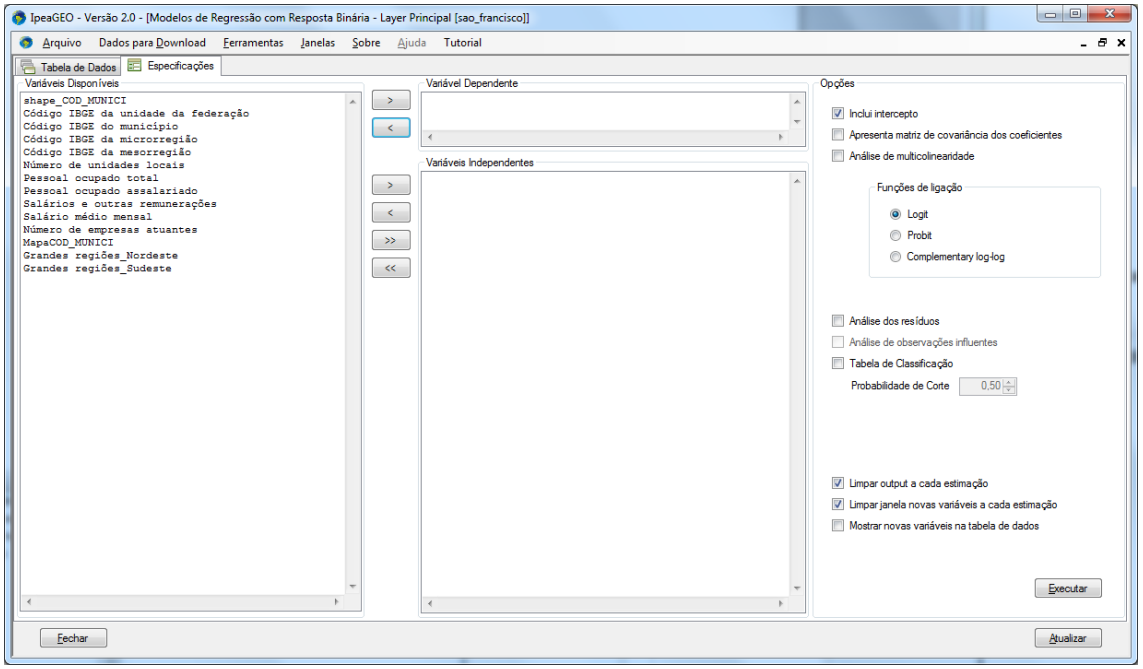
É possível também realizar a análise dos resíduos (com ou sem a análise de observações influentes) e a tabela de classificação (especificando a probabilidade de corte desejada).

Exemplo

Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital com sua tabela de dados (ver tópico de concatenação), clique na opção “Econometria Básica” e em seguida na opção “Regressão com Dados Binários”.



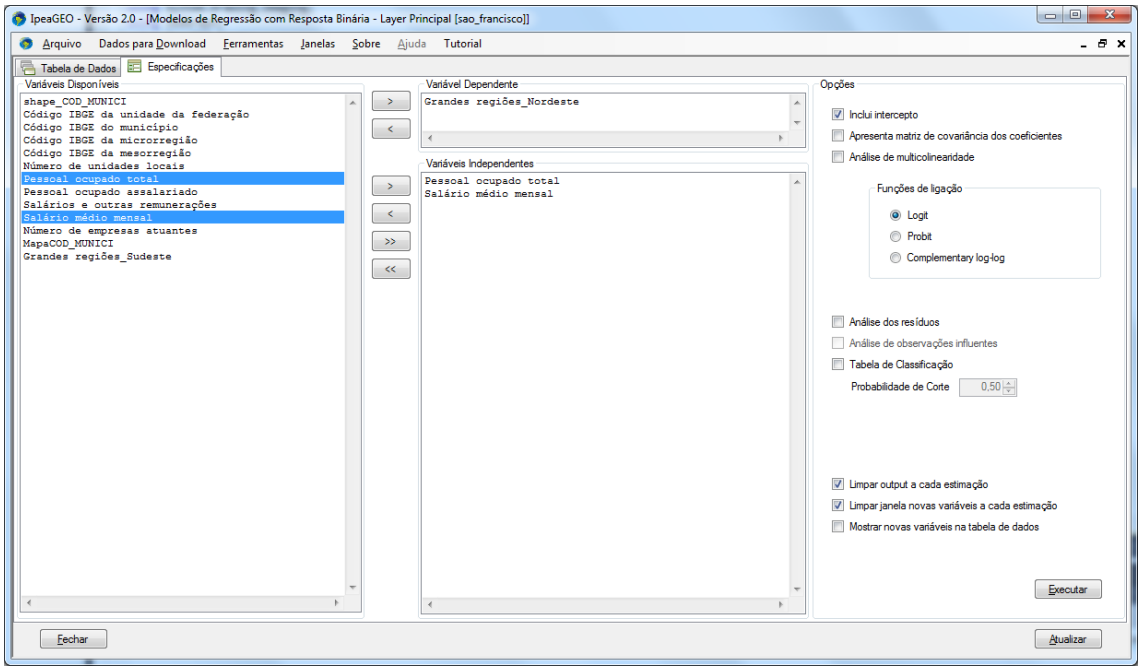
O seguinte formulário irá se abrir:



Perceba que as opções “Incluir Intercepto”, “Logit”, “Limpar output a cada estimação” e “Limpar janela novas variáveis a cada estimação” no lado direito da tela já estão selecionadas, essas opções podem ser alteradas caso seja de interesse do usuário.

Na aba “Especificações”, selecione a “Variável Dependente” de interesse em “Variáveis Disponíveis” e clique no botão [>]. O mesmo procedimento deverá ser feito para as “Variáveis Independentes” que se quer analisar. O botão [>>] transfere todas

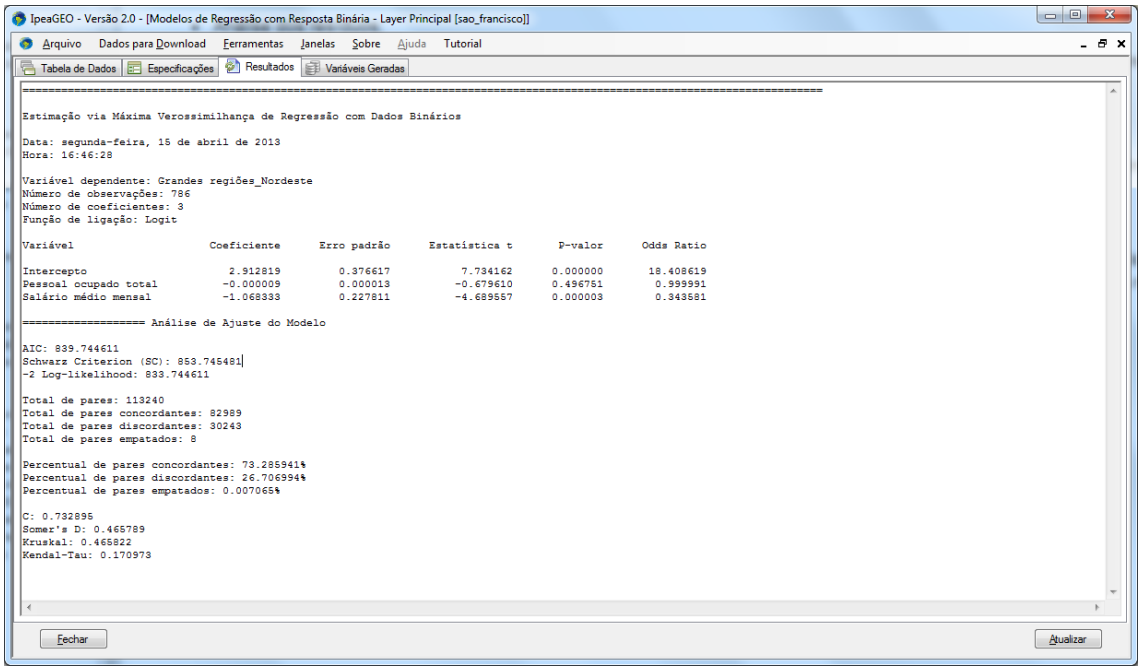
as variáveis disponíveis para a janela de variáveis independentes.

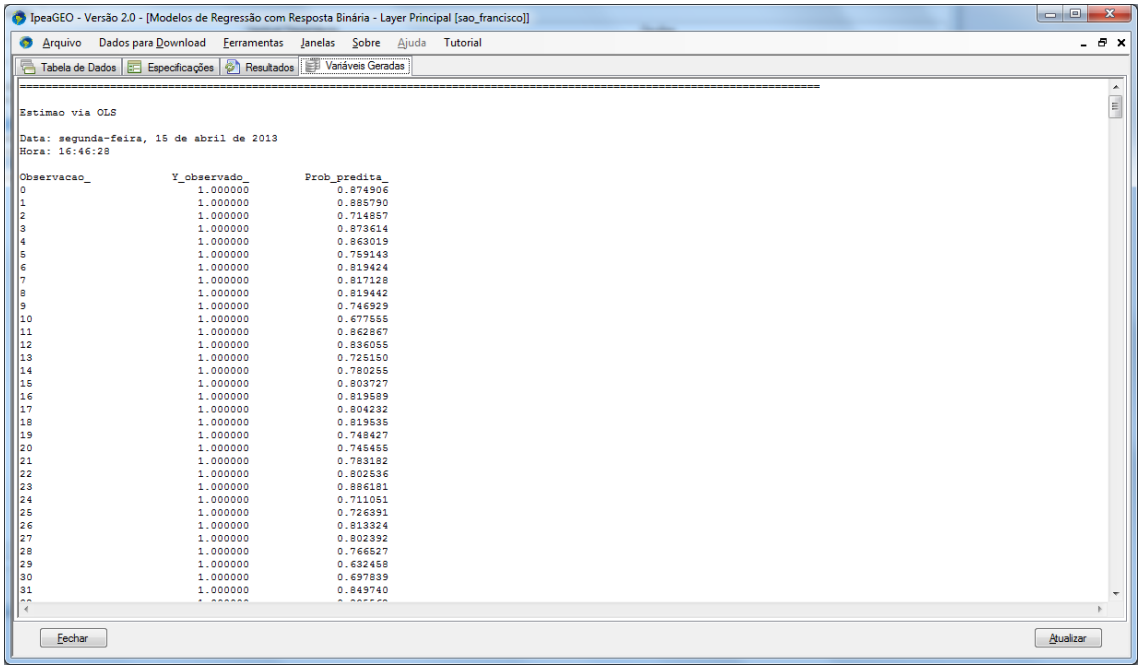


Observe que na aba “Especificações” existem as “Opções para Estimação”, são elas:

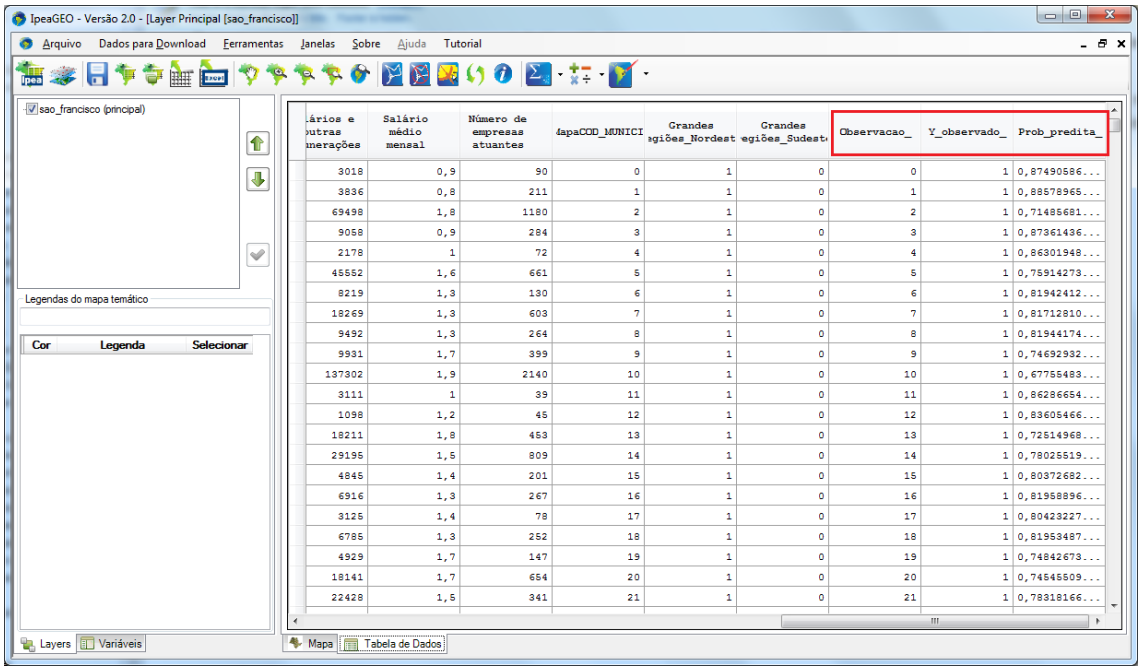
- Incluir Intercepto;
- Apresenta matriz de covariância dos coeficientes;
- Análise de multicolinearidade;
- Funções de ligação (Logit, Probit, Complementary log-log);
- Análise dos resíduos;
- Análise de observações influentes (fica disponível para o usuário apenas quando a “Análise de resíduos” está selecionada);
- Tabela de Classificação (opção “Probabilidade de Corte” varia de 0 a 1).

Depois de selecionadas as opções desejadas, clique no botão “Executar” para realizar a Análise de Regressão. Uma nova aba chamada “Resultados” irá aparecer contendo os resultados da Regressão e as opções selecionadas anteriormente.





Para que os resultados da análise sejam salvos na tabela de dados principal, contida na aba “Tabela de Dados”, basta clicar no botão “Atualizar” localizado no canto inferior direito do formulário. Note que as variáveis “Y_observado”, “Y_predito_” e “Resíduo_” foram incluídas na tabela.



Regressão Linear

Visão Geral

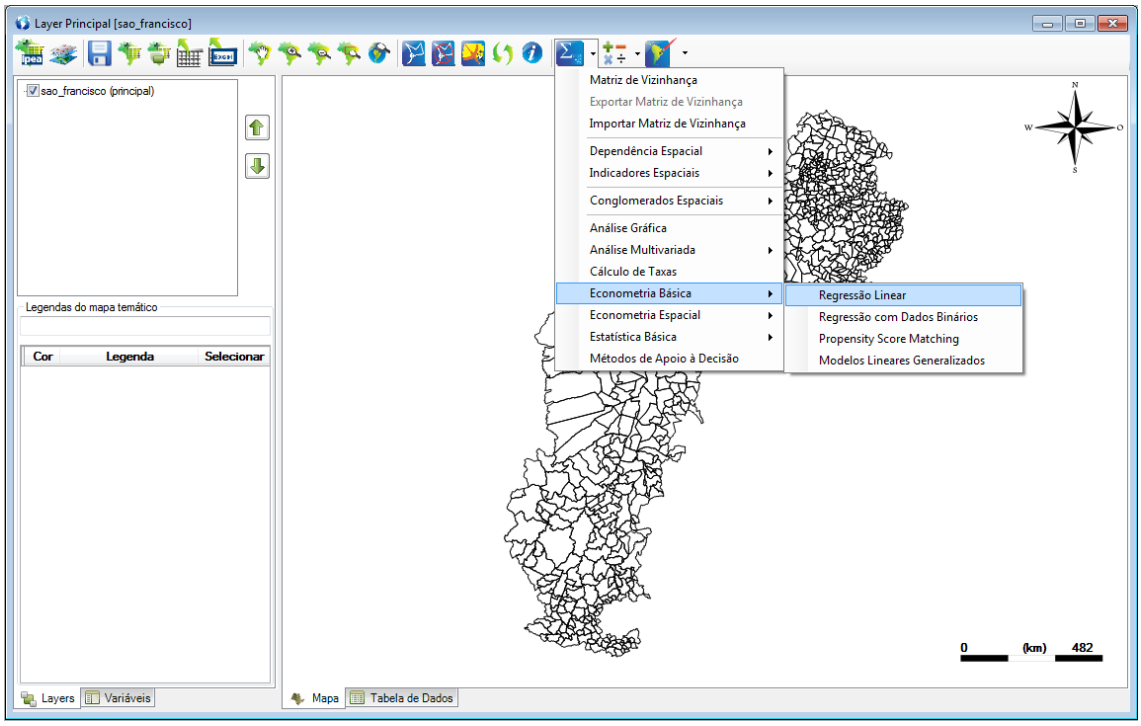
O formulário do IpeaGEO para regressão linear ajusta uma reta de regressão linear – simples(apenas uma variável independente) ou múltipla(mais que uma variável independente). Os coeficientes da equação de regressão são estimados pelo método de mínimos quadrados.

Estão disponíveis, também, as seguintes opções: análise de multicolinearidade e heterocedasticidade (por meio do teste de White, resíduos e presença de observações influentes); e análise gráfica de resíduos (histograma e P-P plot dos resíduos).

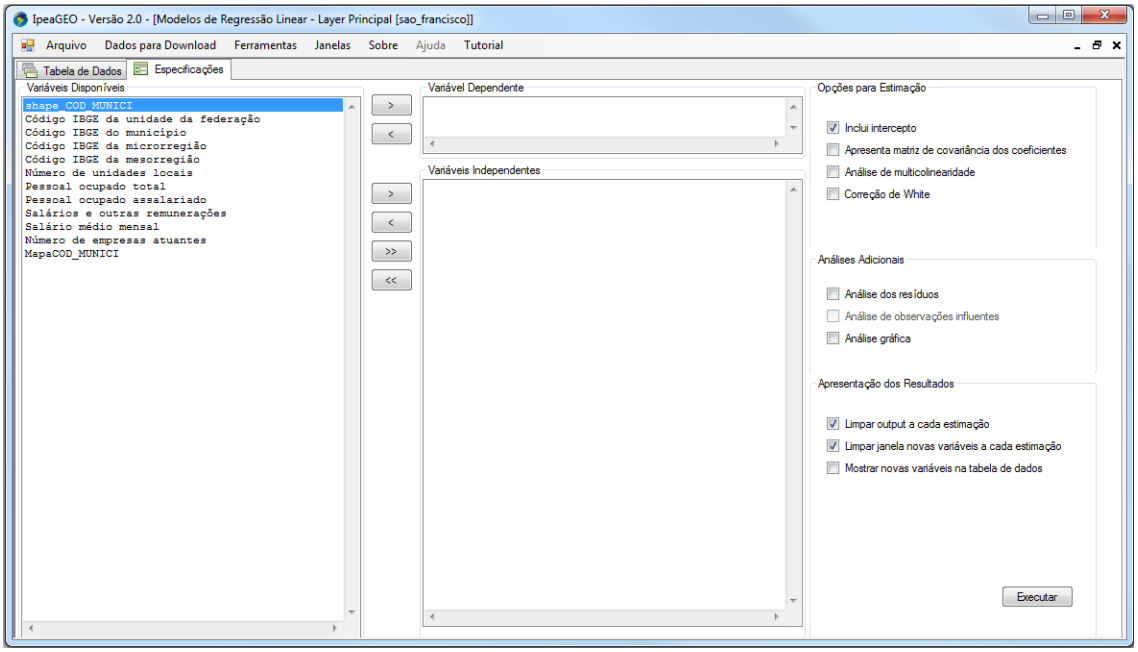
Para a análise de ajuste do modelo são calculados os seguintes coeficientes: coeficiente de determinação (R2), coeficiente de determinação ajustado (R2 ajustado), critério de informação de Akaike (AIC – Akaike Information Criteria), critério de informação de Akaike para pequenas amostras (AICc – Akaike Information Criteria Corrected), critério de informação bayesiano (BIC – Bayesian Information Criteria) e estatística de Wilk (-2 * log-verossimilhança).

Exemplo

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital com a tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , na opção “Econometria Básica” e em seguida na opção “Regressão Linear”.

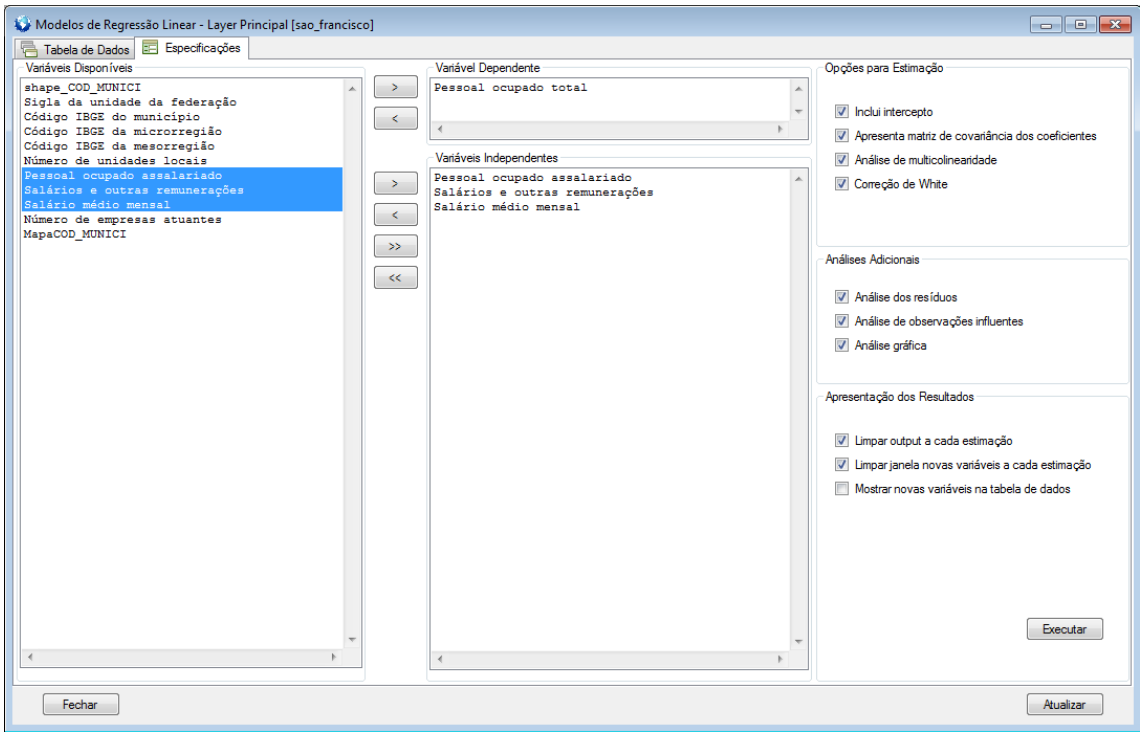


O seguinte formulário irá se abrir:



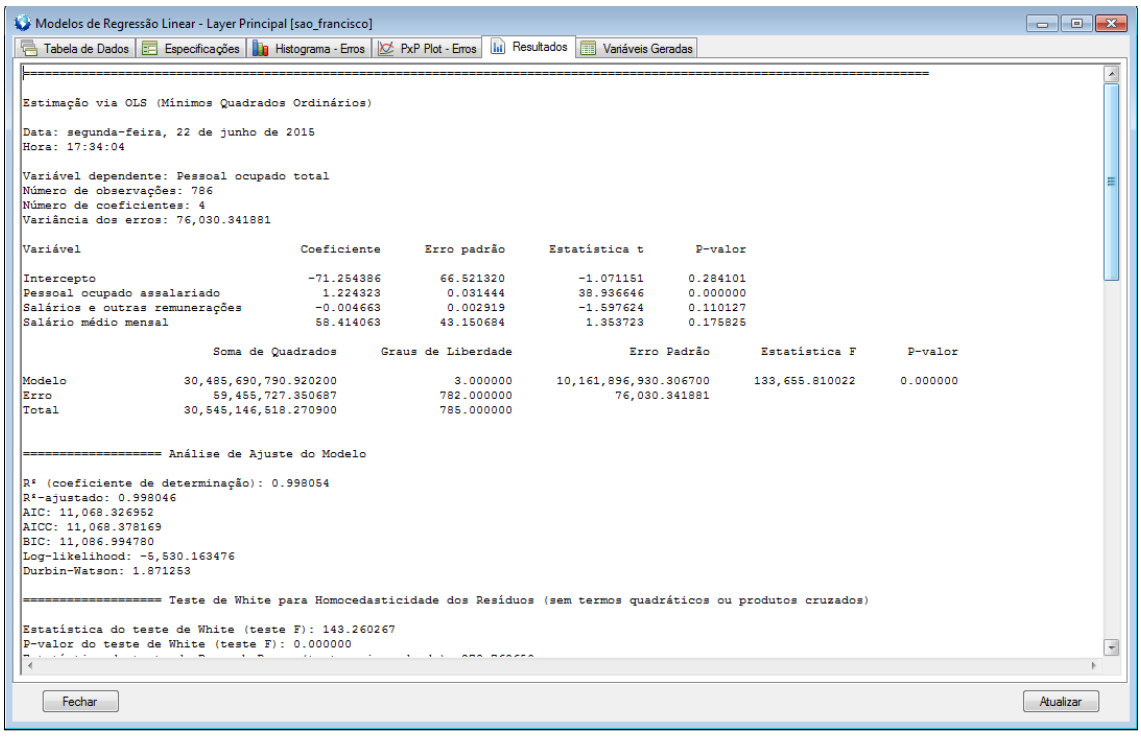
Perceba que as opções “Incluir Intercepto”, “Limpar output a cada estimação” e “Limpar janela novas variáveis a cada estimação” no lado direito da tela já estão selecionadas, essas opções podem ser desmarcadas caso seja interesse do usuário.

Na aba “Especificações”, selecione a “Variável Dependente” de interesse em “Variáveis Disponíveis” e clique no botão [>]. O mesmo procedimento deverá ser feito para as “Variáveis Independentes” que se quer analisar. O botão [>>] transfere todas as variáveis disponíveis para a janela de variáveis independentes.

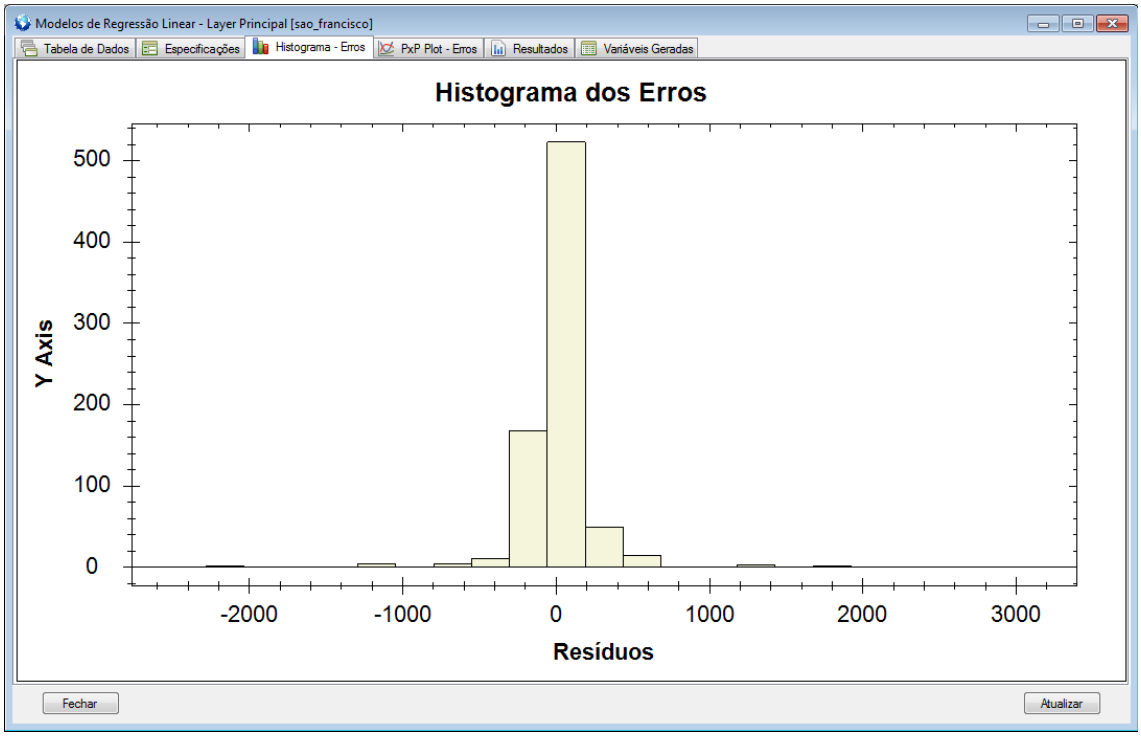


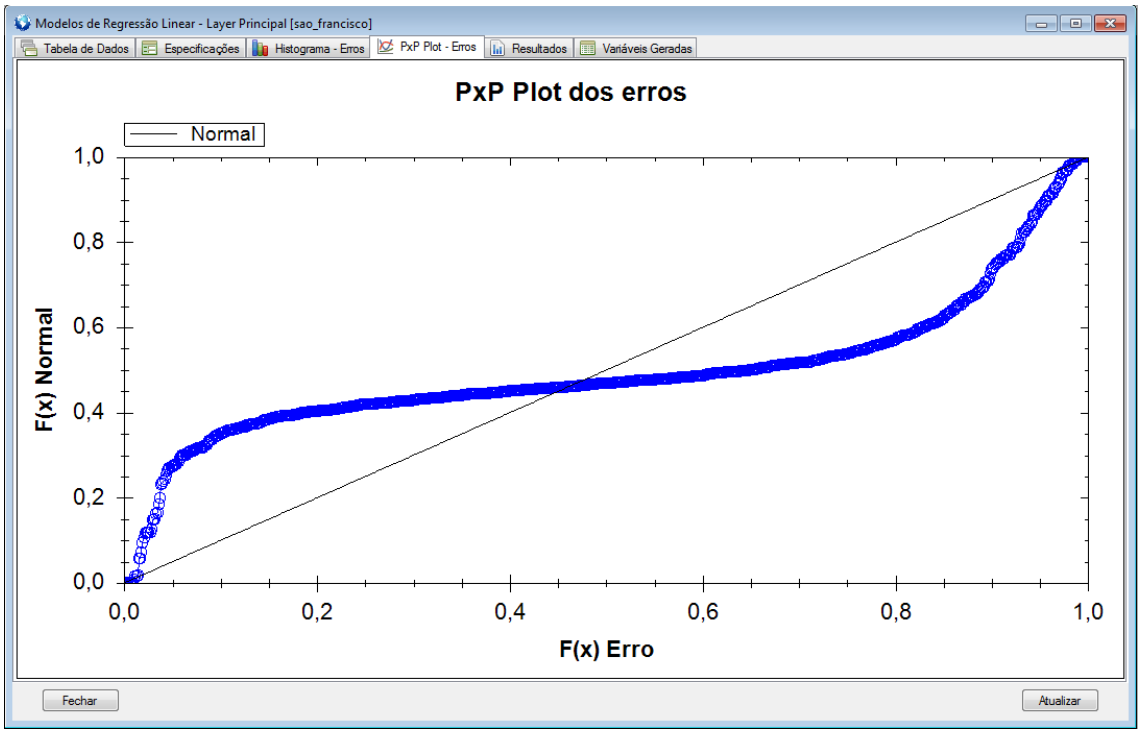
Observe que na aba “Especificações” existem as “Opções para Estimação”, são elas: incluir intercepto, apresentar a matriz de covariância dos coeficientes, efetuar análise de multicolinearidade, aplicar a correção de White para heterocedasticidade. Também é permitido realizar análise de resíduo, análise de observações influentes e análise gráfica.

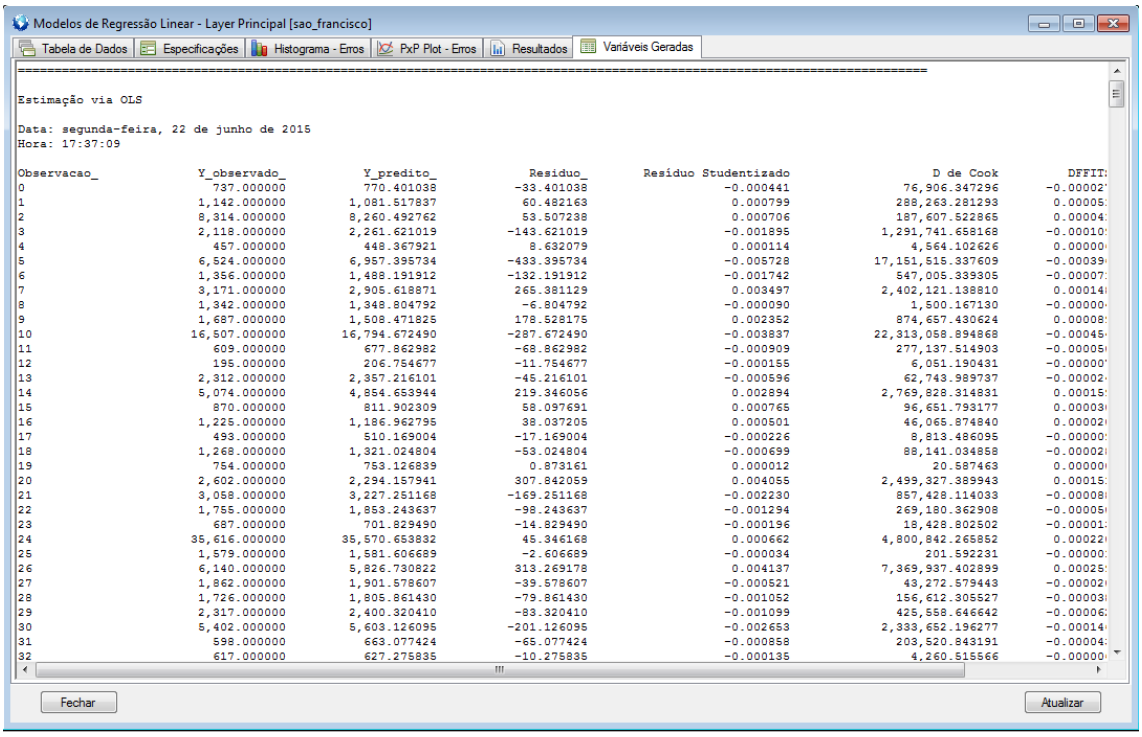
Depois de selecionadas as opções desejadas, clique no botão “Executar” para realizar a Análise de Regressão. Uma nova aba chamada “Resultados” irá aparecer contendo os resultados da Regressão e as opções selecionadas anteriormente.



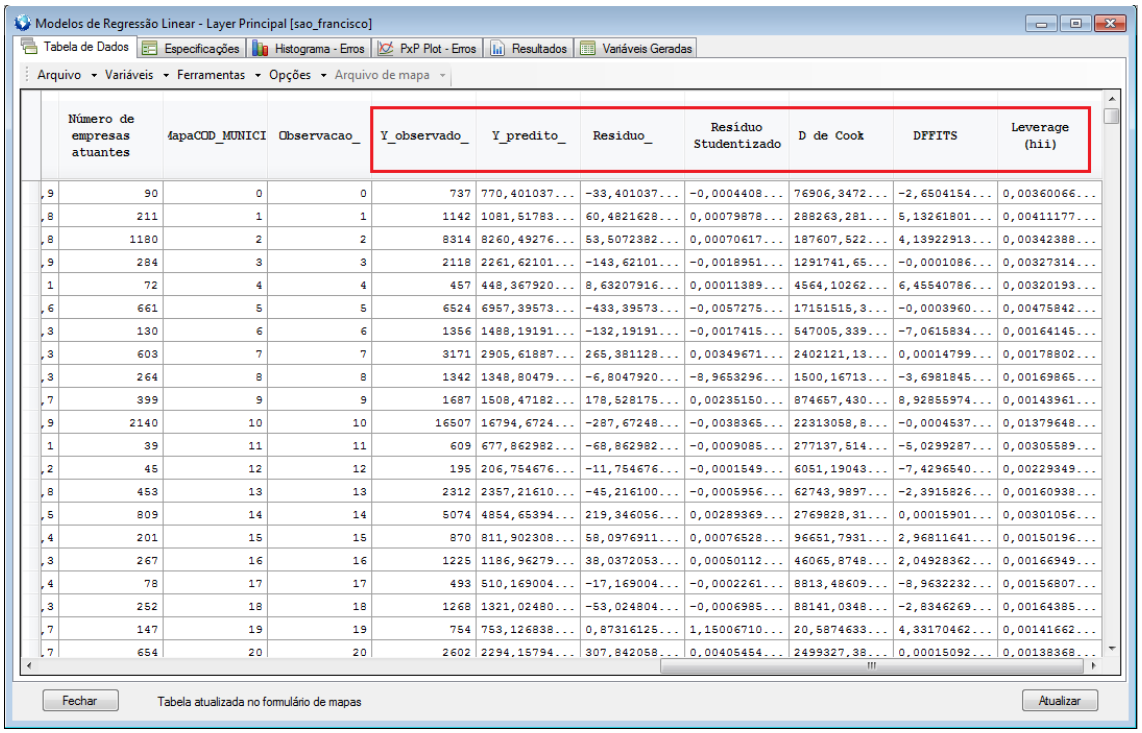
Ao selecionar as opções contidas em “Análises Adicionais” três novas abas são criadas contendo as informações solicitadas. Estas abas estão exemplificadas abaixo.







Para que os resultados da análise sejam salvos na tabela de dados principal, contida na aba “Tabela de Dados”, basta clicar no botão “Atualizar” localizado no canto inferior direito do formulário. Note que, para este exemplo, as variáveis “Y_observado”, “Y_predito_”, “Resíduo_” dentre outras foram incluídas na tabela.



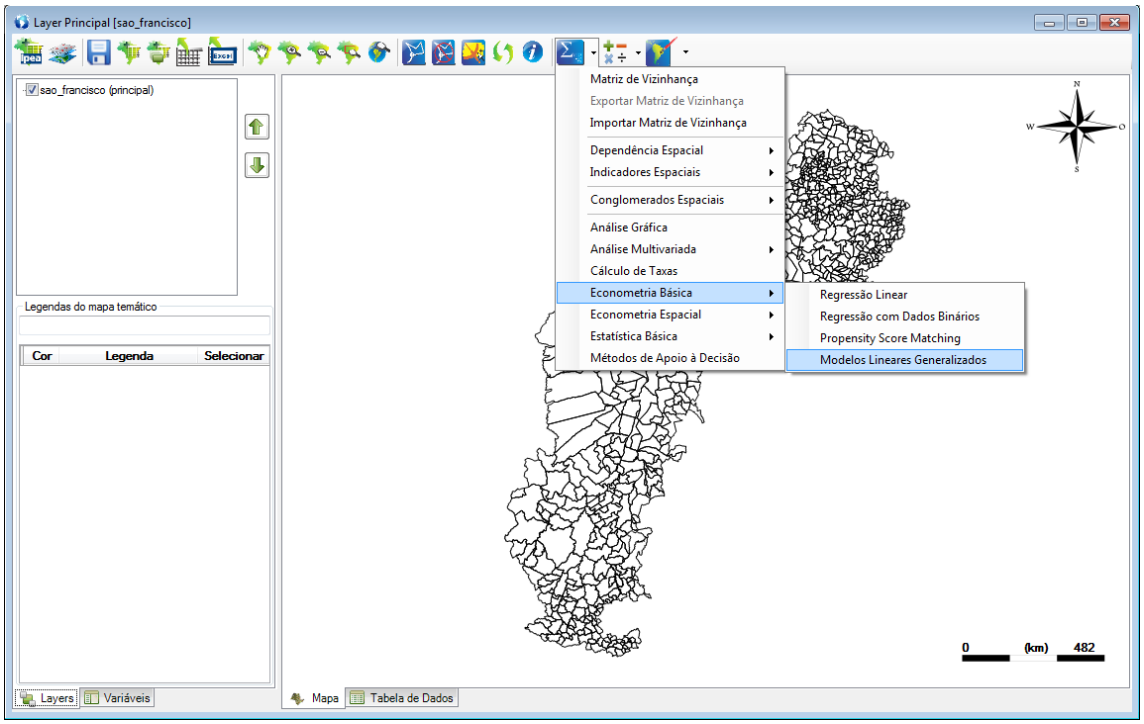
Modelos Lineares Generalizados

Visão Geral

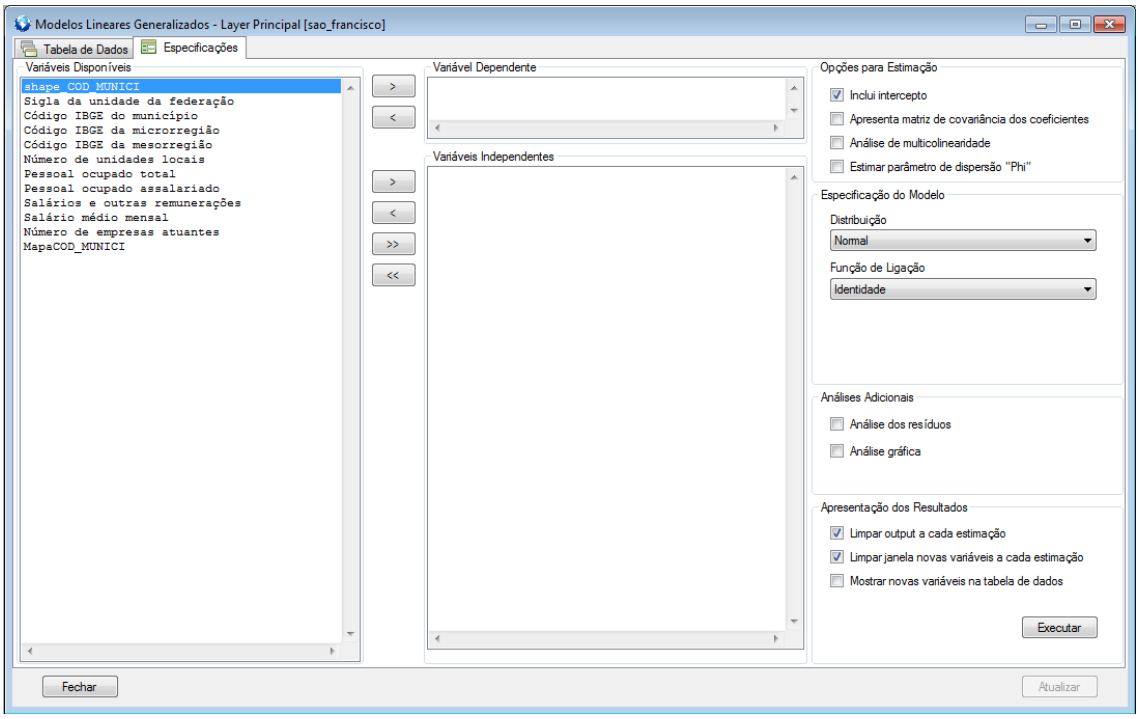
Os modelos Lineares Generalizados, assim como na regressão linear, têm como objetivo analisar a influência de uma ou mais variáveis sobre outra de interesse, com o grande diferencial de expandir esse conceito para gerar modelos com menos pressupostos. Essa técnica relaciona a variável de interesse com as explicativas por meio de funções de ligação

Exemplo

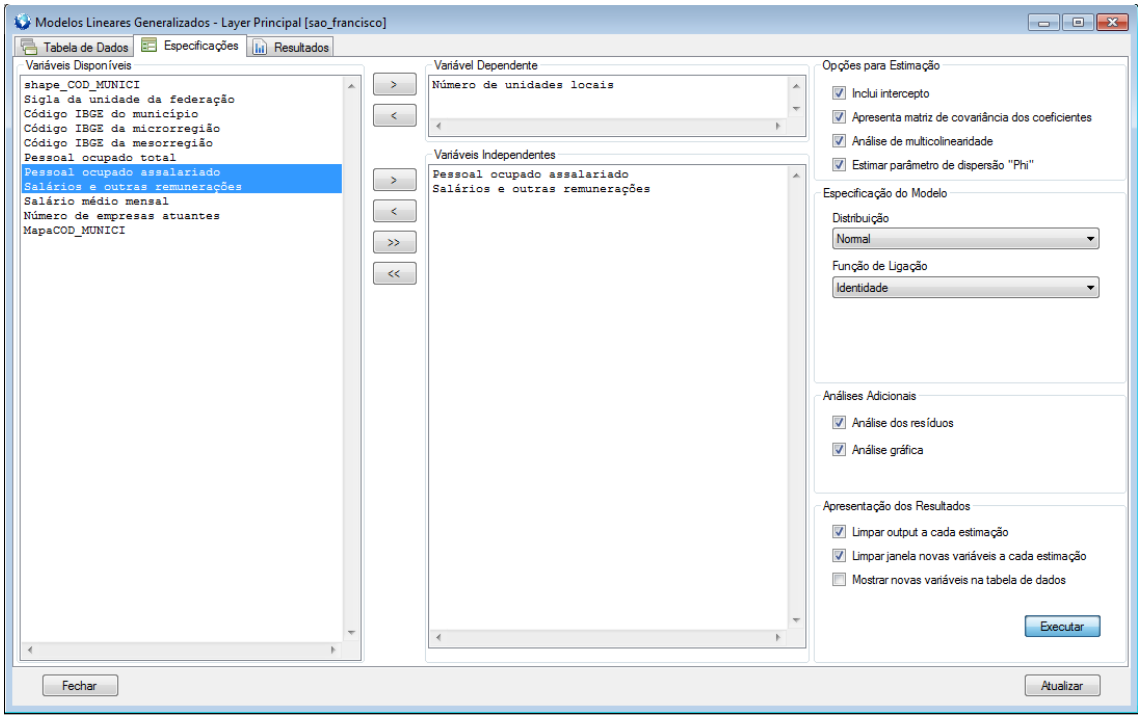
Na janela principal (formulário mapas), selecione Ferramentas → Econometria Básica → Modelos Lineares Generalizados.



O formulário “Modelos Lineares Generalizados” abrirá:



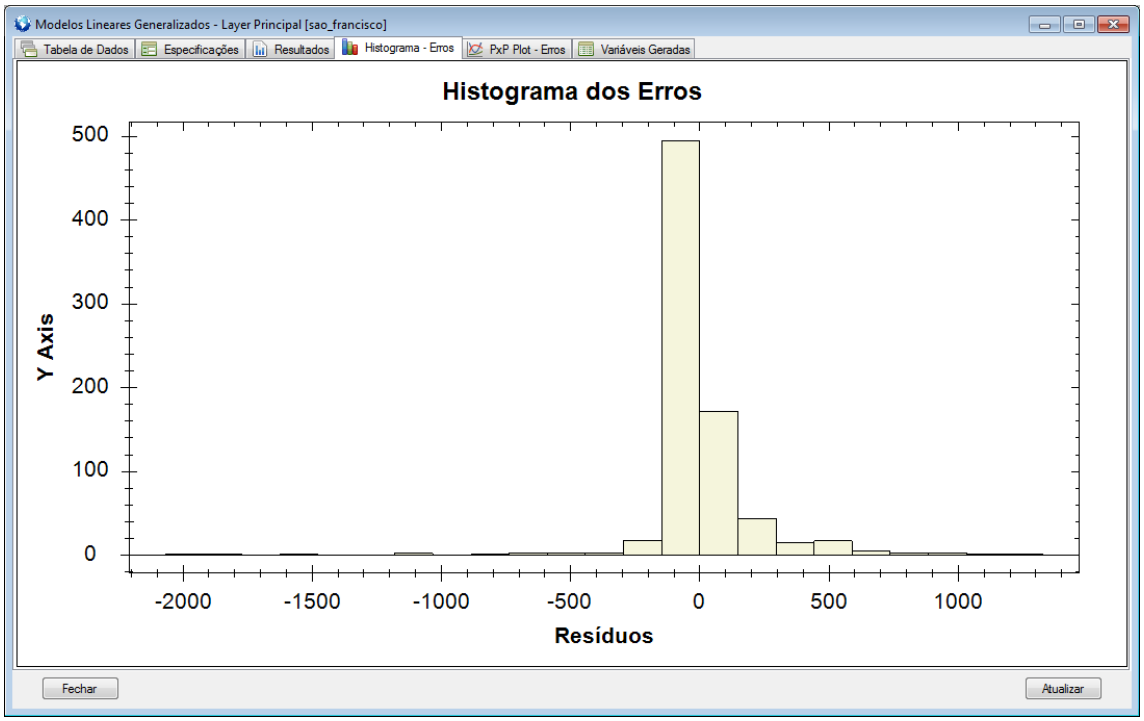
Neste formulário, as Opções para Estimação estão localizadas do lado superior direito e são: Incluir Intercepto, Matriz de Covariância dos Coeficientes, Multicolinearidade e Estimação do Parâmetro de Dispersão. Caso o usuário não selecione a opção de estimar o parâmetro de dispersão, será utilizado um parâmetro com valor igual a 1.



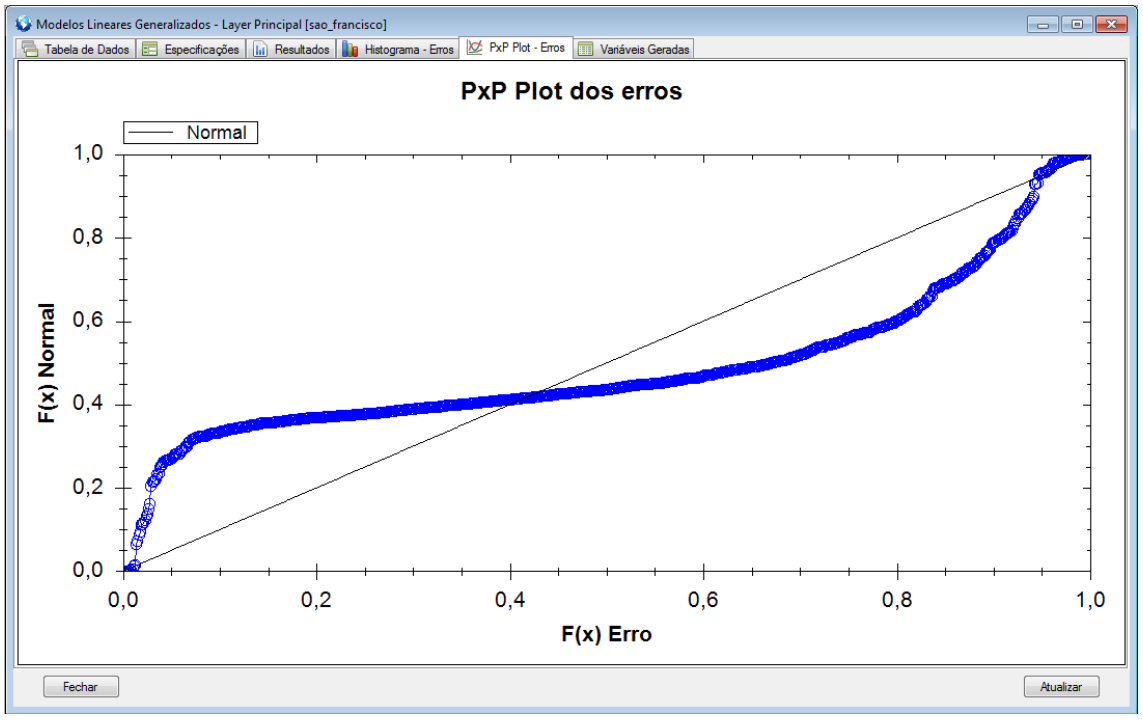
Na Especificação do Modelo, as distribuições de probabilidade são a Normal, Poisson, Bernoulli, Gamma, Geométrica e Binomial Negativa. Junto com a escolha da

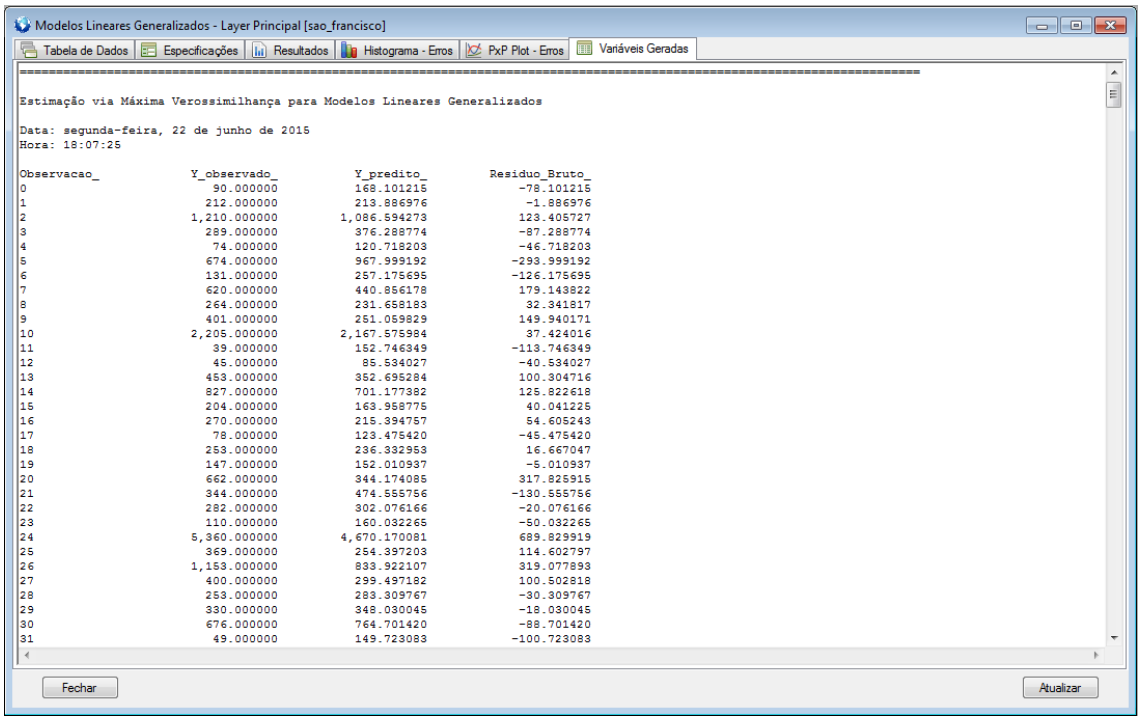
distribuição existe a escolha da função de ligação do modelo, composta por: Identidade, Logaritma, Logit, Probit, Power e Complementary log-log.

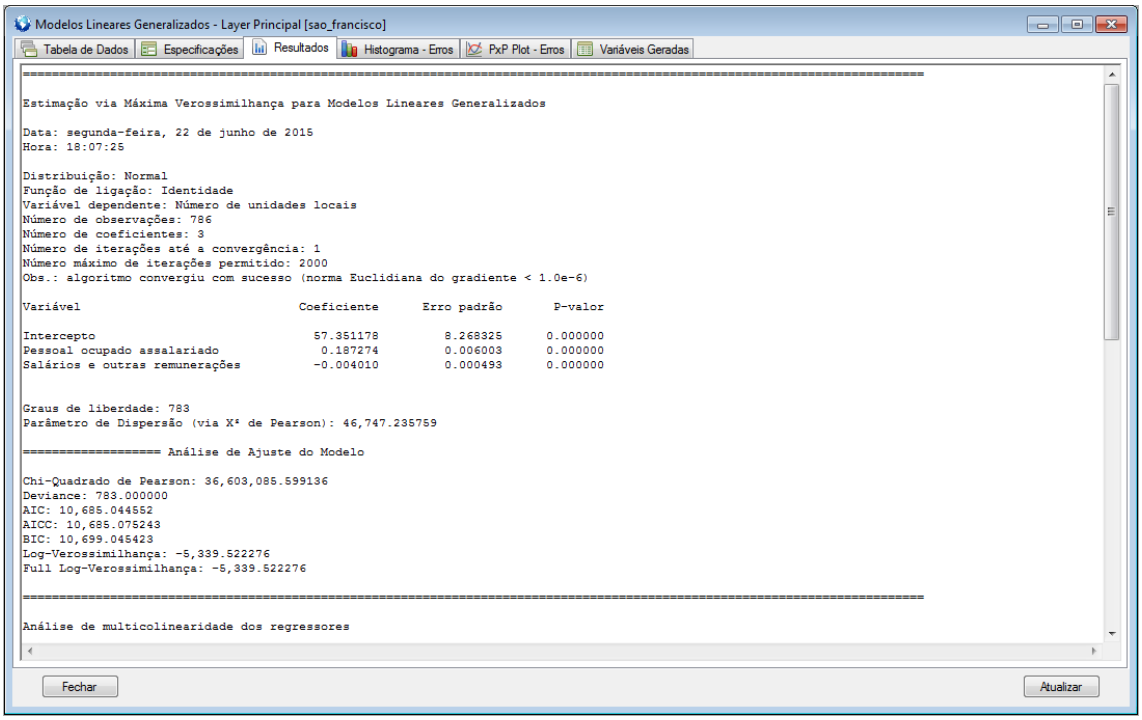
Logo abaixo aparecem as análises adicionais que o usuário tem a disposição e em seguida como o usuário quer que o resultado seja apresentado.



Ao selecionar todas as caixas de opções, algumas abas são criadas. A aba Histograma – Erros mostra como os resíduos estão distribuídos e a aba PxP Plot (que aparece quando se escolhe a distribuição Normal) exibe o gráfico PxP plot dos resíduos. As duas últimas abas exibem o resultado e as variáveis geradas.







Estatística Básica

Análise e testes para médias

Visão Geral

Esse formulário do IpeaGEO permite ao usuário a aplicação de testes estatísticos paramétricos, não paramétricos e testes de ajuste.

De acordo com **Bussab & Morettin (2002)**, os testes de hipóteses utilizam amostras para verificar a adequação desta hipótese ao universo real, ou seja, à população da qual foi amostrada. Têm como objetivo fornecer uma metodologia que permita o usuário verificar se os dados amostrais trazem evidências que apoiam ou não uma hipótese estatística formulada.

Estes testes servem essencialmente para descrever uma população a partir de uma amostra, comparar essa população a um valor hipotético, comparar duas ou mais amostras independentes ou pareadas, comparação de médias ou de variâncias, dentre outros.

O IpeaGEO possui as seguintes opções de testes: Testes de ajuste (aderência), comparação não paramétrica de médias e comparação paramétrica de médias.

Na opção de “Testes de ajuste*”* (aderência), o usuário pode realizar os seguintes testes:

- Teste de Kolmogorov-Smirnov para as distribuições Normal, Exponencial, Uniforme e Poisson.
- Teste de qui-quadrado;

Na opção *comparação não paramétrica de médias* é possível realizar os testes de Kruskall-Wallis, Mann-Whitney e Wilcoxon.

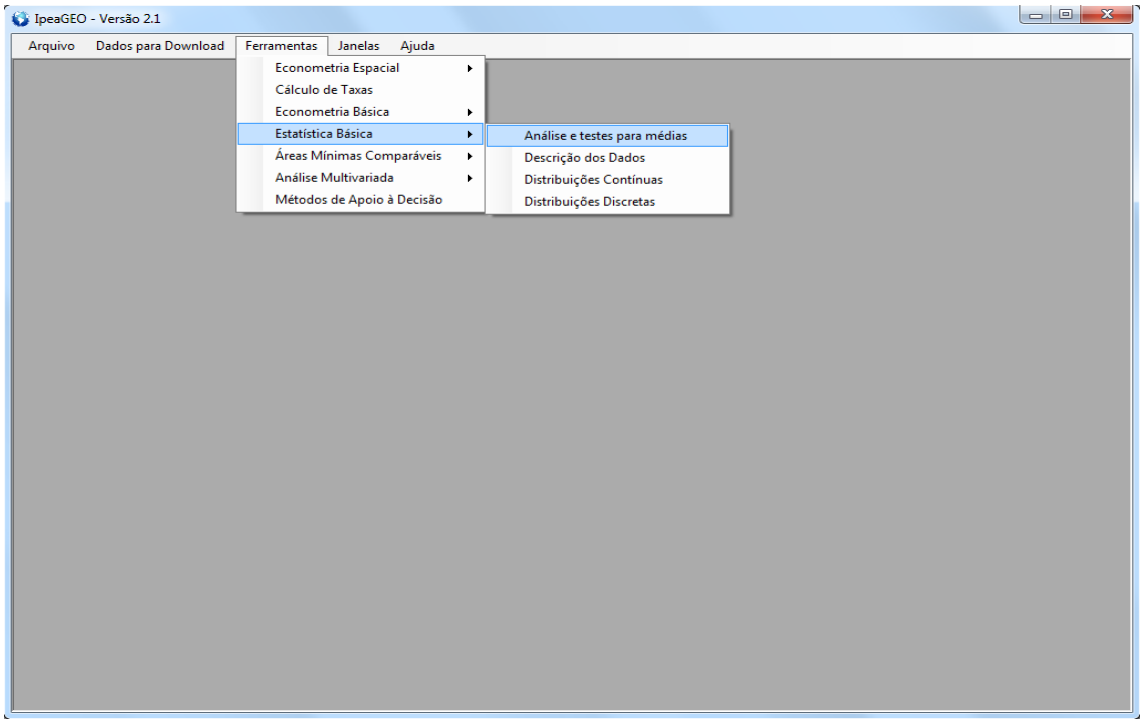
Na opção “comparação paramétrica de médias”, é possível realizar os testes t-Student para amostras independentes (com variâncias conhecidas ou desconhecidas) e para amostras pareadas; além do modelo de análise de variância (ANOVA – Analysis of Variance).

Exemplo

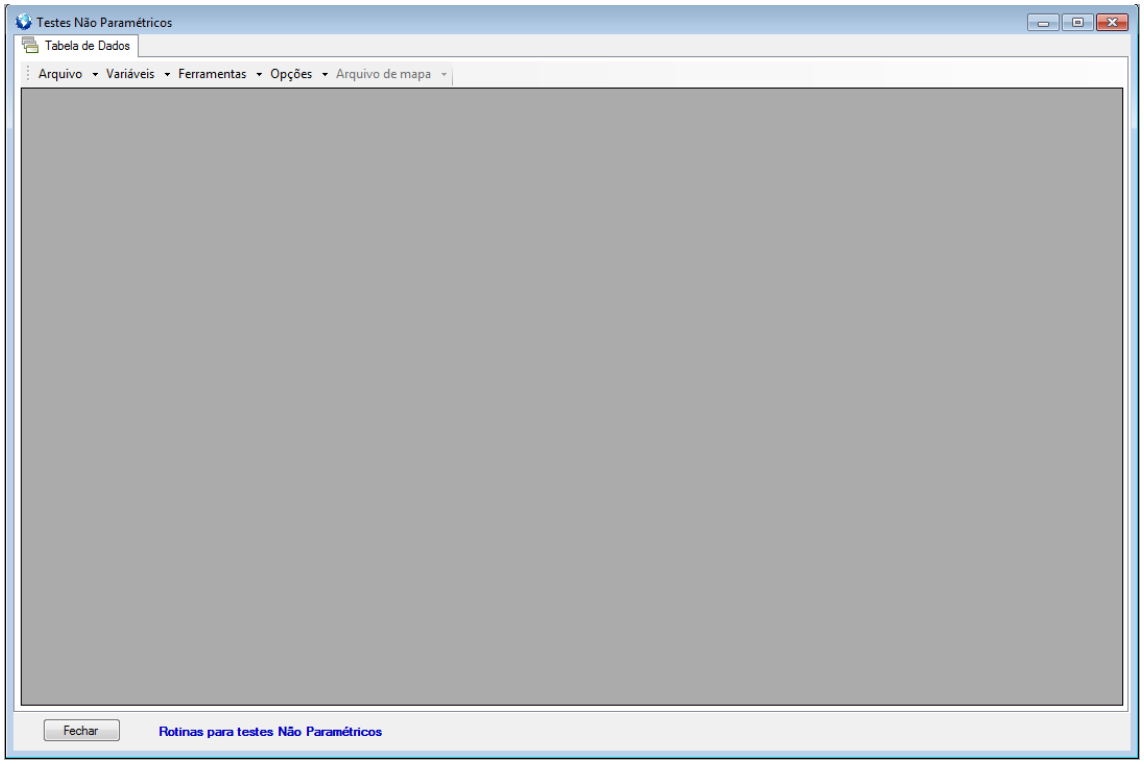
Existem duas maneiras para realizar tais testes no IpeaGEO:

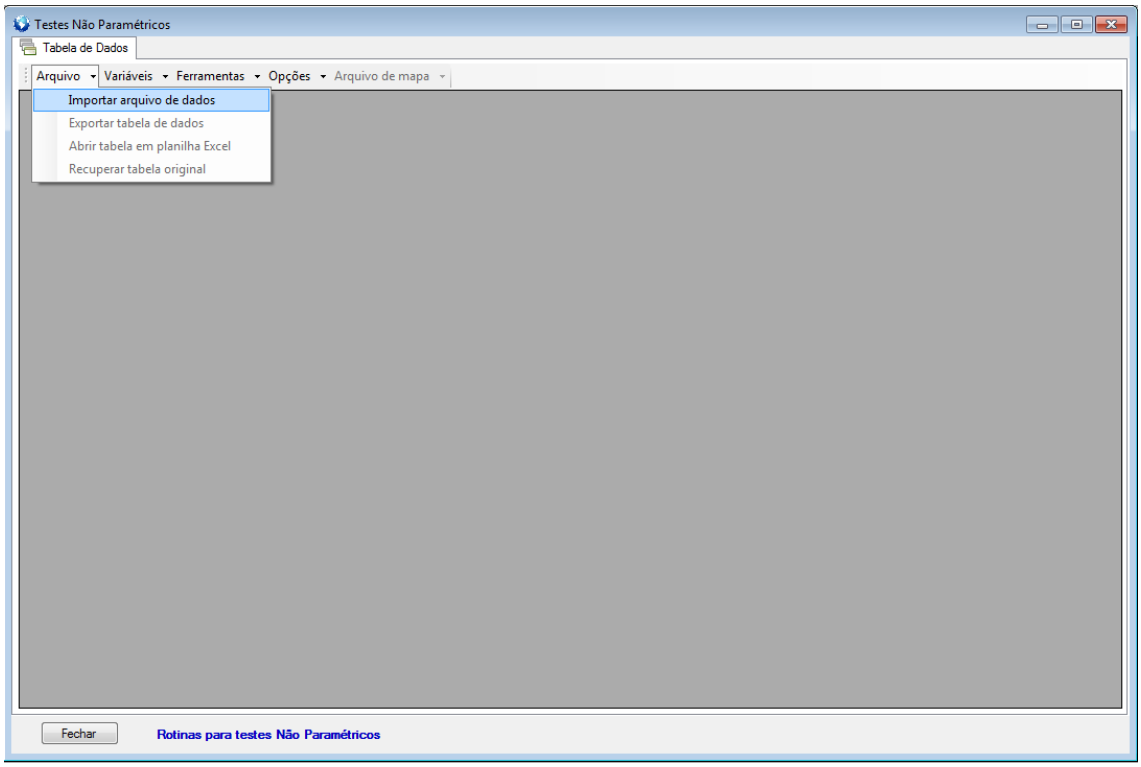
- Direto do menu principal:

No ícone "Ferramentas", clique na opção “Estatística Básica” e em seguida na opção “Análise e Testes para Médias”.



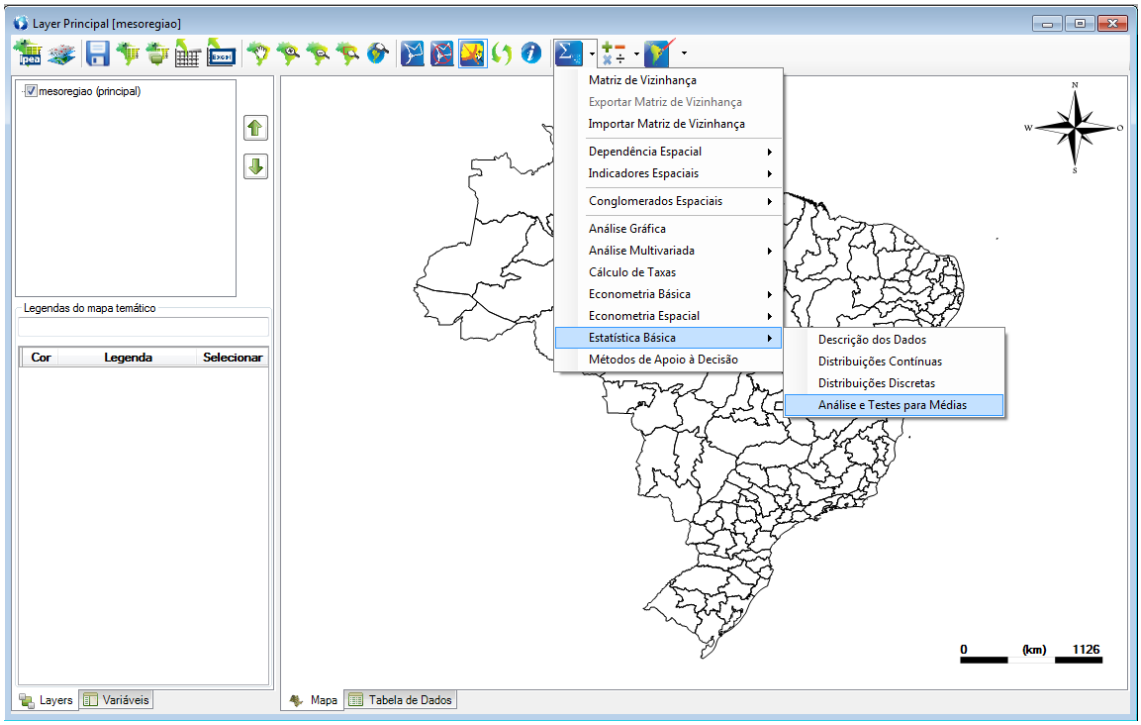
Ao clicar em “Análise e Testes para Médias” abrirá uma aba chamada “Tabela de Dados”, na qual o usuário tem que importar um arquivo de dados do seu computador, clicando em "Arquivo" e em seguida em "Importar arquivo de dados".



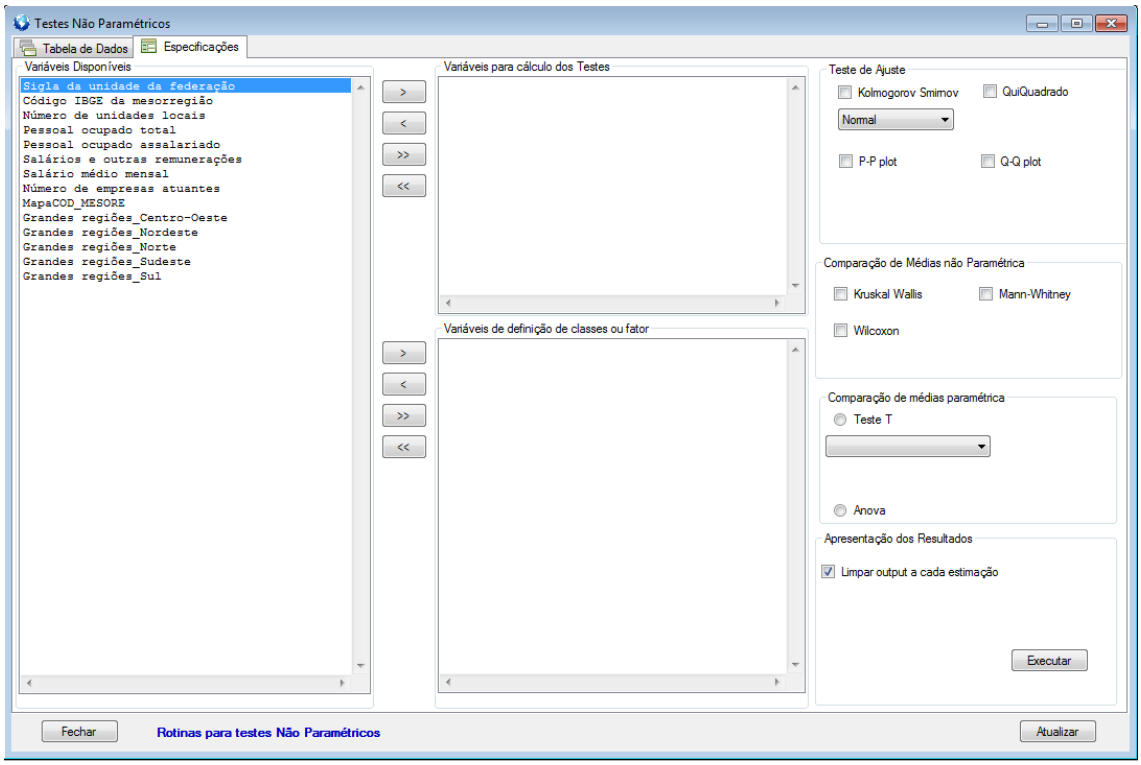


- <u>A partir da importação de uma malha digital (shape):</u>

Clique no ícone de “Ferramentas” , opção “Estatística Básica” e em seguida na opção “Análise e Testes para Médias”.

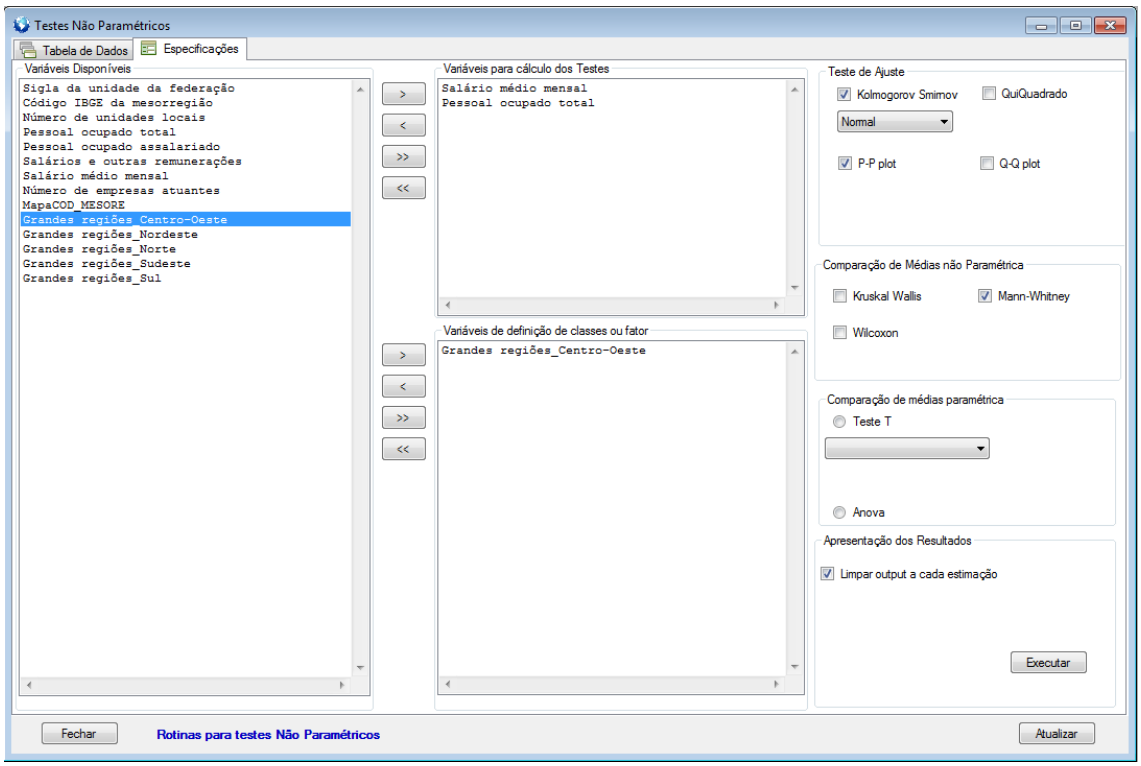


O seguinte formulário irá se abrir:



Do lado esquerdo da aba “Especificações” estão as "Variáveis Disponíveis", que serão usadas tanto para o cálculo dos testes como para a definição de classe ou fator.

Para selecionar uma variável, clique no botão [>]. O botão [>>] seleciona todas as variáveis disponíveis.

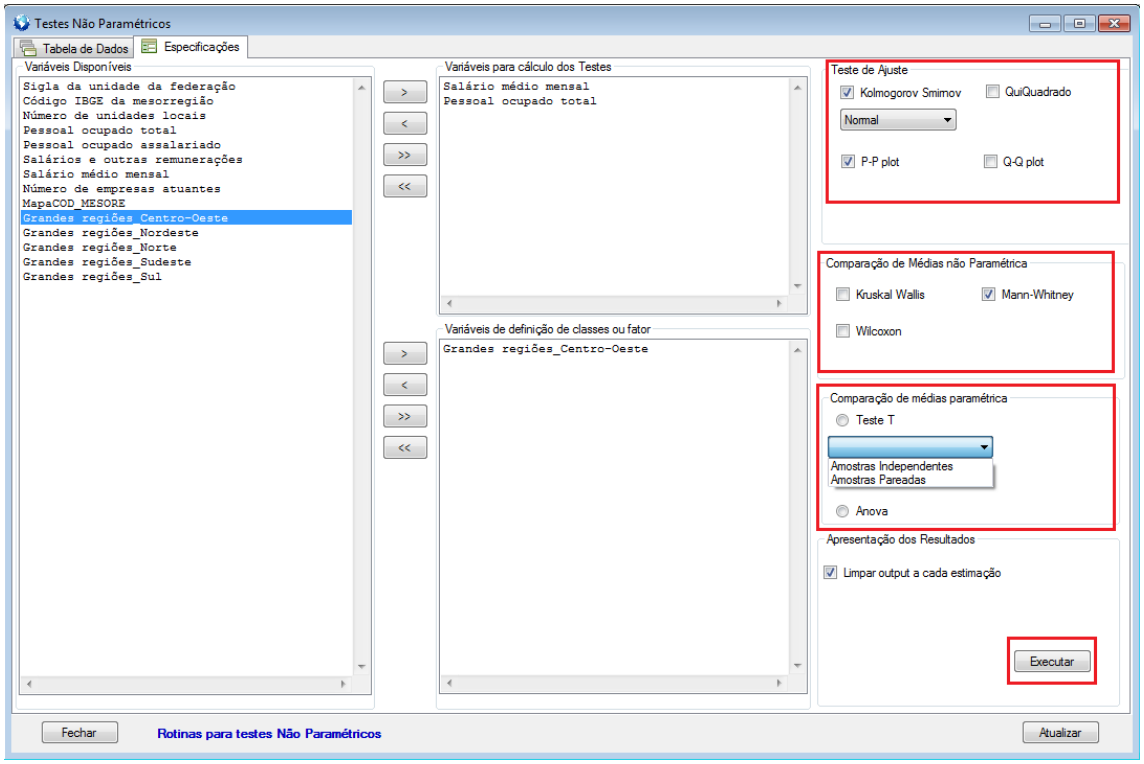


Do lado direito da janela, estão localizados as seguintes opções de "Testes de

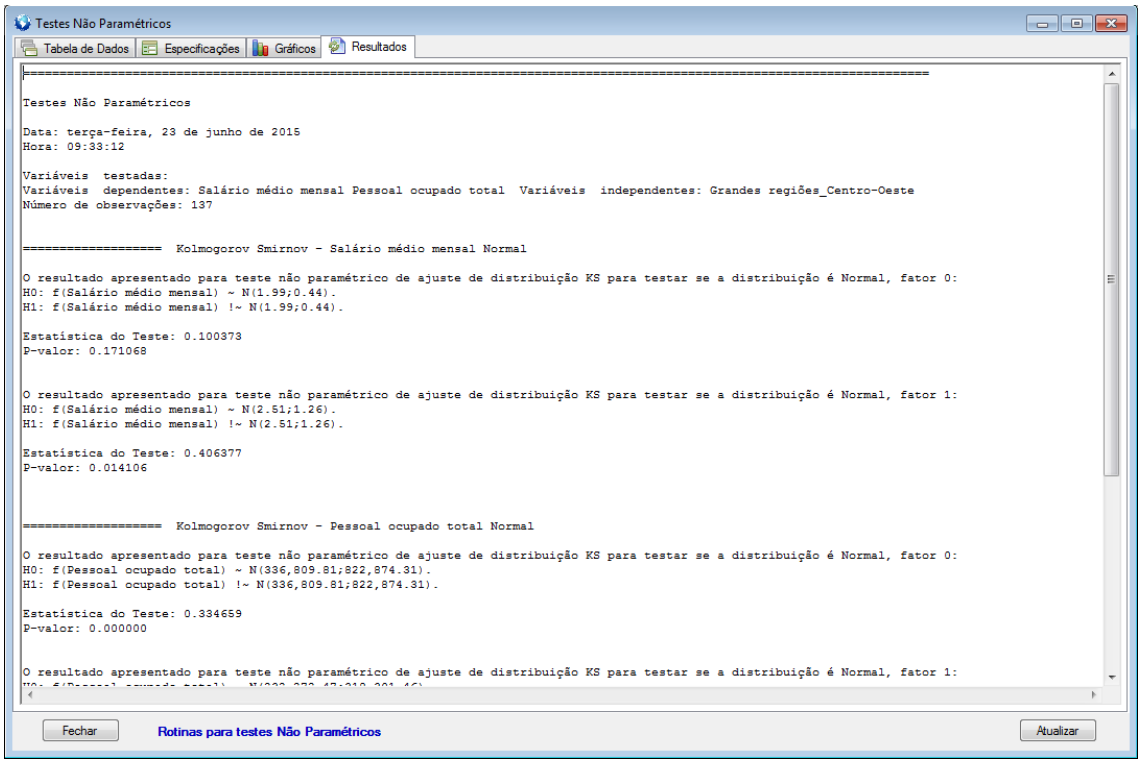
Ajuste": Kolmogorov-Smirnov, Qui-Quadrado, P-P plot e Q-Q plot

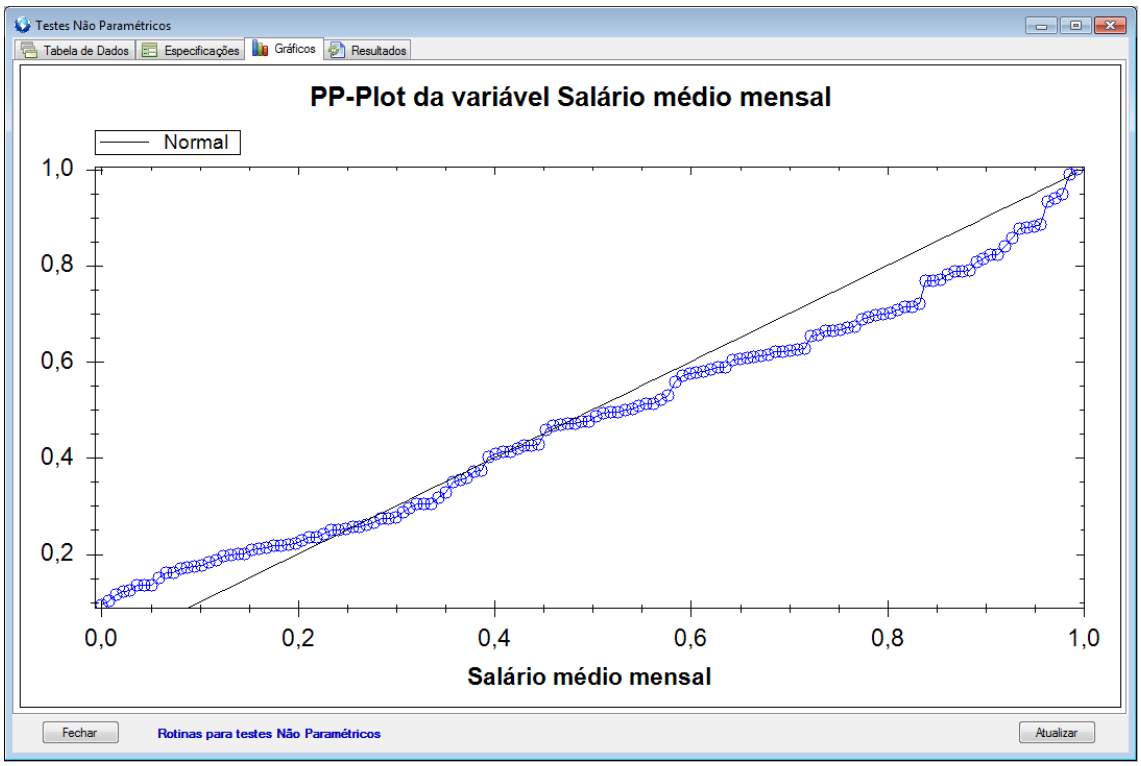
Existem as seguintes opções de testes não paramétricos para a comparação de médias: Kruskal Walls, Mann-Whitney e Wilcoxon.

Além destas, existem as seguintes opções para os testes paramétricos para comparação de médias: Teste T e análise de variância (ANOVA).



Após selecionar as variáveis desejadas e as opções de análise, clique no botão "Executar". De acordo com as escolhas do usuário duas novas abas chamadas “Resultados” e “Gráficos” irão aparecer contendo as análises realizadas.





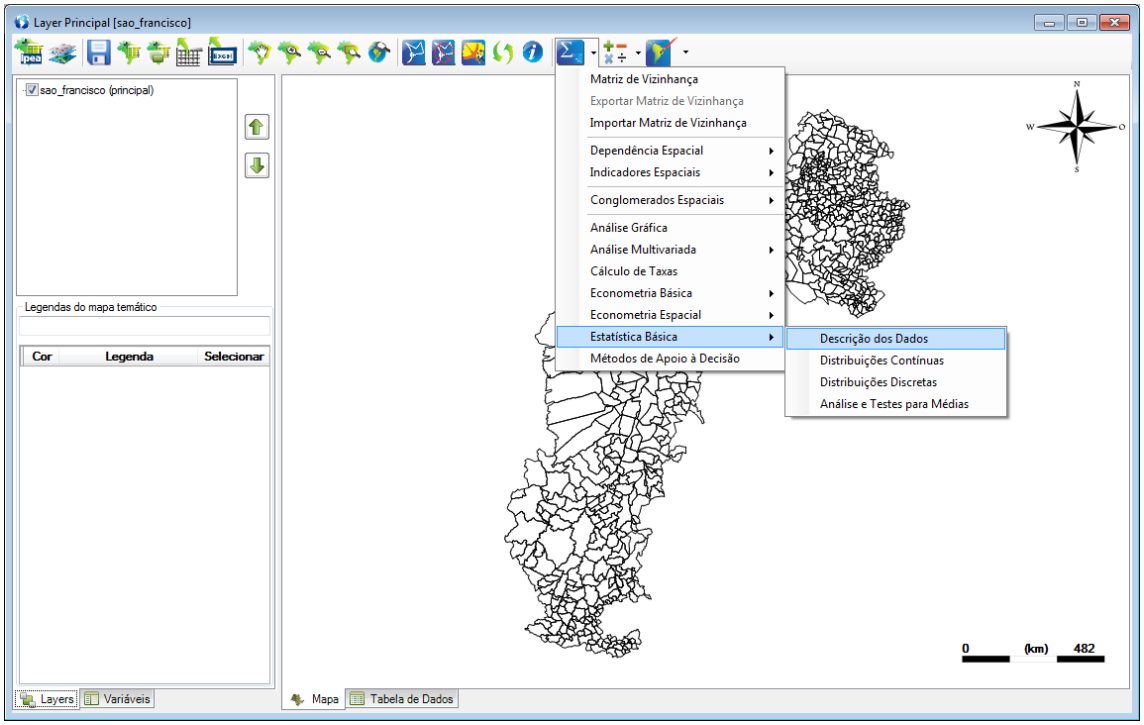
Descrição dos Dados

Visão Geral

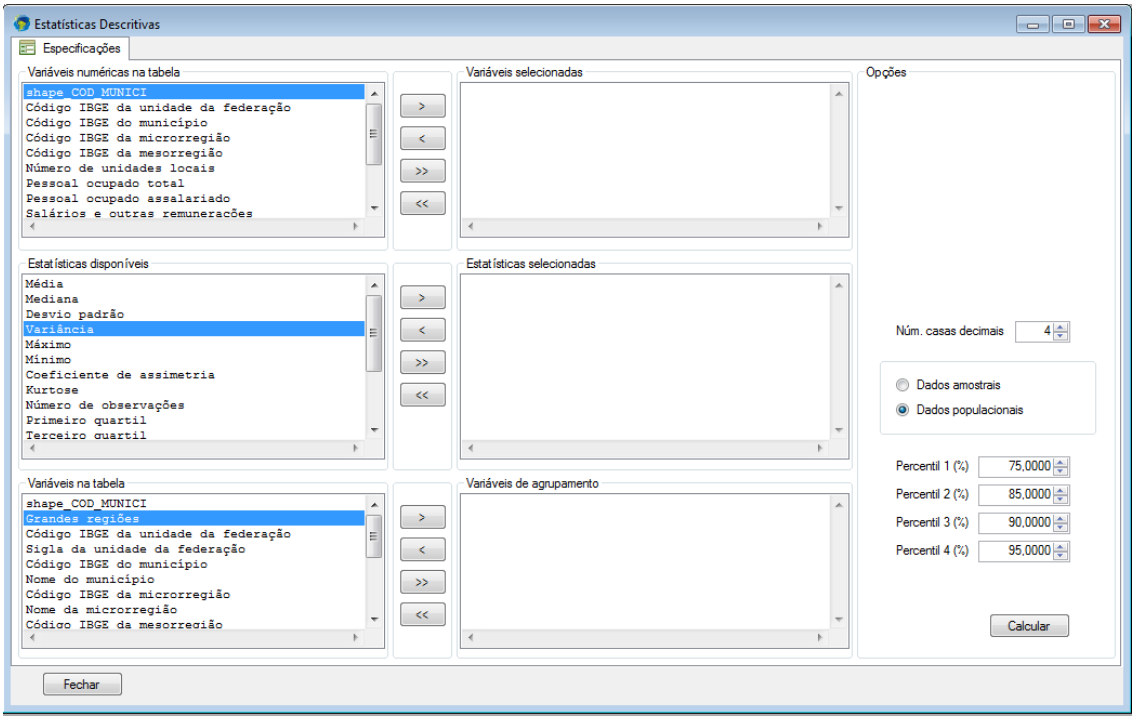
Esse formulário é idêntico ao formulário de Estatística Descritiva.

Exemplo

Após importar sua tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar seu arquivo de malha digital (*shape*) com sua tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , opção “Estatística Básica” e em seguida na opção “Descrição dos Dados”.



O seguinte formulário irá se abrir:

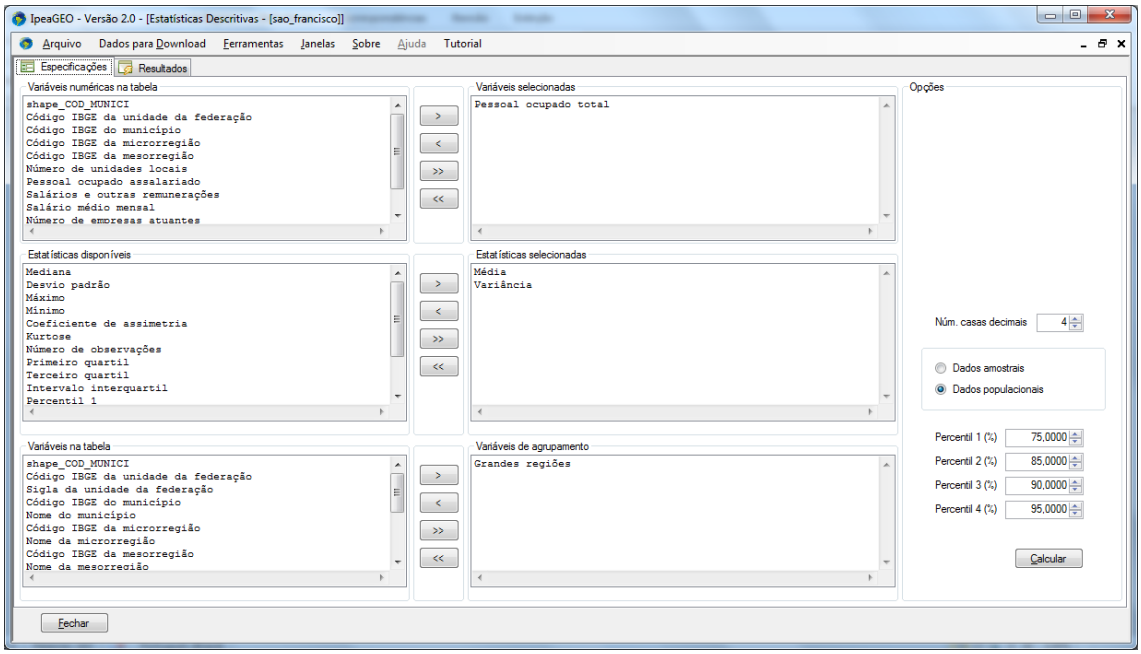


Escolha as variáveis desejadas em “Variáveis numéricas na tabela”, assim como uma estatística para o processo em “Estatísticas disponíveis”. O usuário também deve selecionar em “Variáveis na tabela” as variáveis que serão usadas em “Variáveis de agrupamento”.

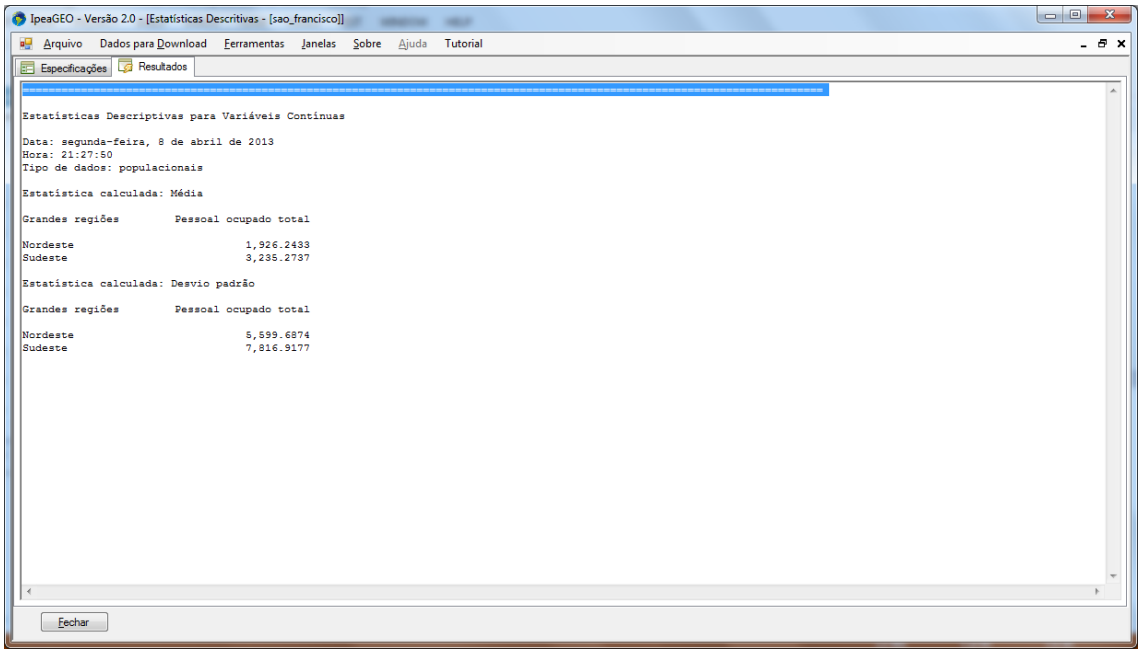
As seguintes opções estão disponíveis:

- Número de casas decimais.
- Classificação dos dados.
- Tamanho de cada percentil.

Depois de feitas as escolhas, clicar em “Calcular”.



Uma aba abrirá contendo os resultados referentes às análises escolhidas.



Caso deseje voltar à aba anterior para mais testes, basta clicar na aba “Especificações”.

Distribuições Contínuas

Visão Geral

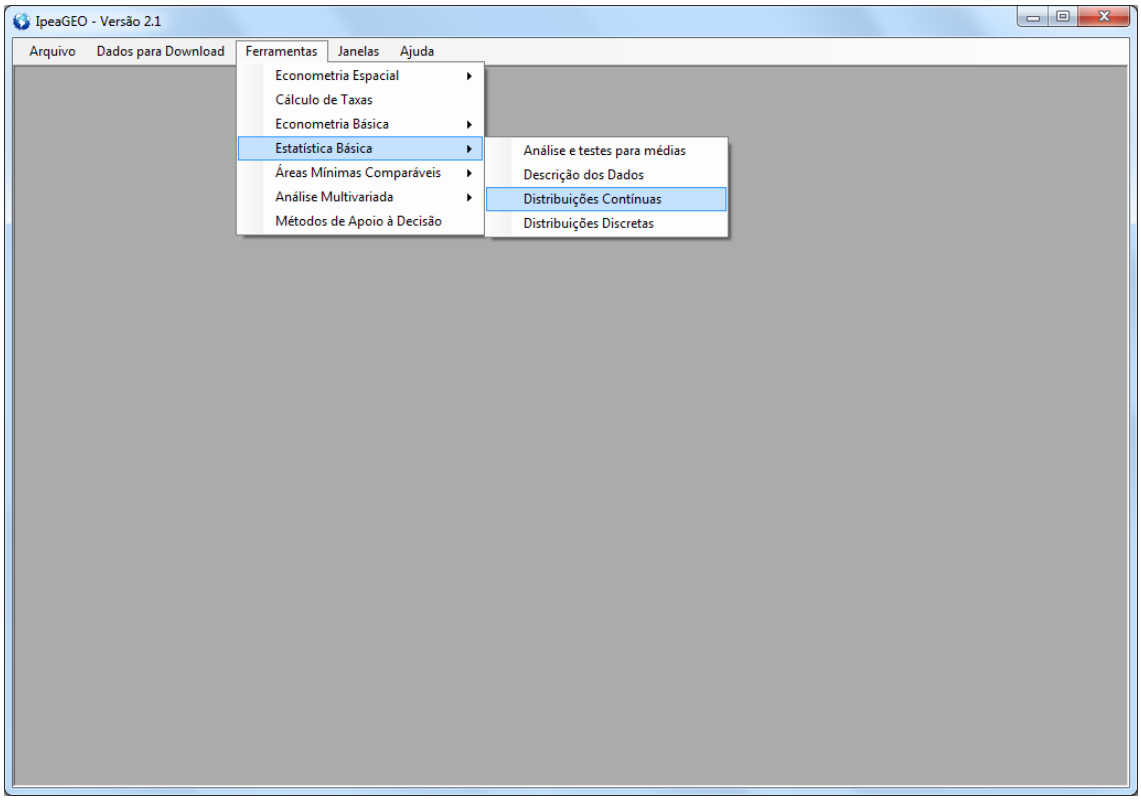
O IpeaGEO permite verificar o ajuste de variáveis aleatórias contínuas com as seguintes distribuições: Normal, Exponencial, Gama, Cauchy, F, Qui-quadrado, Beta, Qui (Chi), Weibull e Logística.

Exemplo

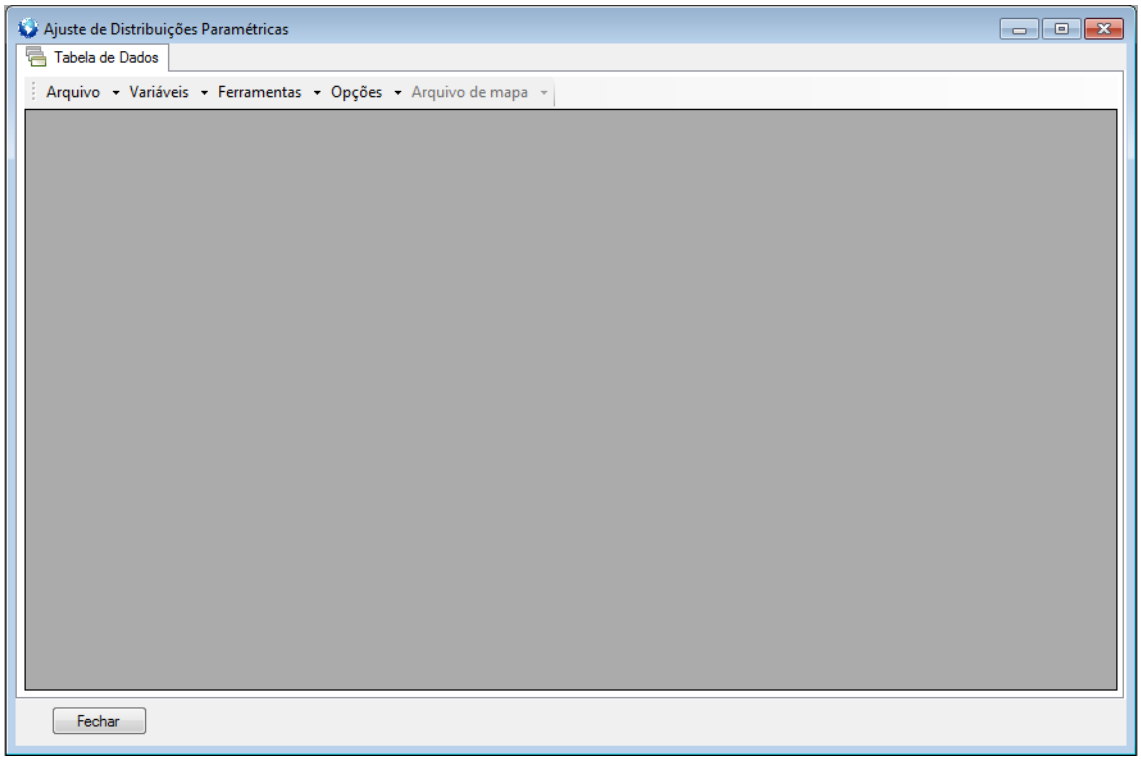
Existem dois modos para realizar essa análise no IpeaGEO:

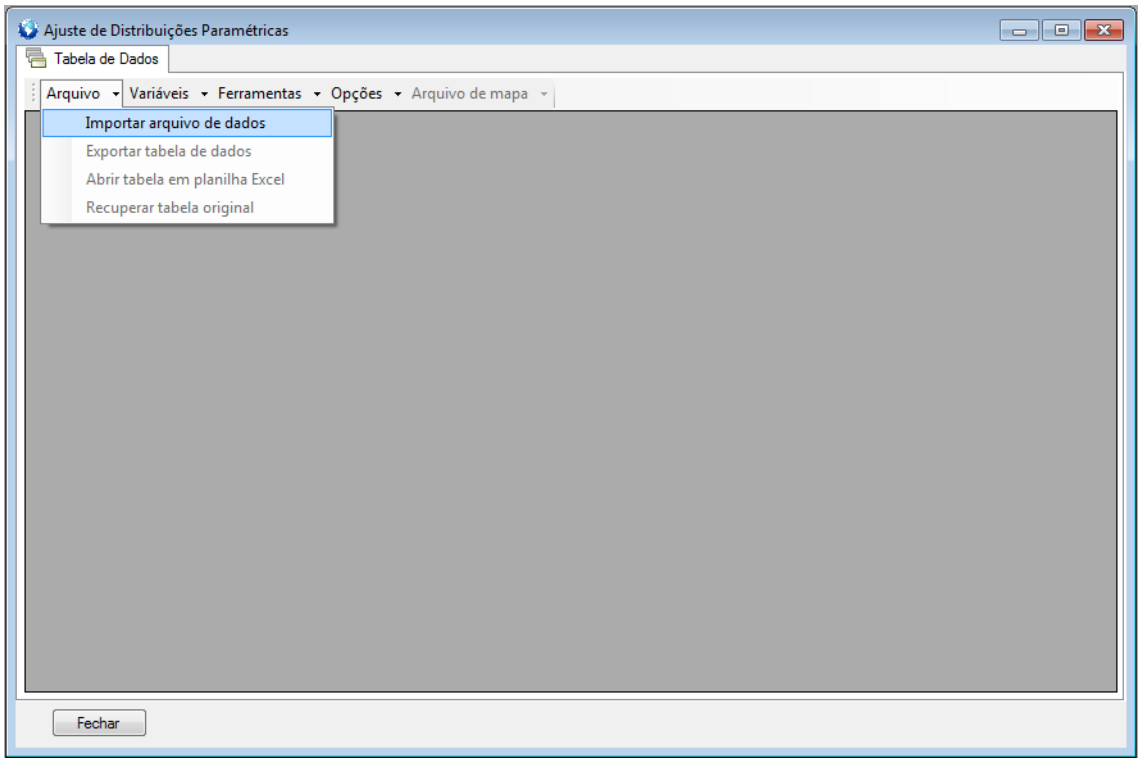
- Direto do menu principal:

No ícone "Ferramentas", clique na opção “Estatística Básica” e, em seguida, na opção “Distribuições Contínuas”.



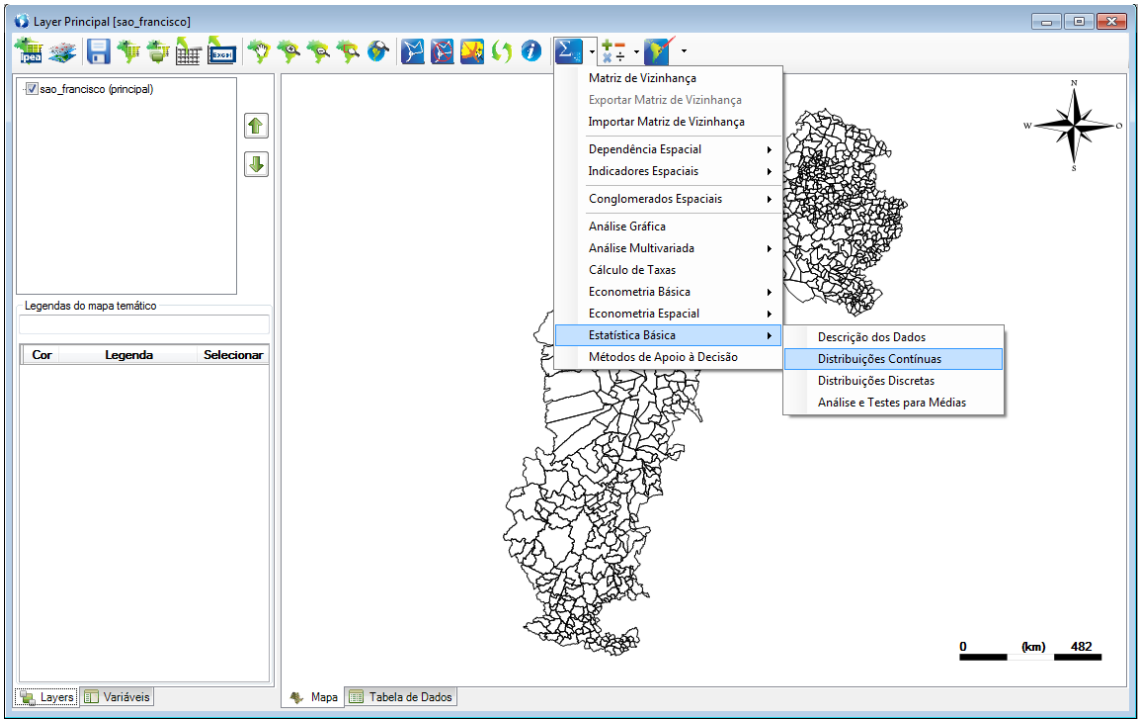
Ao clicar em “Distribuições Contínuas” se abrirá uma aba chamada “Tabela de Dados”, na qual será necessário importar um arquivo de dados. Para isto, clique em "Arquivo" e, em seguida, em "Importar arquivo de dados".



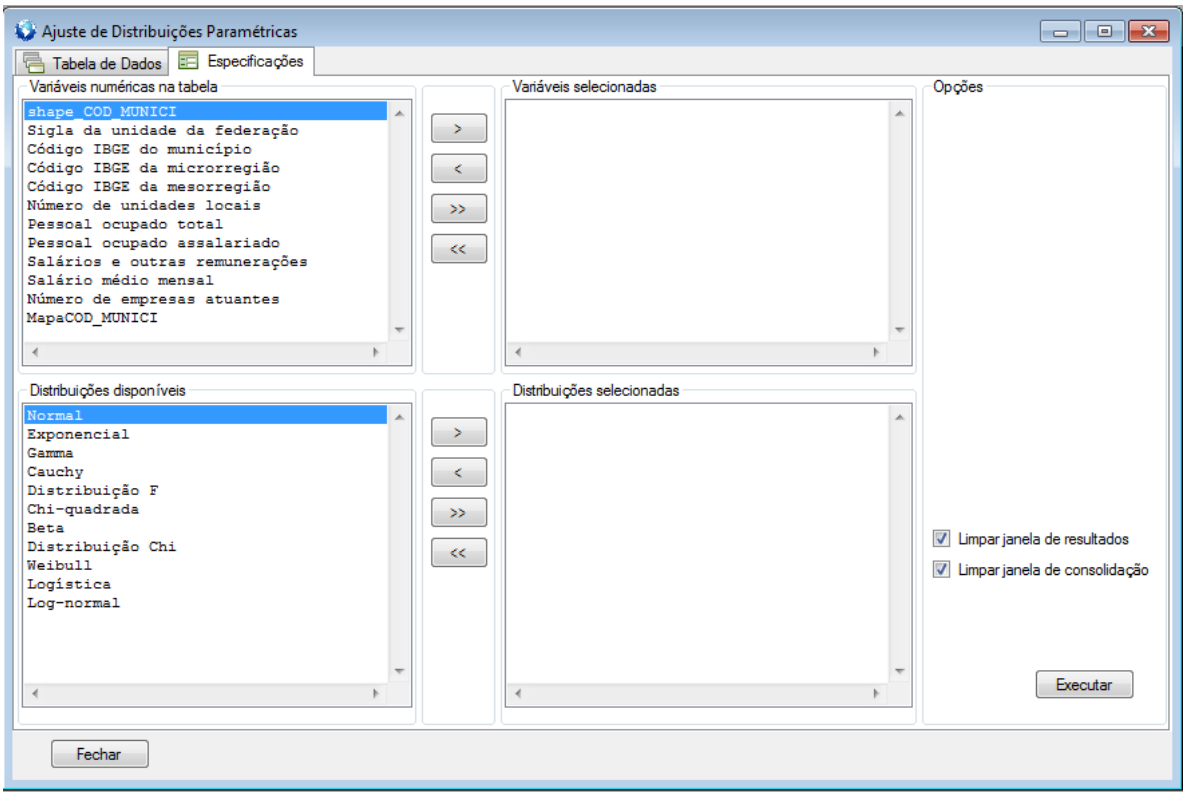


- <u>A partir da importação de uma malha digital (shape)</u>:

Clique no ícone “Ferramentas” , opção “Estatística Básica” e, em seguida, na opção “Distribuições Contínuas”.

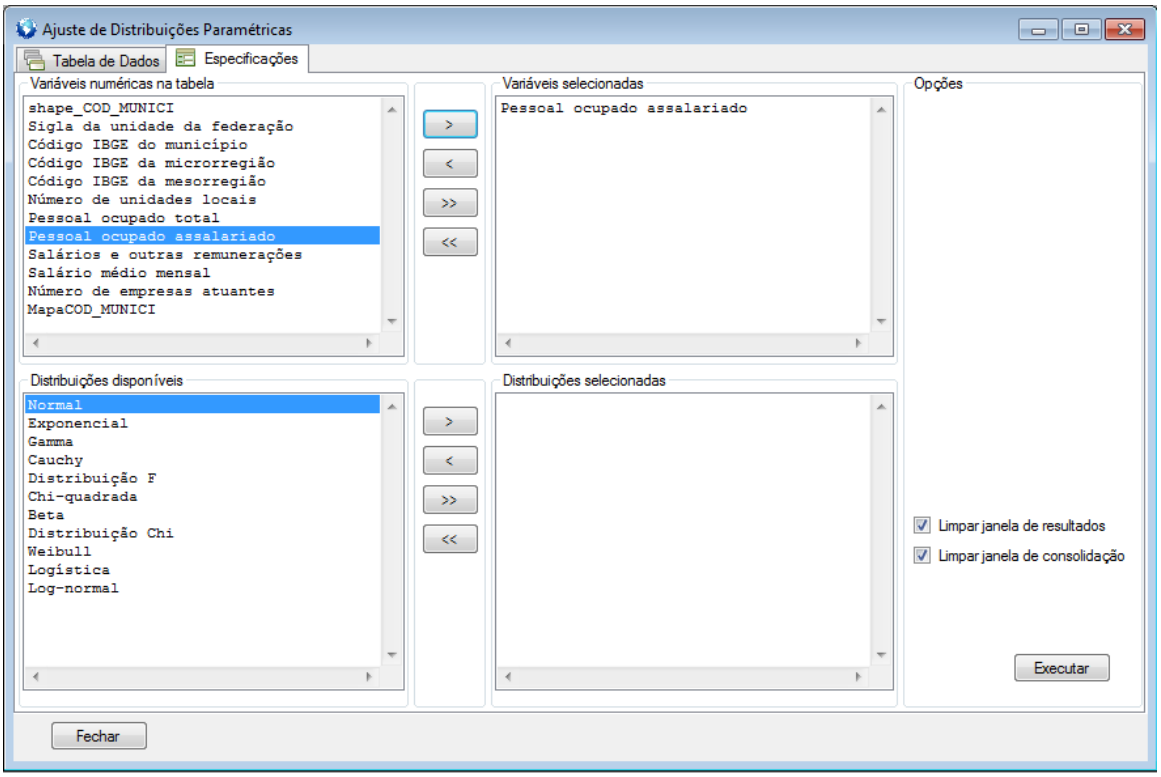


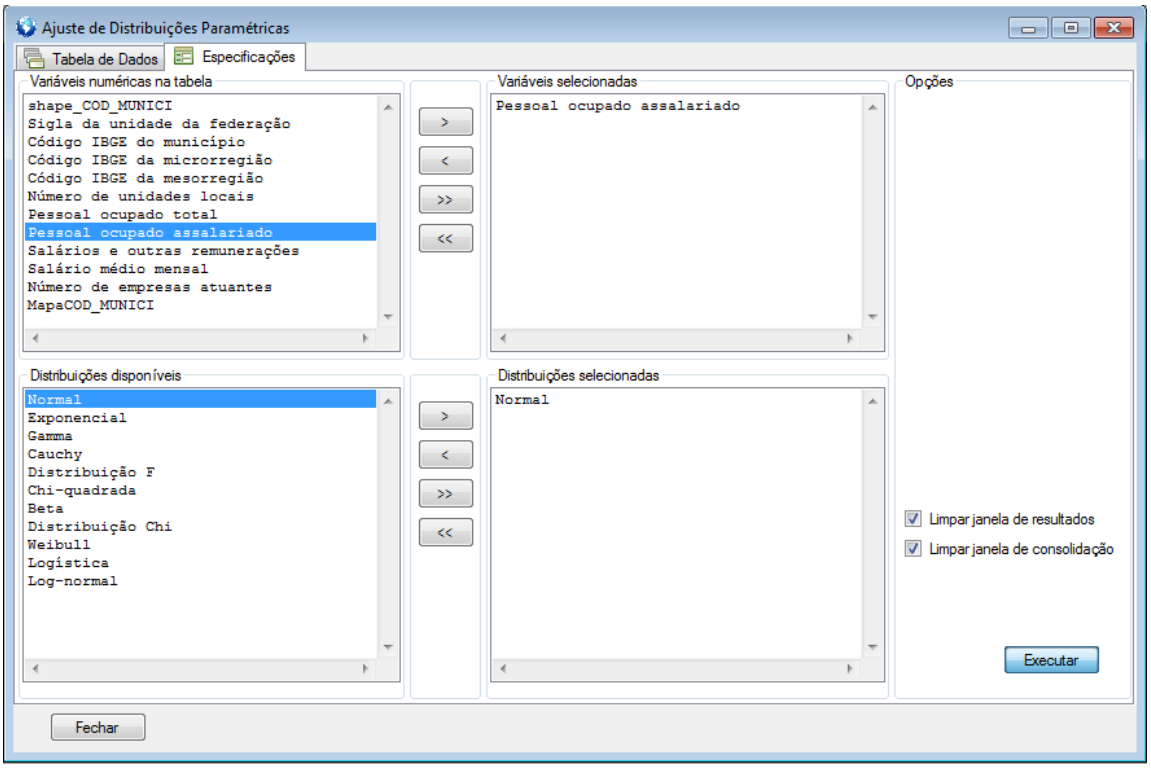
O seguinte formulário irá se abrir:



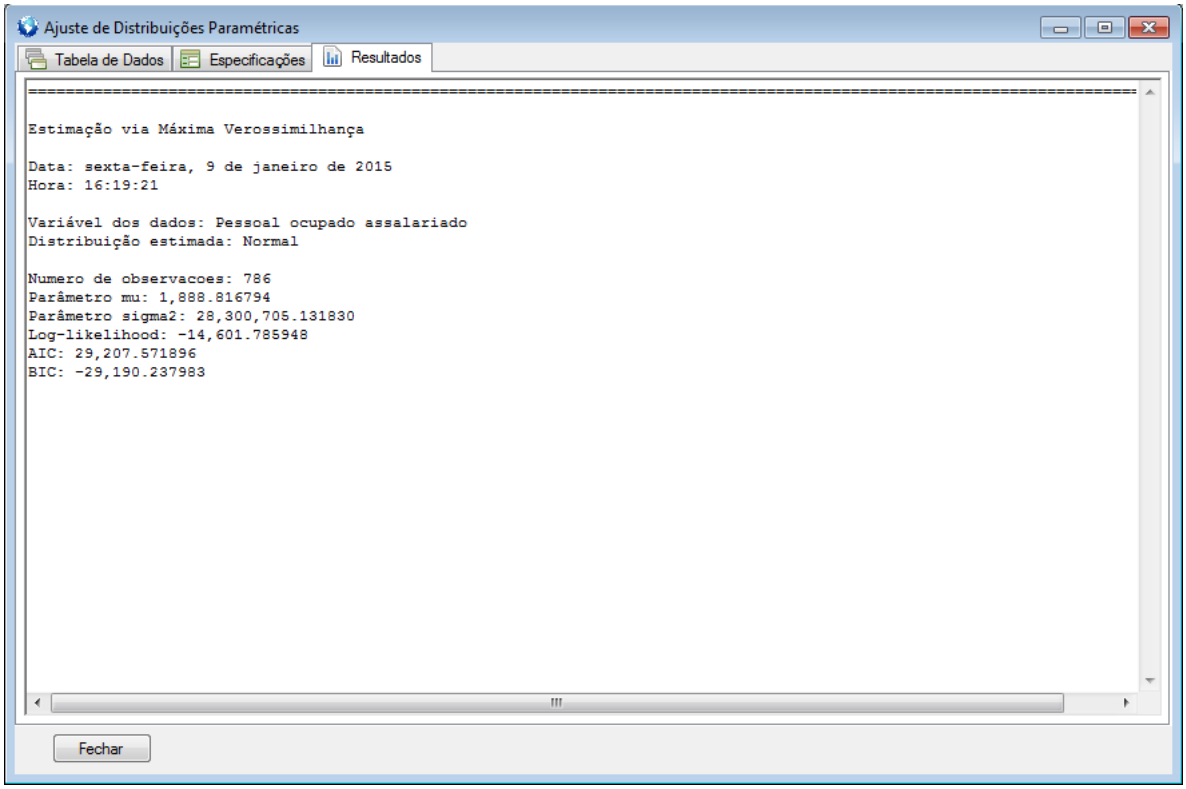
Do lado esquerdo da aba “Especificações” estão as "Variáveis numéricas na tabela" e as "Distribuições disponíveis" no IpeaGEO. As caixas no lado direito armazenam as variáveis numéricas selecionadas e as distribuições selecionadas, respectivamente.

Para selecionar uma variável ou distribuição, clique no botão [>]. O botão [>>] seleciona todas as variáveis ou distribuições disponíveis.





Após a escolha das variáveis numéricas e as distribuições disponíveis, clique no botão "Executar". Uma nova aba chamada “Resultados” aparecerá contendo os resultados das análises realizadas.



Distribuições Discretas

Visão Geral

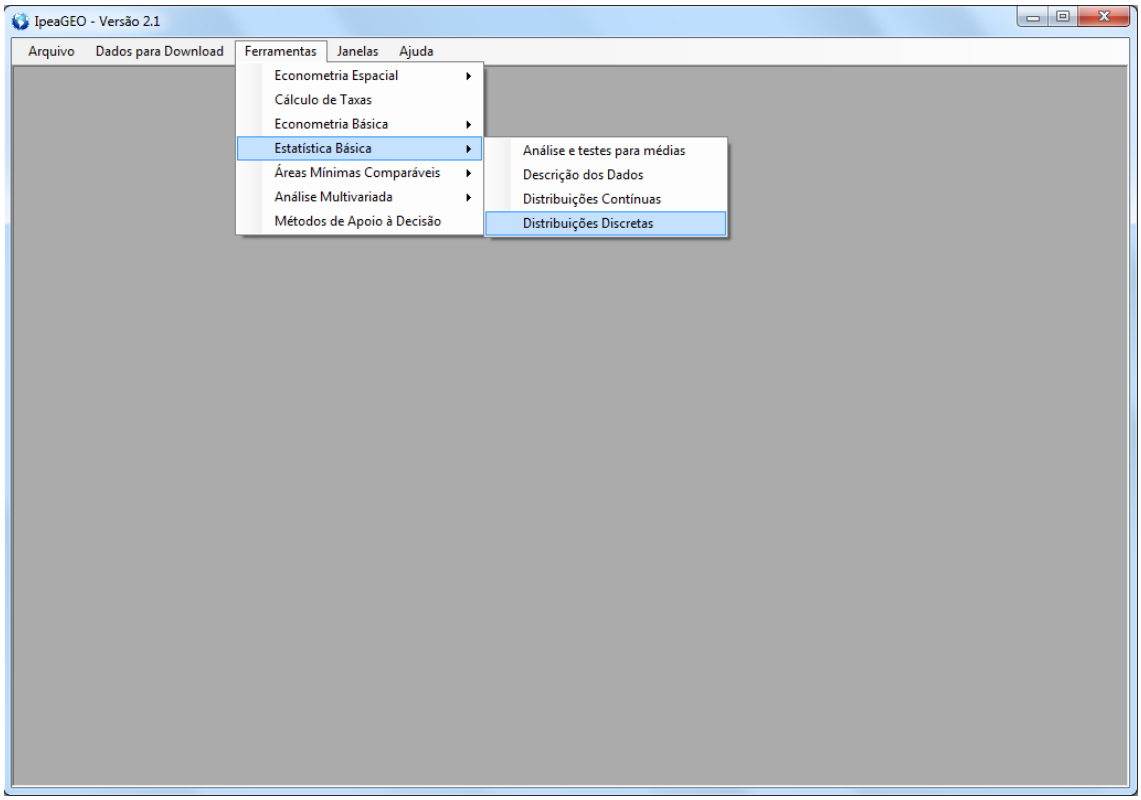
O IpeaGEO permite verificar o ajuste de variáveis aleatórias discretas com as seguintes distribuições: Bernoulli, Binomial, Geométrica, Pascal e Poisson.

Exemplo

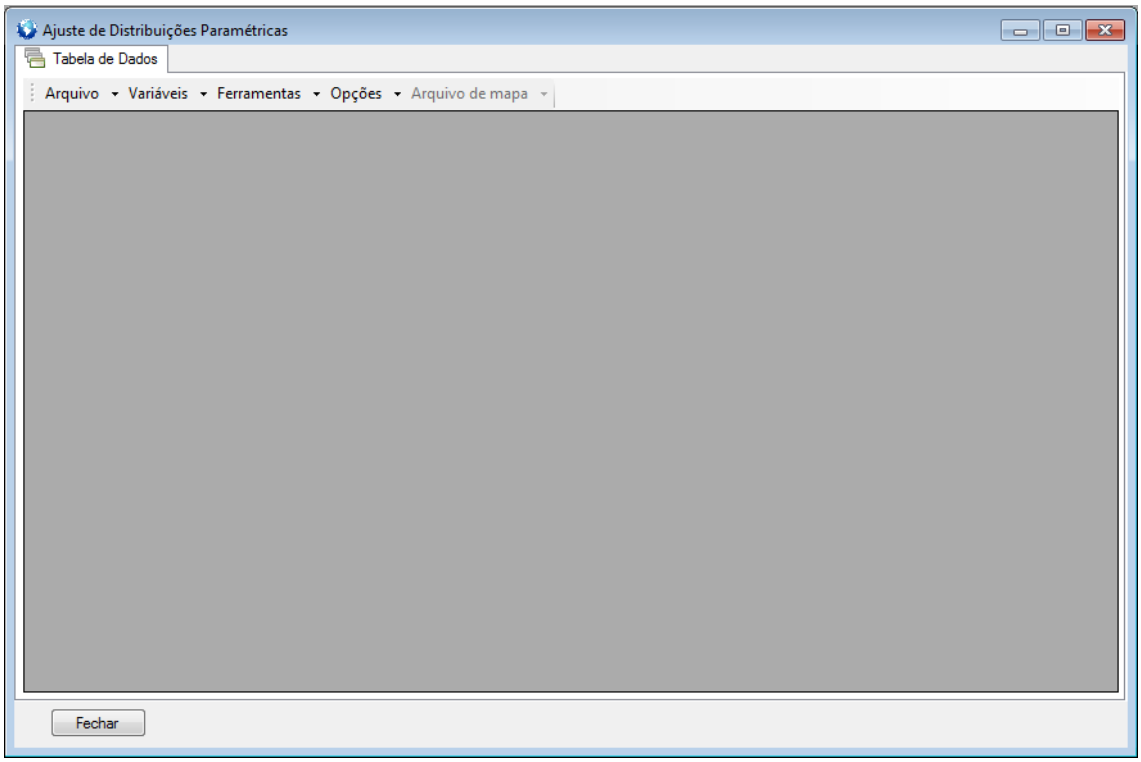
Assim como para as distribuições contínuas, existem dois modos para realizar essa análise no IpeaGEO:

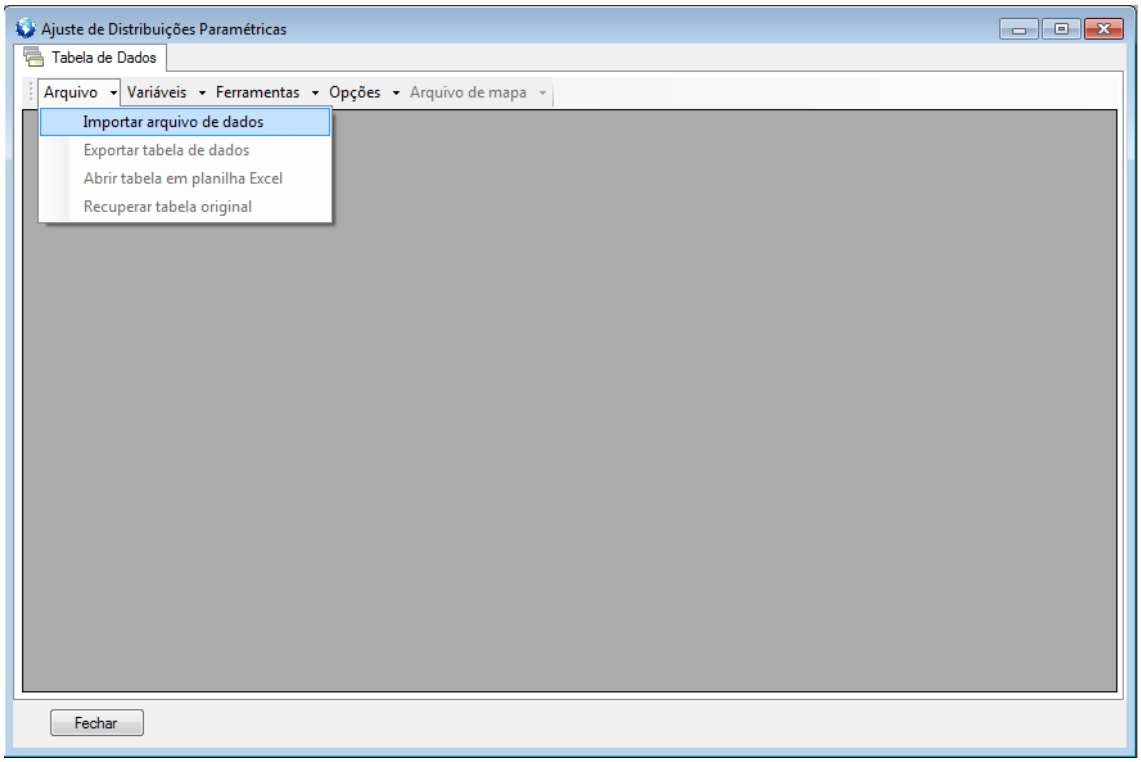
- Direto do menu principal:

No ícone "Ferramentas", clique na opção “Estatística Básica” e, em seguida, na opção “Distribuições Discretas”.



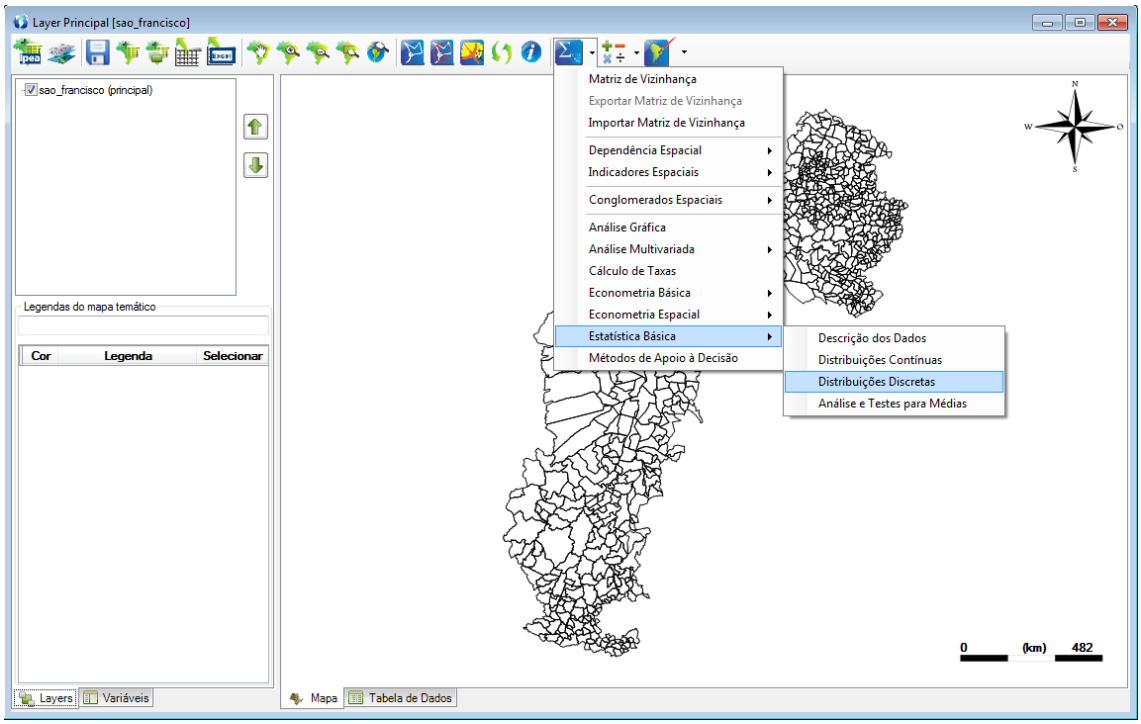
Ao clicar em “Distribuições Discretas” se abrirá uma aba chamada “Tabela de Dados”, na qual será necessário importar um arquivo de dados. Para isto, clique em "Arquivo" e, em seguida, em "Importar arquivo de dados".



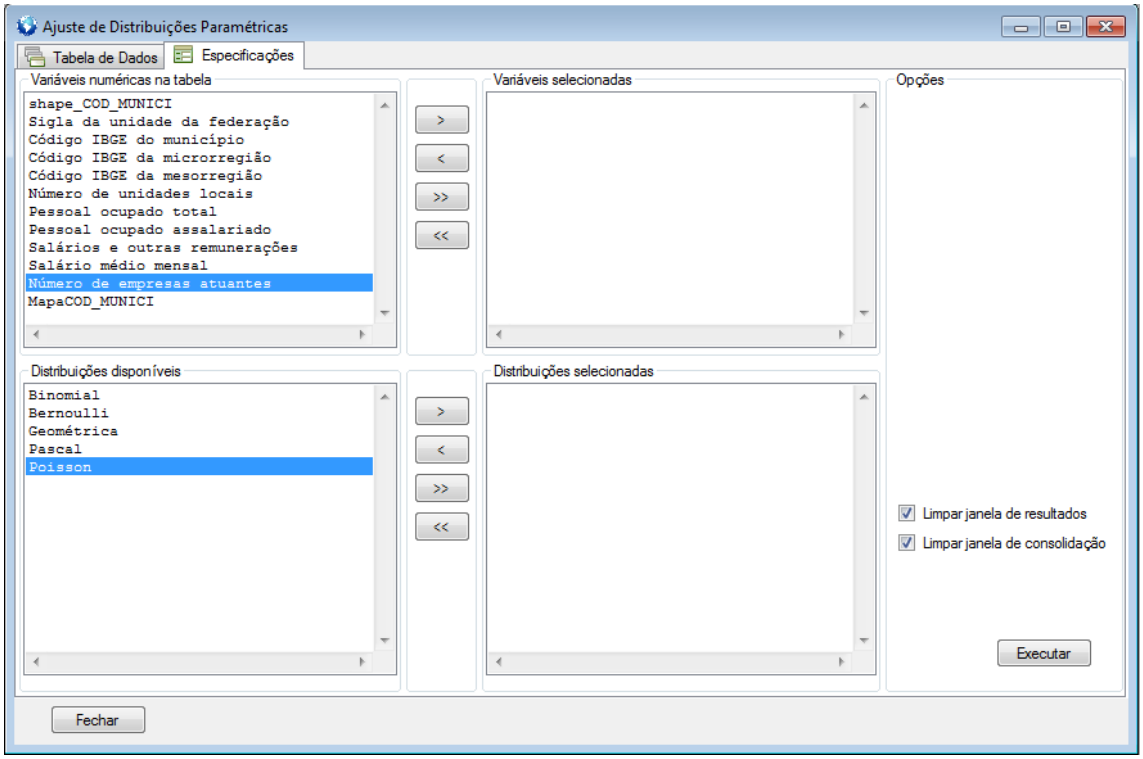


- A partir da importação de uma malha digital (shape):

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , opção “Estatística Básica” e, em seguida, na opção “Distribuições Discretas”.

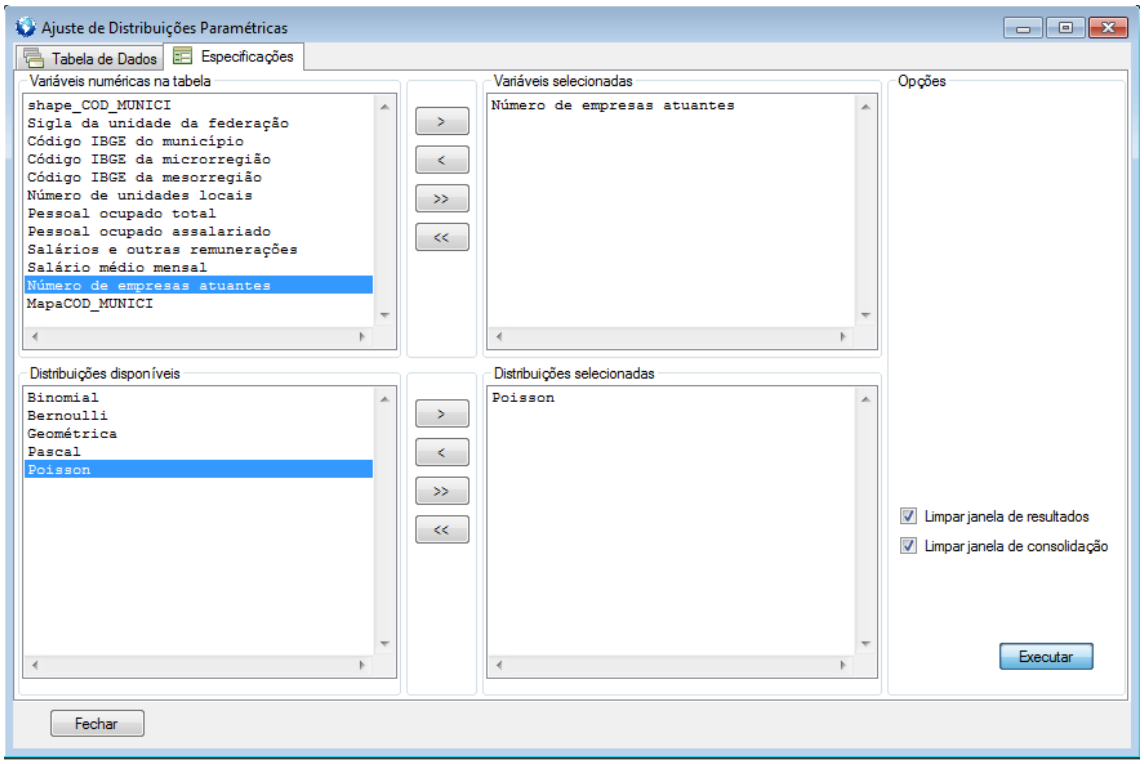


O seguinte formulário irá se abrir:



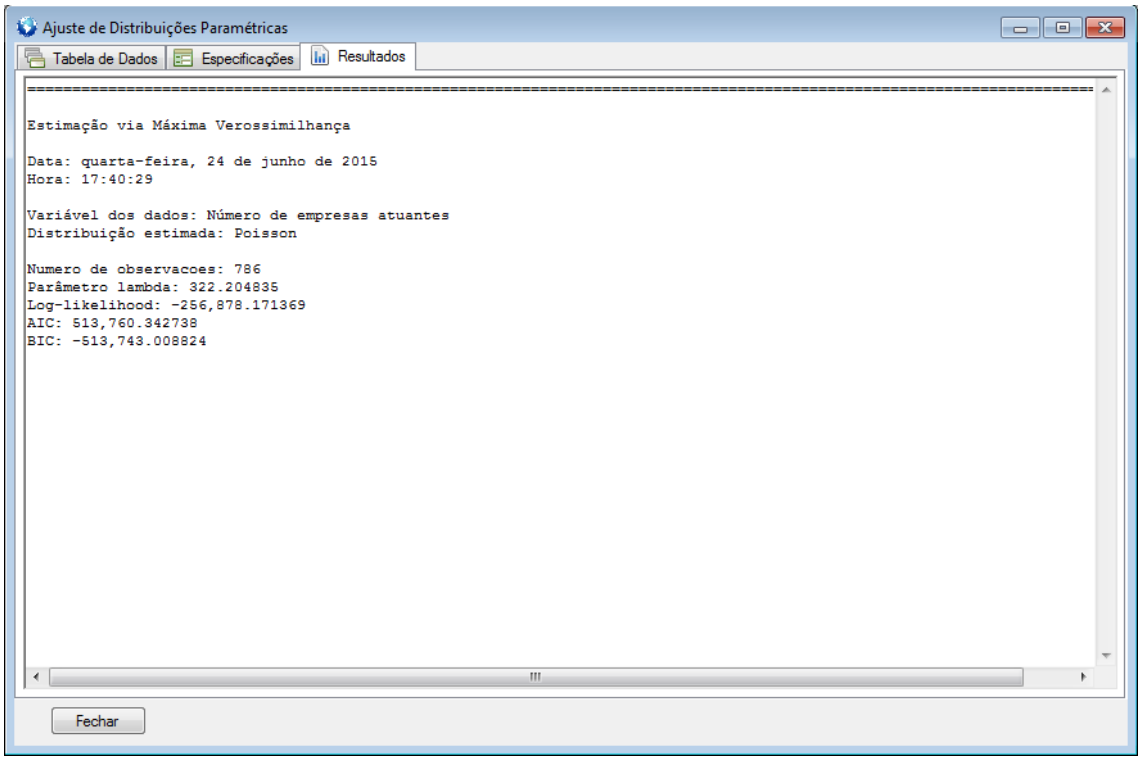
Na nova janela, escolha as variáveis desejadas em “Variáveis numéricas na tabela”, bem como uma distribuição de probabilidade para se testar em “Distribuições disponíveis”. Caso mais de uma distribuição seja selecionada, processos separados serão feitos para cada distribuição selecionada. Se o tipo de variável não for compatível com a distribuição selecionada, uma mensagem de erro aparecerá. Após as seleções

das variáveis e das distribuições, clique em “Executar”.



É permitido ainda limpar a janela de resultados e limpar janela de consolidação. Caso contrário, resultados gerados anteriormente continuarão na janela.

Após clicar em “Executar” se abrirá uma aba contendo os resultados dos testes.



É possível retornar à aba de especificações para testar outras distribuições. O resultado será exibido abaixo do resultado anterior na aba de resultados. É possível escolher por manter todos os resultados ou selecionar e apagar os que não mais estiverem sendo usados.

Métodos Multicritérios de Apoio as Decisões

Visão Geral

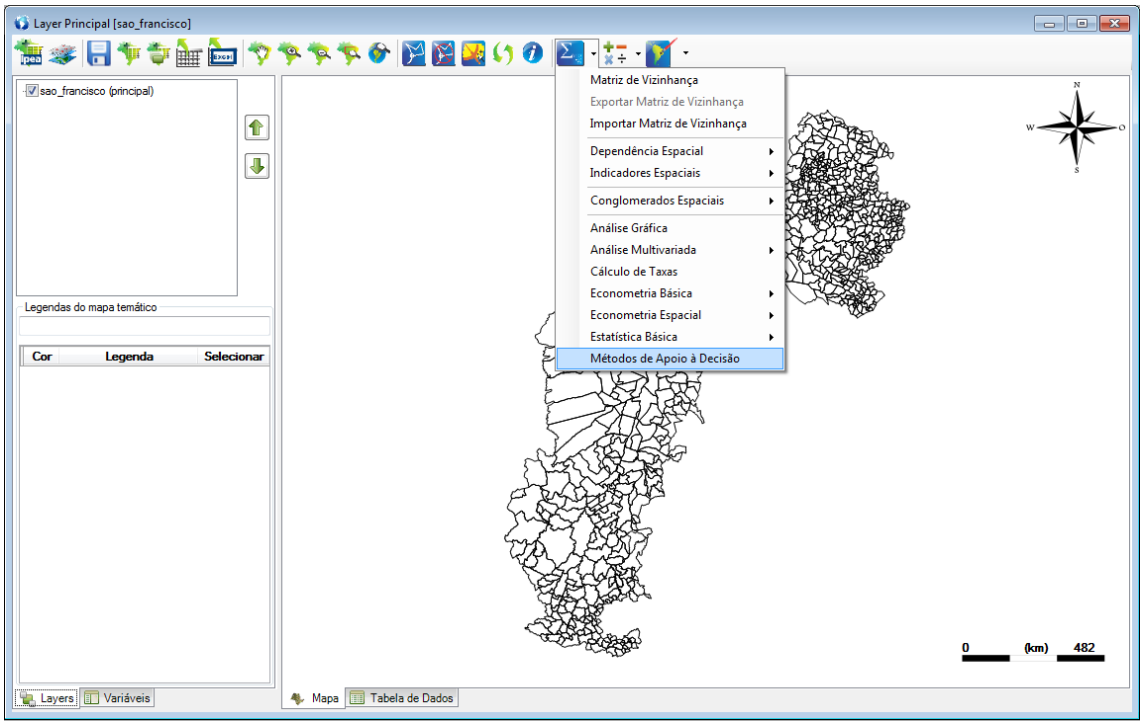
Nos problemas de decisão, o analista deve escolher a melhor alternativa dentre várias decisões possíveis, considerando, para tanto, como tais alternativas se comportam de acordo com alguns critérios. Suponha, por exemplo, que dentre múltiplas possibilidades, deseja-se escolher o melhor automóvel a ser comprado. Para que essa decisão tenha um bom embasamento deve-se, antes de tudo, definir os critérios que levarão ao modelo mais indicado. O melhor automóvel, por exemplo, pode ser aquele que tenha o menor preço, o menor consumo de combustível e a melhor aparência. Muitas vezes, entretanto, não é possível encontrar uma alternativa ótima. Isso significa que, nem sempre, há uma alternativa que cumpra, simultaneamente, todos os critérios definidos na análise.

As técnicas de apoio multicritério à decisão surgiram no inicio da década de 1970 com o objetivo de sistematizar processos de escolha, no caso em que o decisor pretende alcançar múltiplos objetivos. Essa metodologia não apresenta uma solução definitiva, mas auxilia o processo de decisão recomendando cursos de ação.

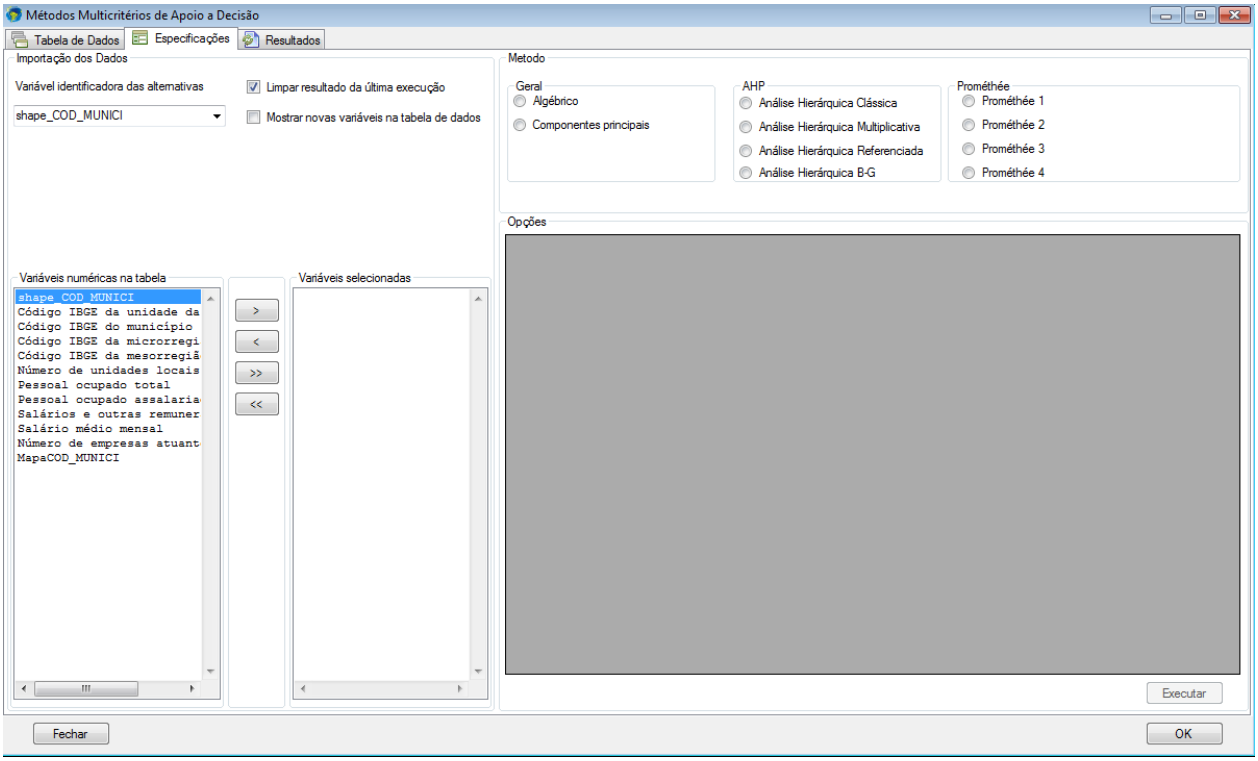
O IpeaGEO oferece ao usuário as seguintes opções de técnicas para analise de decisão multicritério: métodos gerais – algébrico e componentes principais; família AHP – análise hierárquica clássica, multiplicativa, referenciada e B-G; família Prometheé – Prométhée 1, 2, 3 e 4.

Exemplo

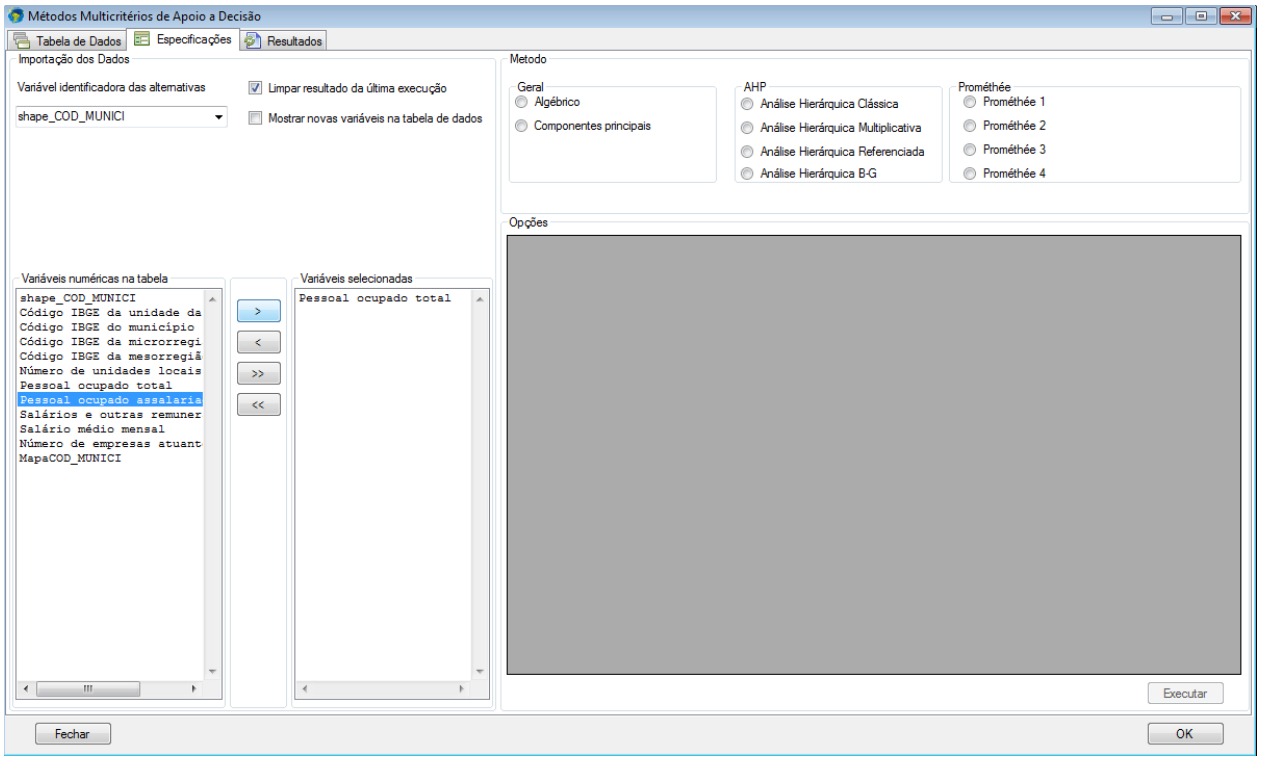
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo *shape* com a tabela de dados (ver tópico de concatenação), clique na opção “Métodos de Apoio à Decisão” localizada na caixa “Ferramentas” .



Do lado esquerdo no canto inferior da aba “Especificações” estão localizadas as variáveis numéricas disponíveis.



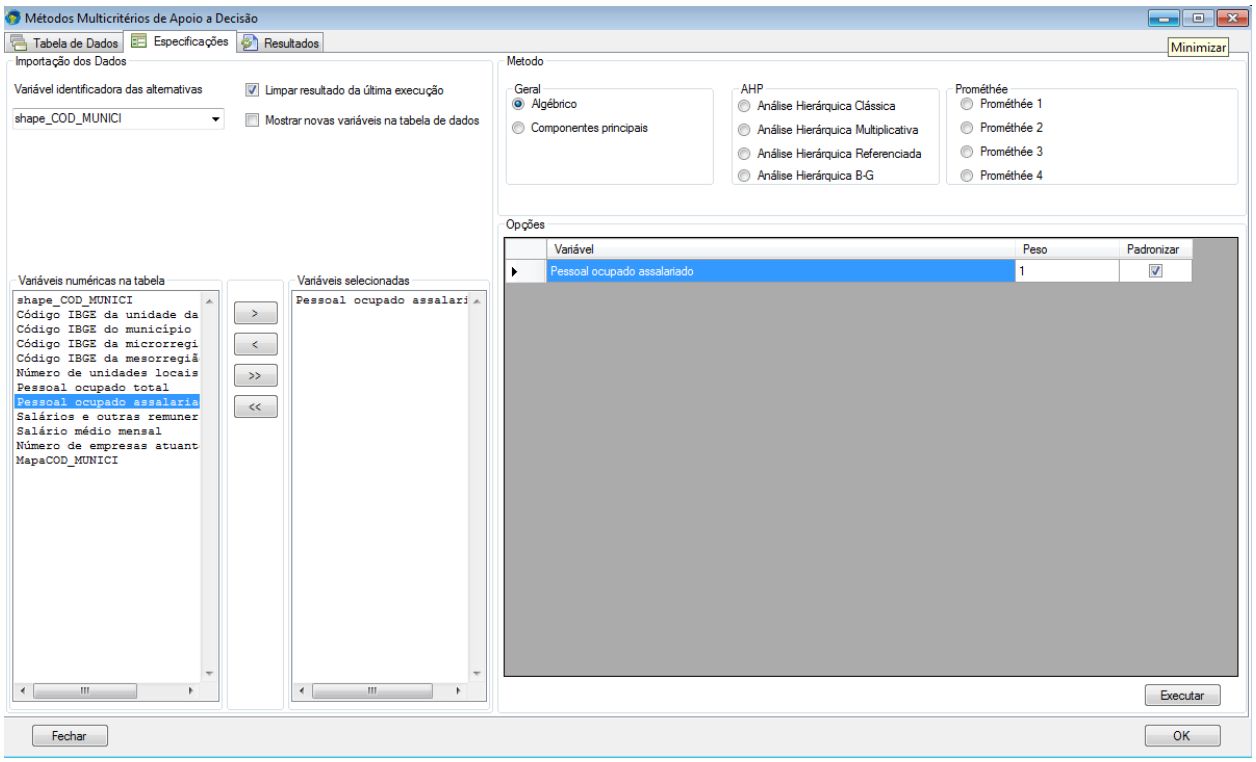
O botão > permite selecionar uma única variável, enquanto que o botão >> permite selecionar todas as variáveis disponíveis.

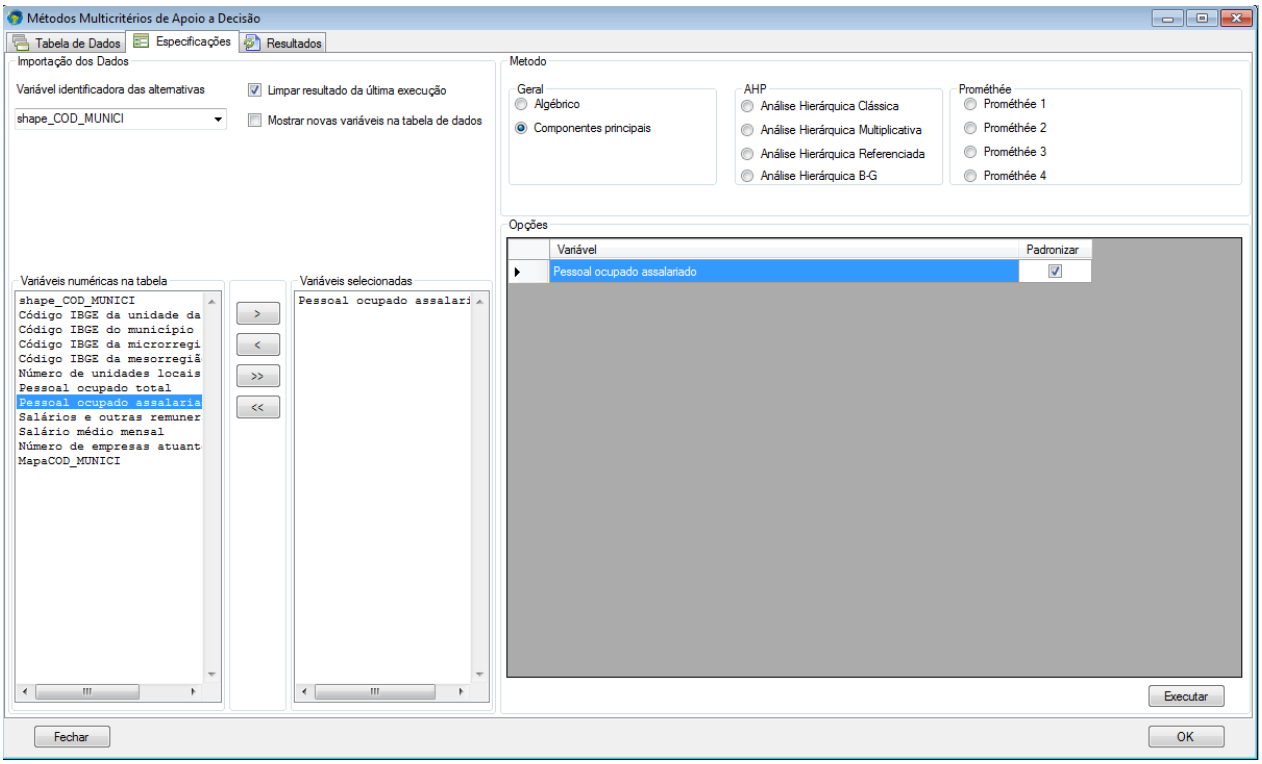


Os métodos estão divididos em três famílias: Geral, AHP, Prométhée. Ao selecionar um dos métodos, surgirão algumas opções de configuração da variável. As opções disponíveis para cada método são:

Família Geral:

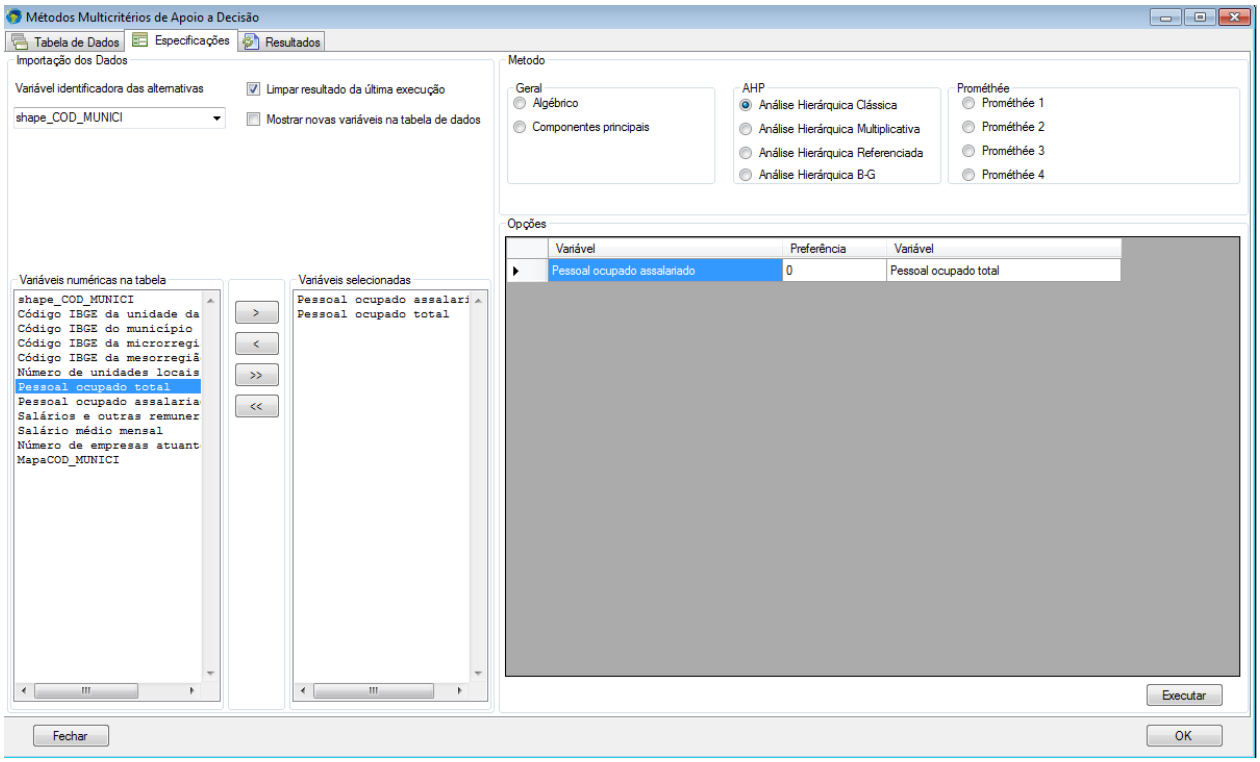
- Algébrico;
- Componentes principais.

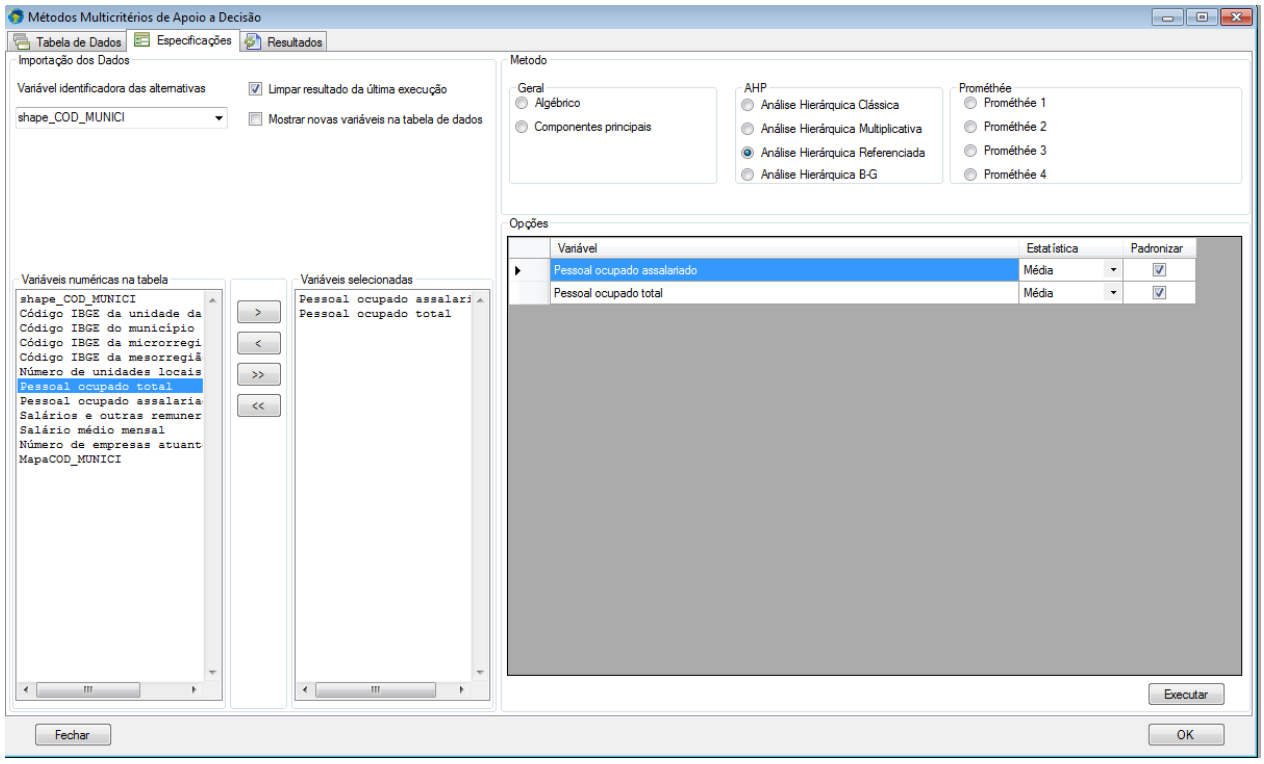




Família AHP:

- Análise Hierárquica Clássica;
- Análise Hierárquica Multiplicativa;
- Análise Hierárquica Referenciada;
- Análise Hierárquica B-G.

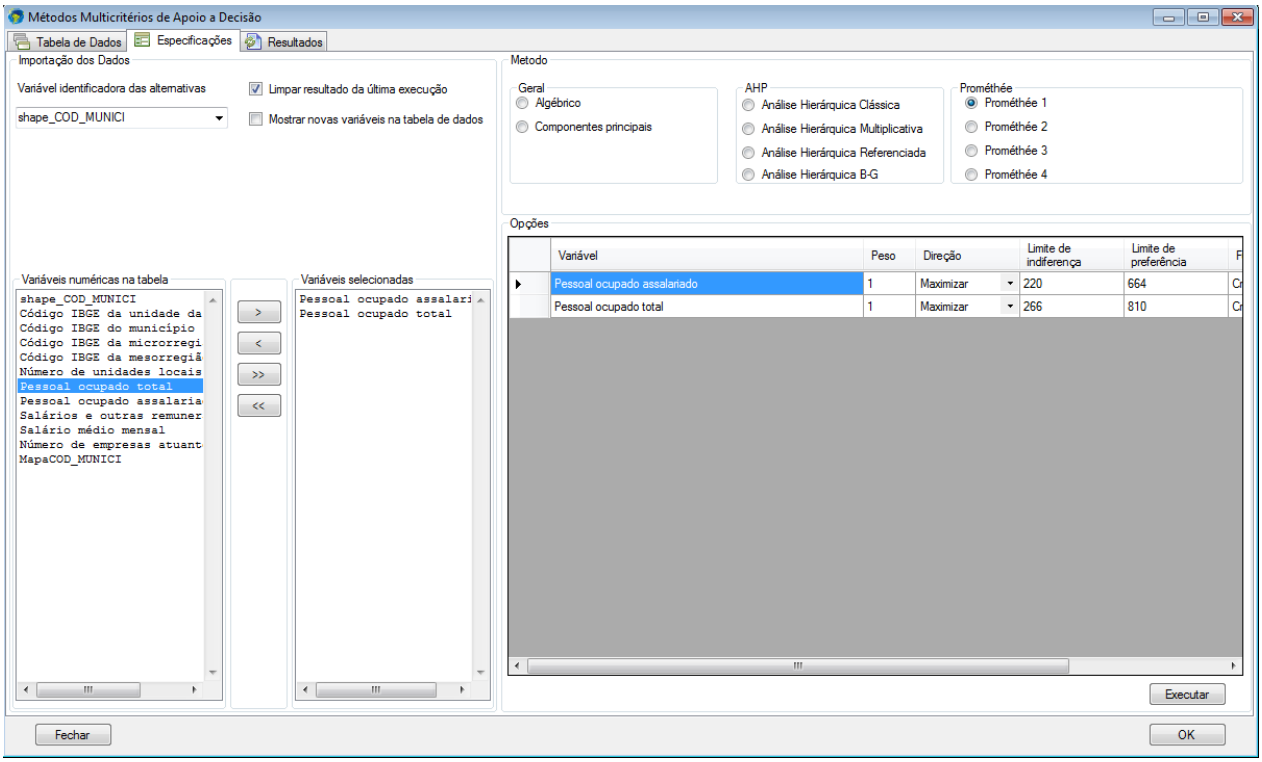


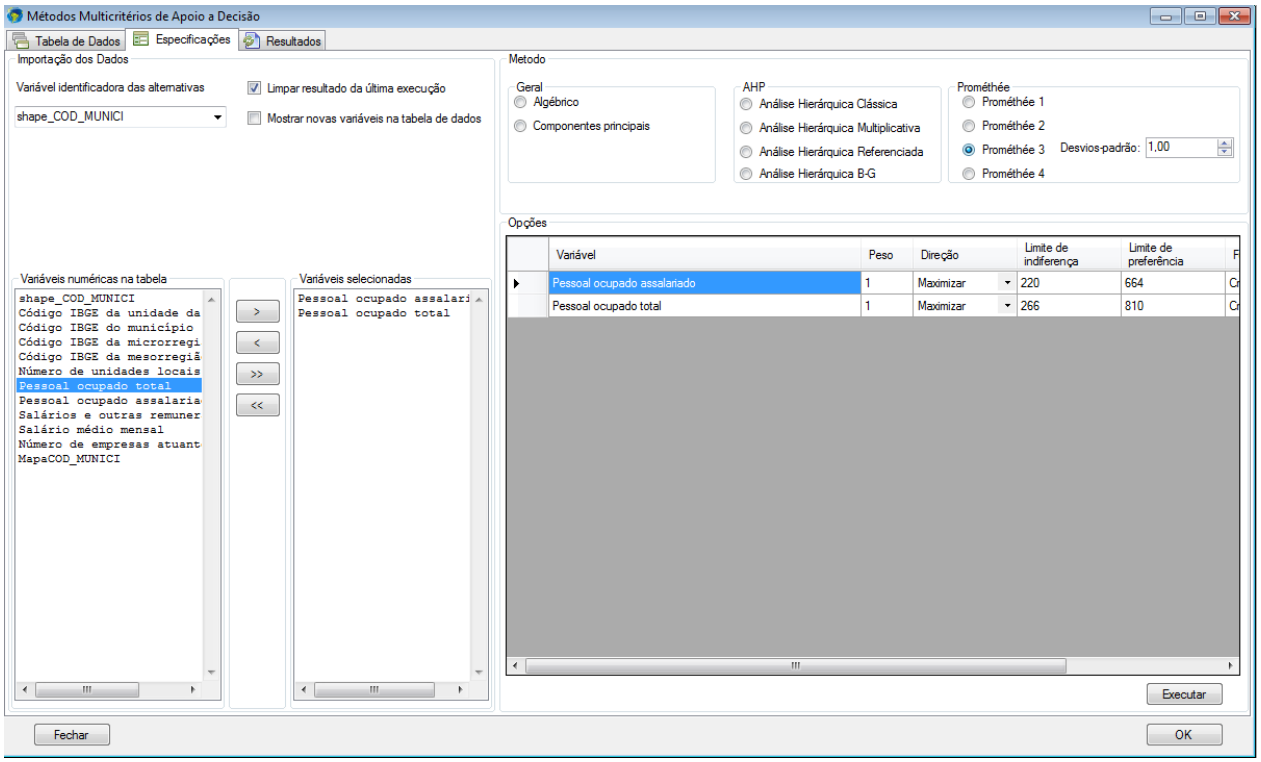


Família Prométhée:

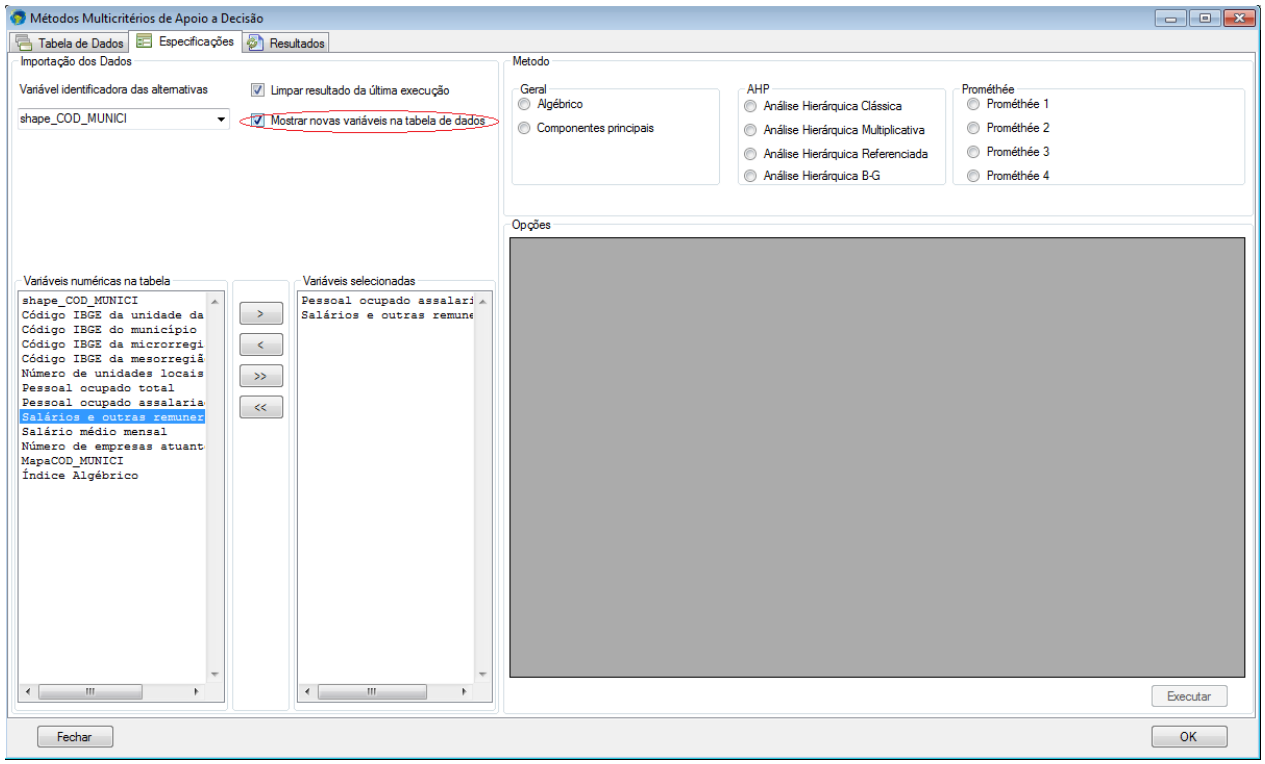
- Prométhe 1;
- Prométhée 2;
- Prométhée 3;

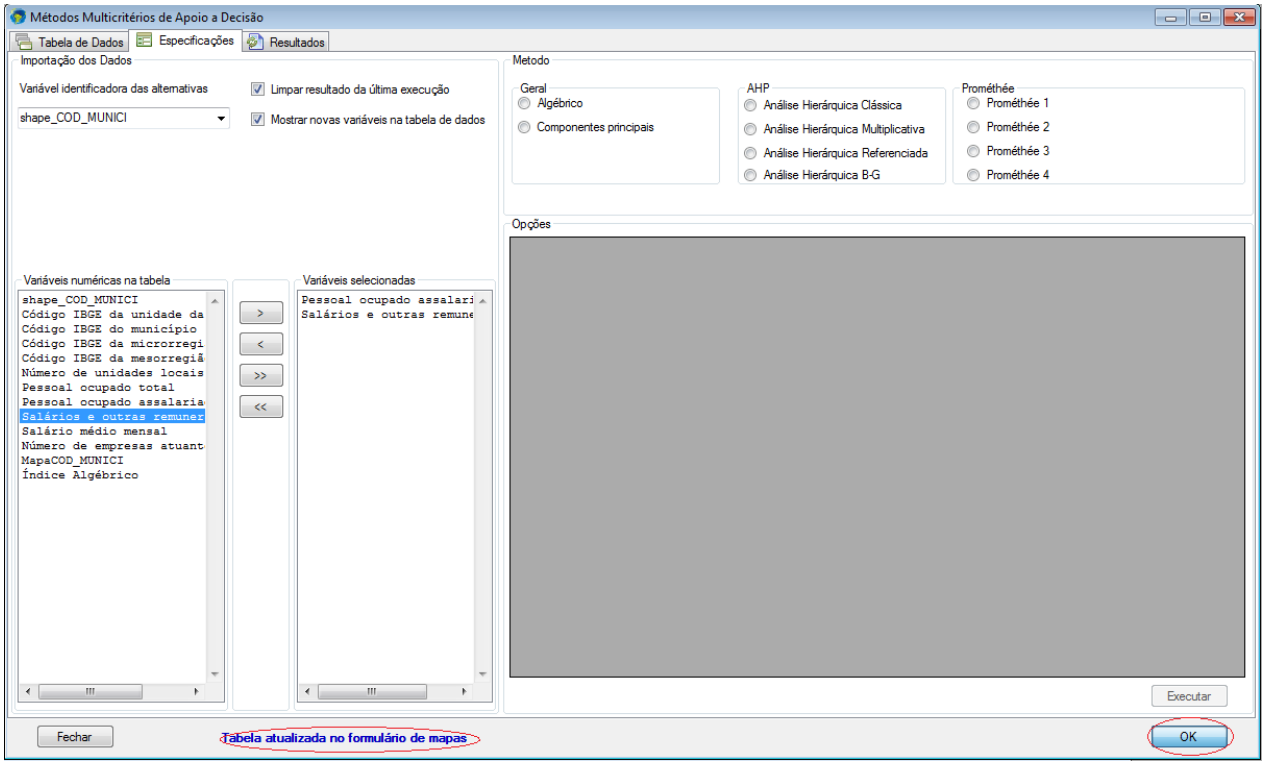
- Prométhée 4.



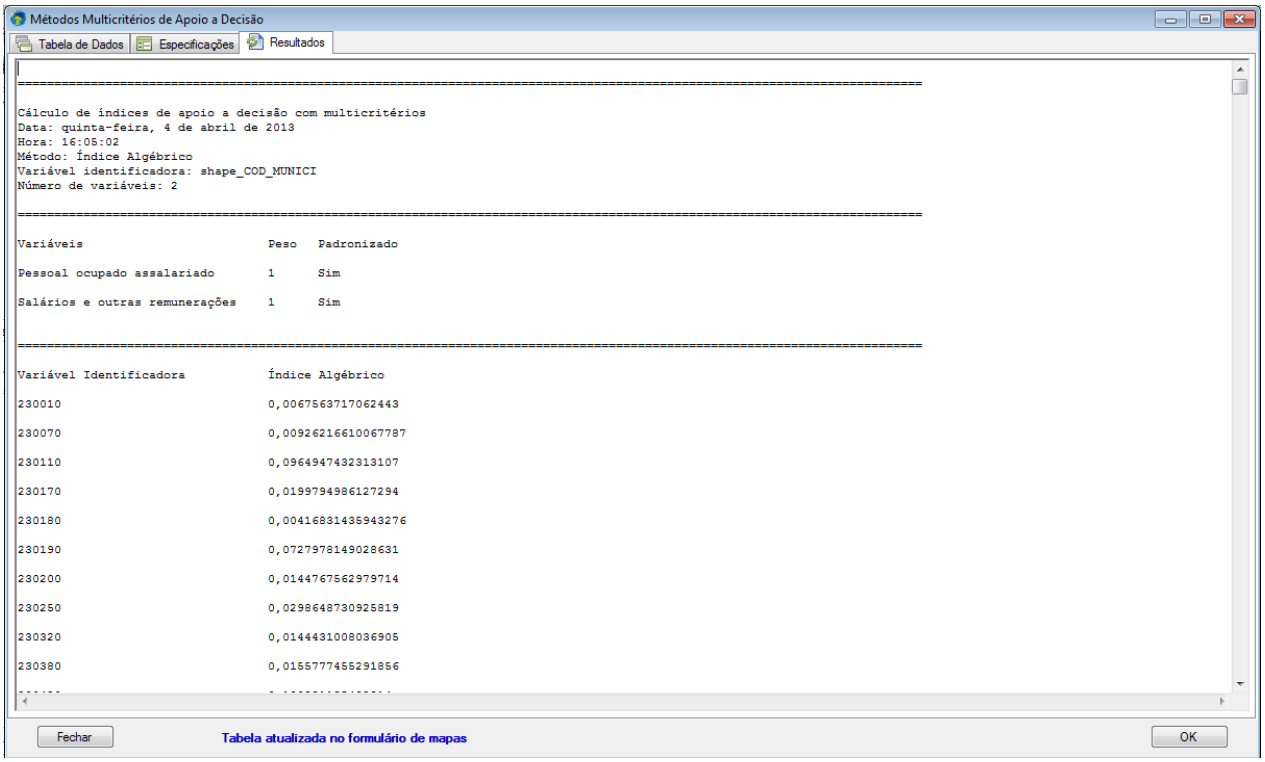


Após clicar na caixa localizada do lado superior esquerdo “Mostrar novas variáveis na tabela de dados” e, posteriormente, no botão “Ok”, o IpeaGEO atualiza a tabela no formulário de mapas.





Clique no botão “Executar” para mostrar o resultado na aba “Resultados”.



Funções Espaciais do IpeaGEO

Áreas Mínimas Comparáveis

Compatibilização de Variáveis

Visão Geral

Um dos problemas mais comuns para os pesquisadores que trabalham com dados municipais, com informações para diferentes anos, é o problema de compatibilização das malhas de municípios. Entre 1990 e 2010, quase mil e quinhentos municípios novos foram criados. De fato, é difícil encontrar dois anos nos quais a malha de municípios seja a mesma. Para maiores detalhes, o leitor pode recorrer a **Carvalho et al (2009)**. Do ponto de vista de análise com dados geográficos, o analista pode encontrar, por exemplo, que a população de um determinado município está diminuindo, simplesmente por que o município original foi dividido em dois novos, sendo que um deles manteve o mesmo nome (e mesmo código) do município de origem.

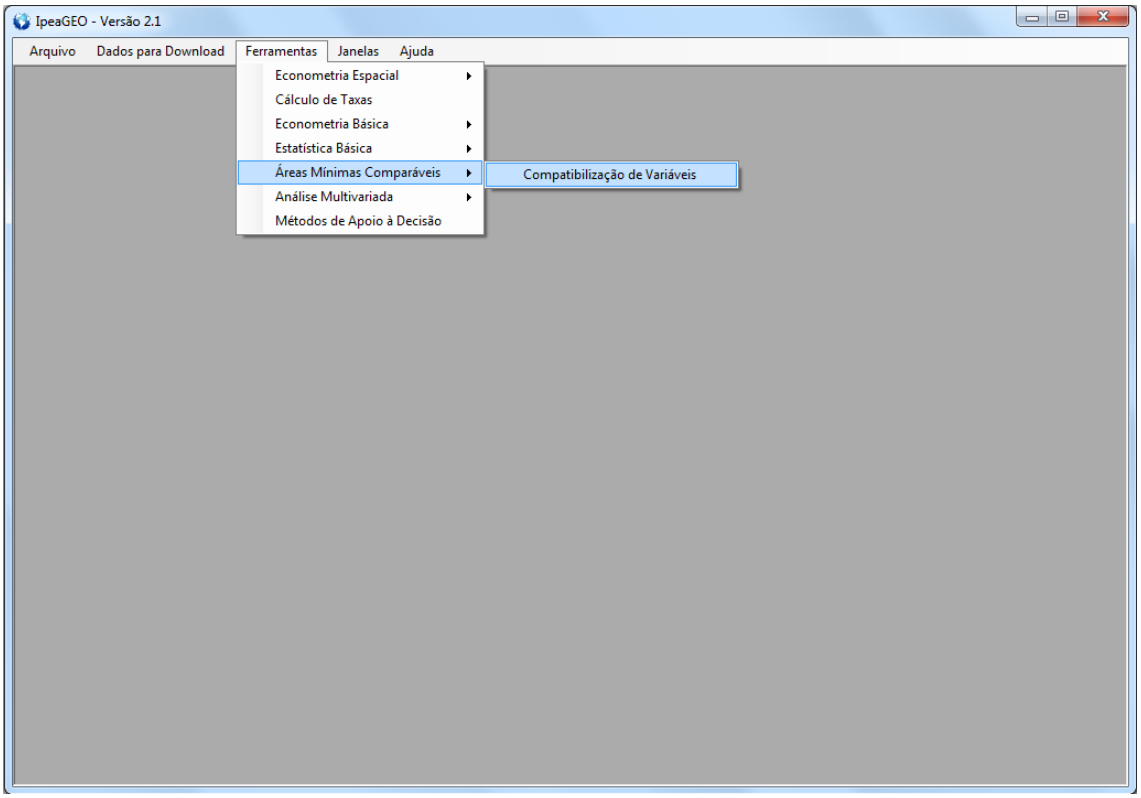
Para contornar esse problema, o Ipea e o IBGE, em um esforço conjunto, desenvolveram o conceito de Áreas Mínimas Comparáveis – AMC. As áreas mínimas comparáveis são uniões de um ou mais municípios, de forma que o analista tenha áreas mínimas de referências, que sejam imutáveis ao longo dos anos. Com isso, o analista poderá agregar informações, em um determinado ano, para os municípios de cada AMC. Como as AMC são constantes ao longo dos anos, todas as análises de comparação do dinamismo dos municípios, por exemplo, podem ser feitos ao nível de AMC's.

O IpeaGEO traz, na sua versão 2.0, uma funcionalidade específica para

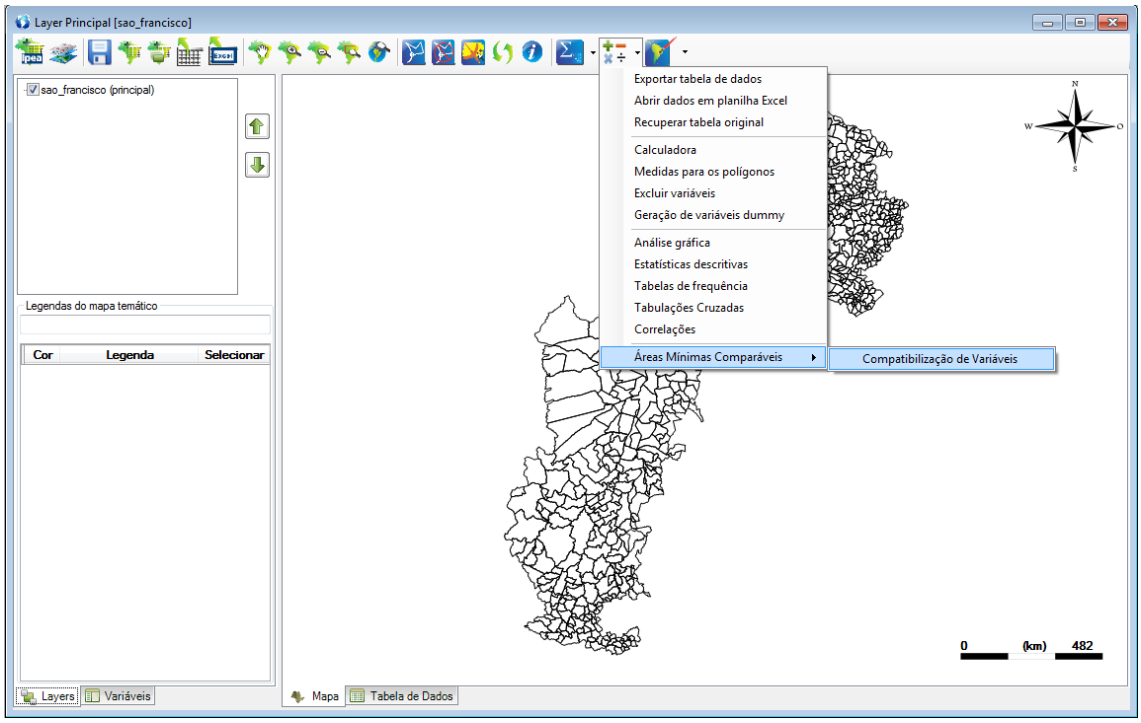
tratamento de informações municipais, possibilitando a sua agregação por áreas mínimas comparáveis. A nova versão do programa já vem com os mapeamentos entre municípios e AMC's para cada ano. Além disso, o sistema IpeaGEO permite que o usuário faça agregações de diferentes formas: somas, médias simples, médias ponderadas, mínimos, máximos, etc.

Exemplo

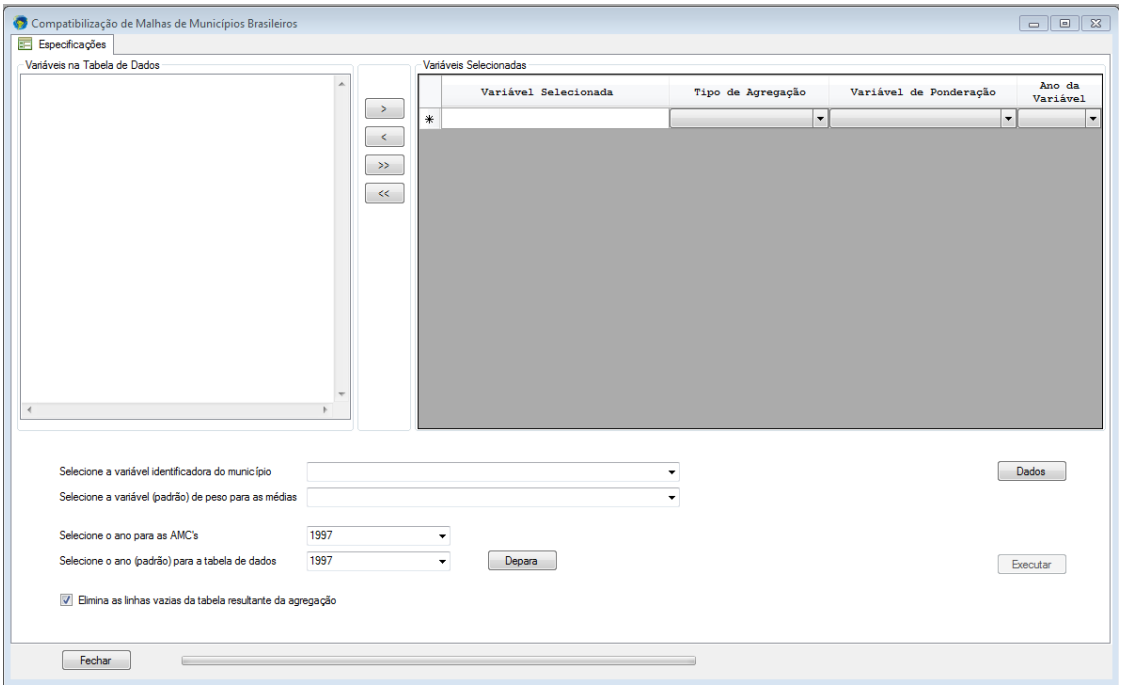
No botão "Ferramentas"**,** clique na opção "Áreas Mínimas Comparáveis" e, em seguida, na opção "Compatibilização de Variáveis".



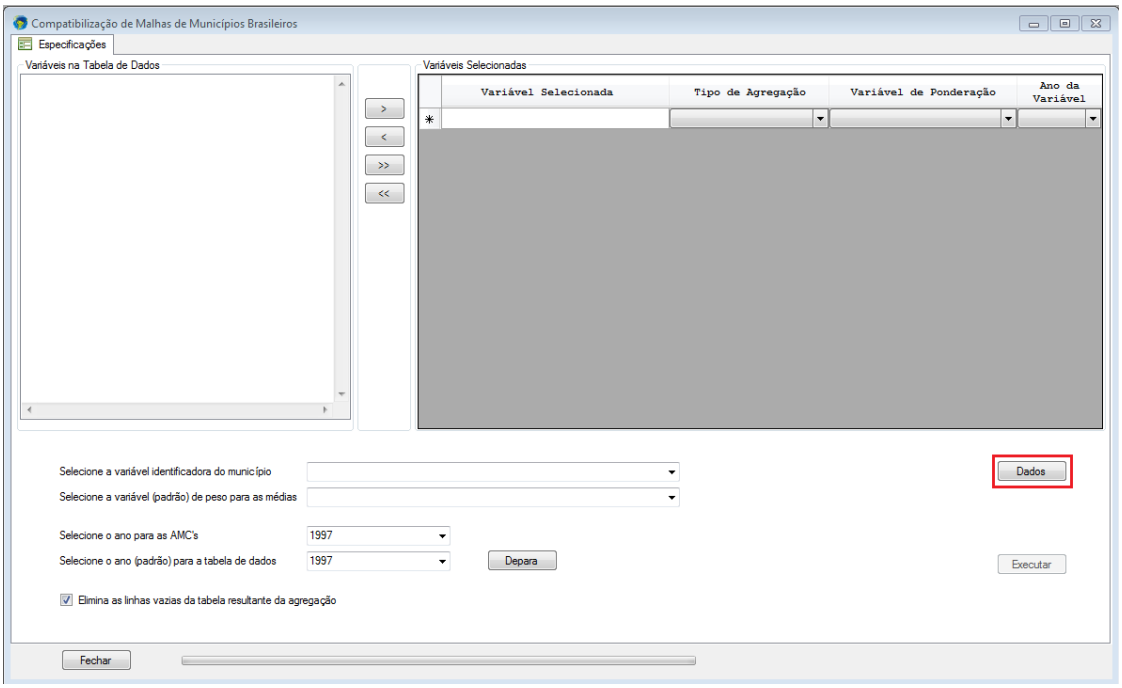
Uma maneira alternativa para acessar o formulário de Compatibilização de Variáveis é, após concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clicar no ícone "Análise exploratória e Manipulação de dados" , opção "Áreas Mínimas Comparáveis" e, em seguida, "Compatibilização de Variáveis":



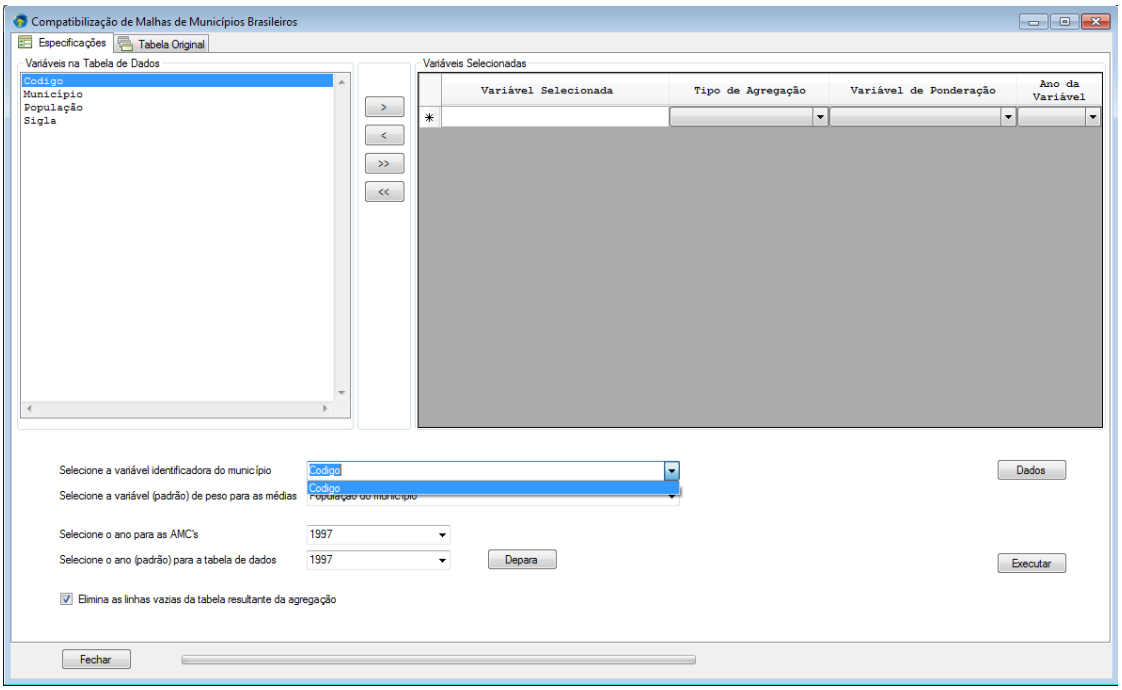
O seguinte formulário irá se abrir:



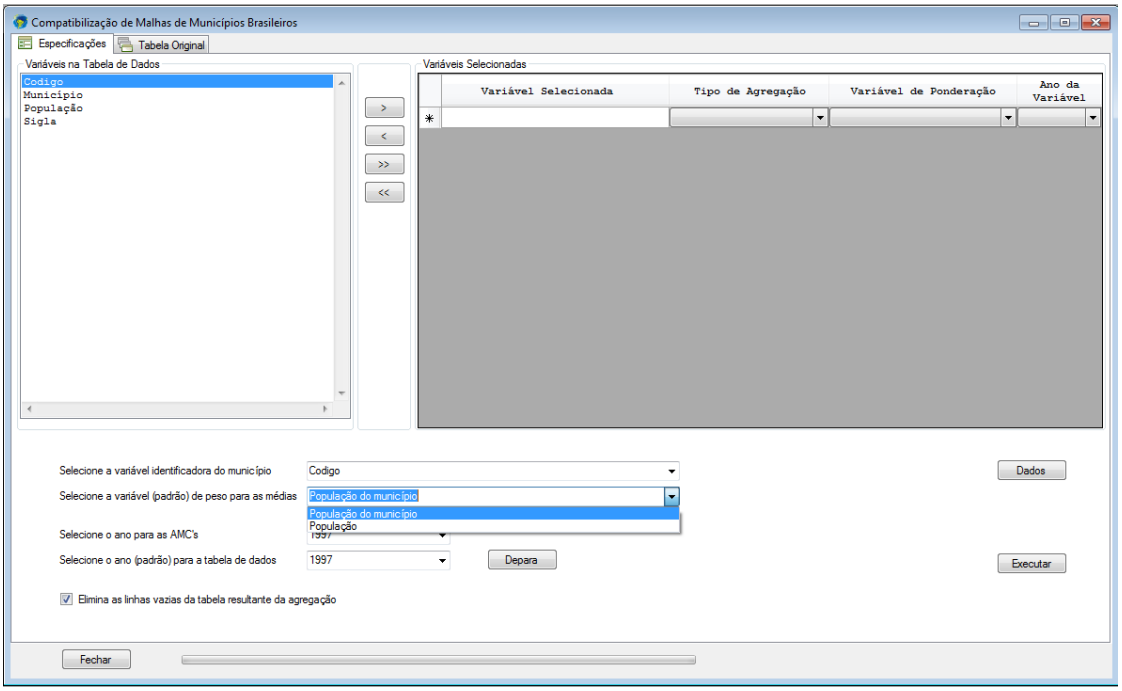
Para importar a tabela que será usada no formulário clique no botão “Dados”. O formulário de importação de tabela de dados irá se abrir.



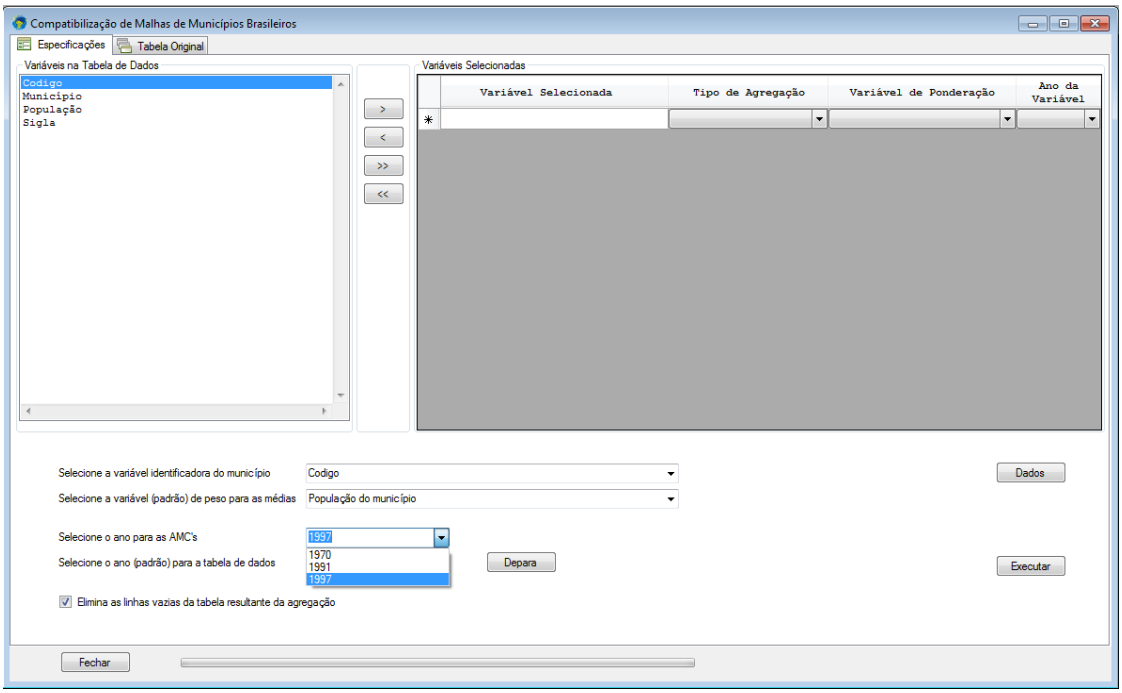
Após importar a tabela a ser utilizada, selecione a variável identificadora do município.

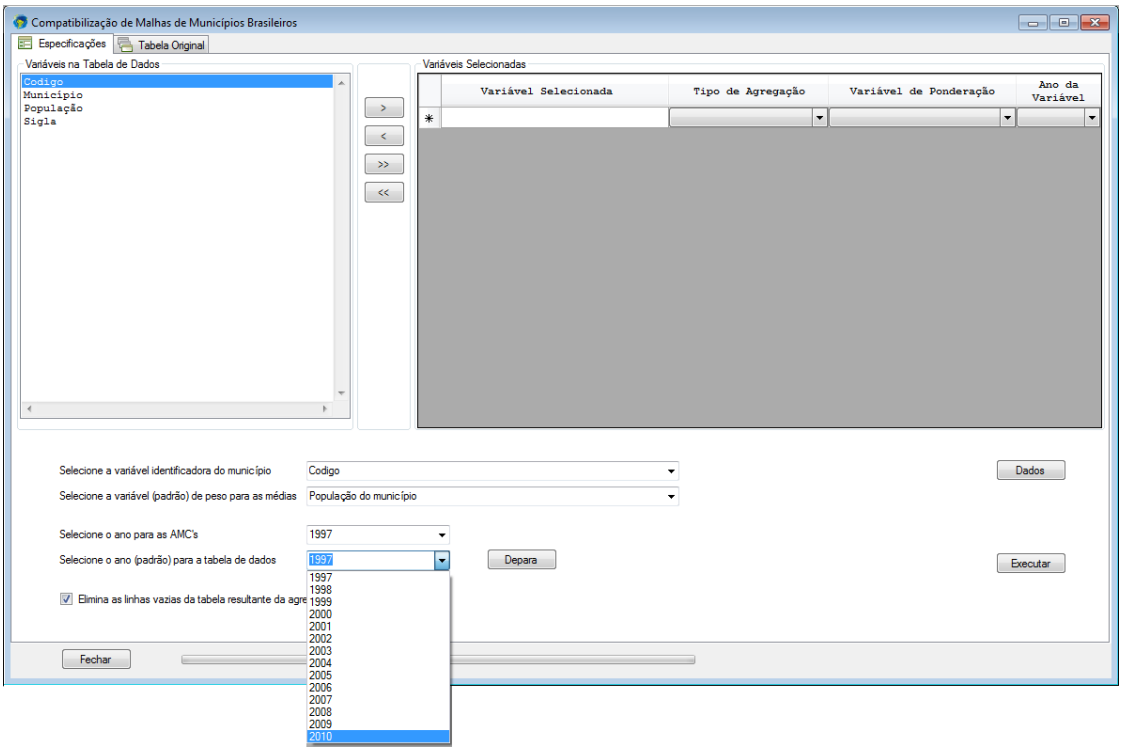


Em seguida selecione a variável (padrão) de peso para as médias.

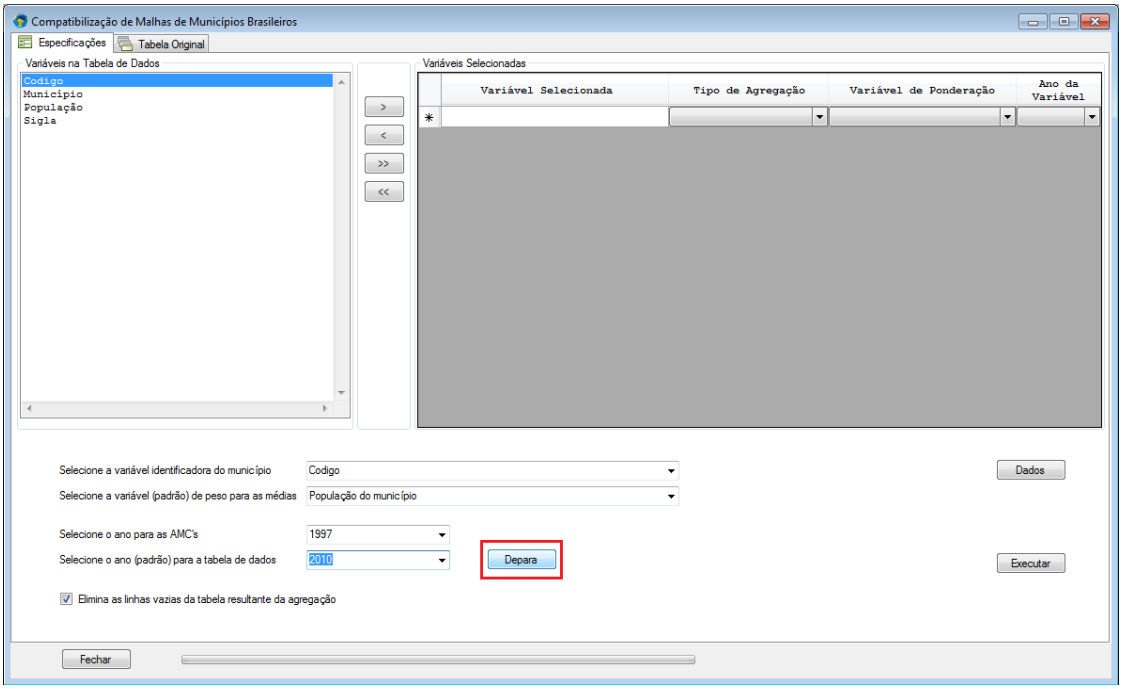


Posteriormente, deve ser escolhido o ano para as AMC's e o ano (padrão) para a tabela de dados.

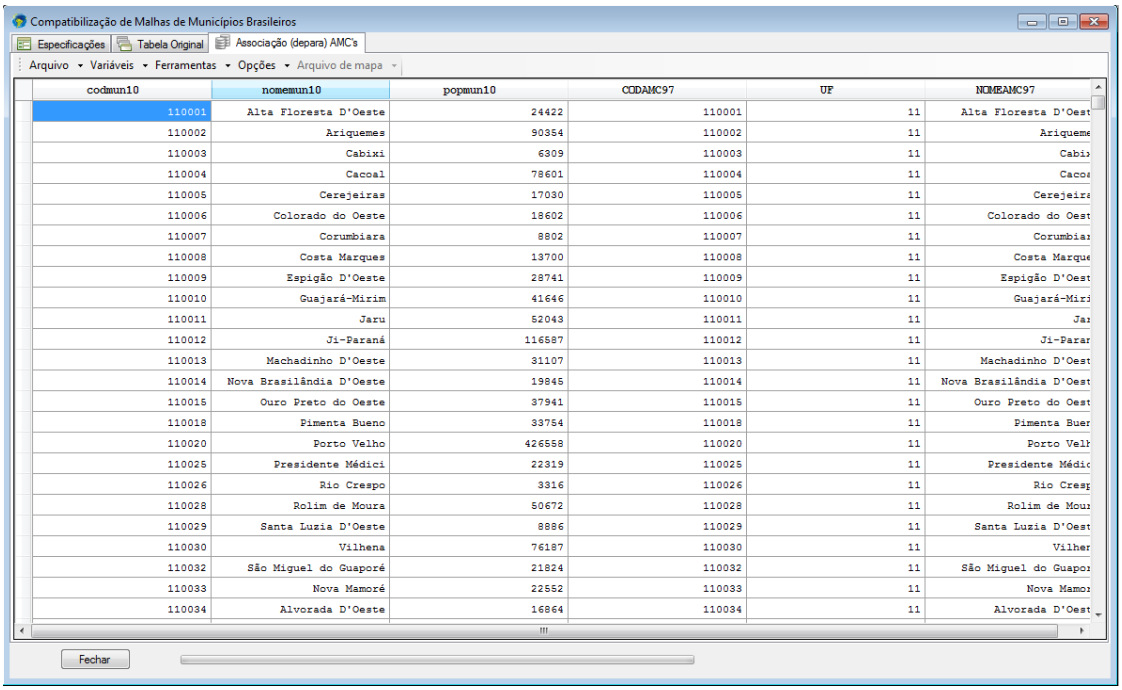




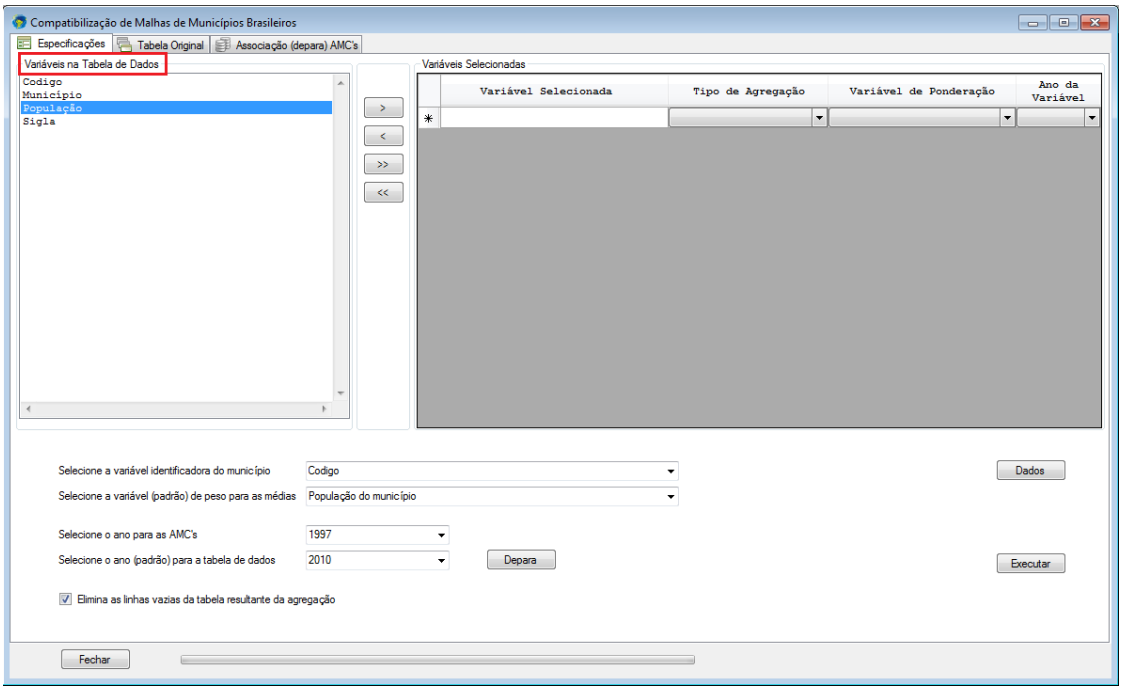
Selecionados os anos desejados, clicar no botão “Depara”.



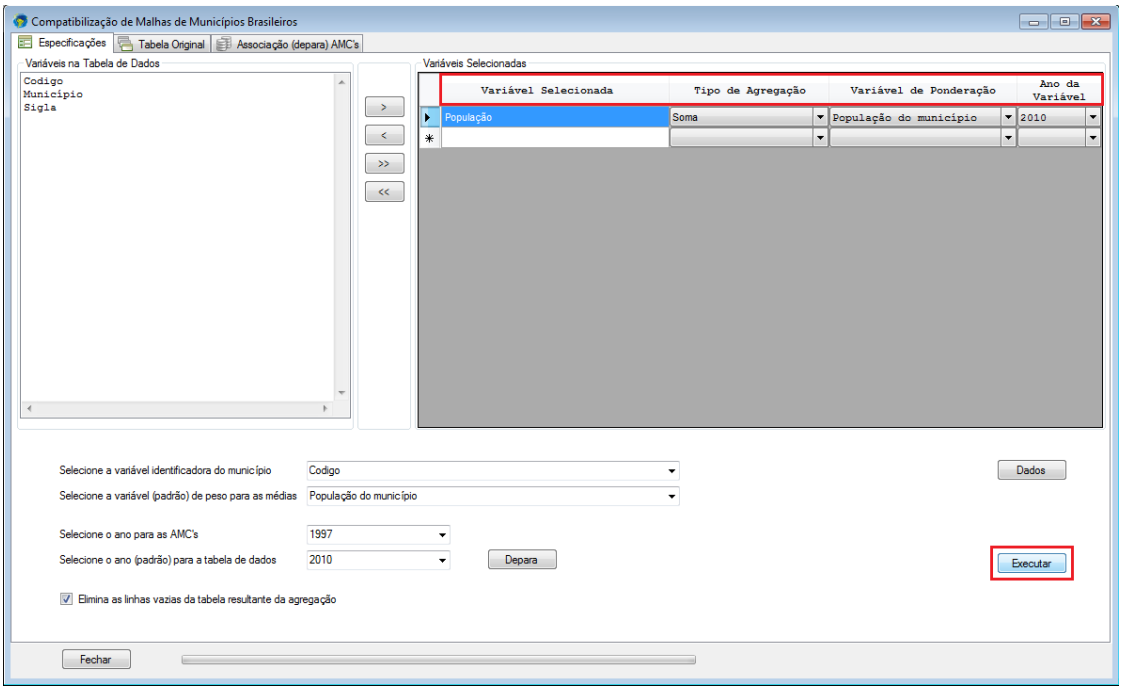
Uma nova aba contendo a associação das AMC’s irá aparecer com os dados das variáveis e anos selecionados.



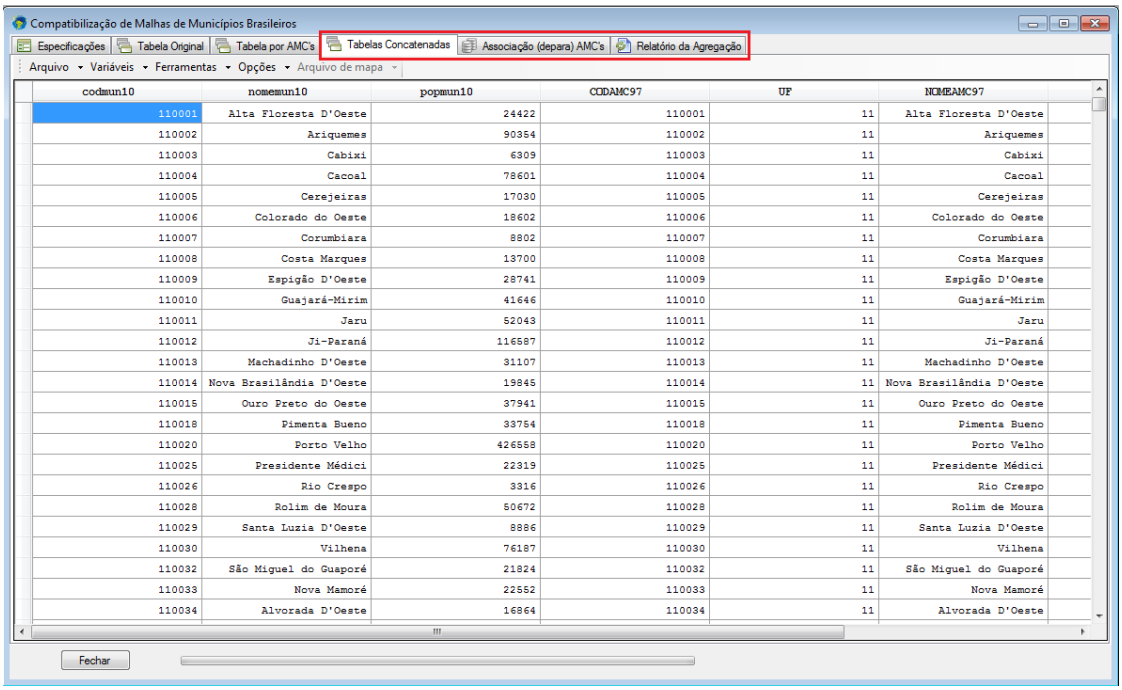
Voltando para a aba "Especificações", selecione as variáveis a serem utilizadas em “Variáveis na Tabela de Dados”.



Ao selecionar a variável desejada estão disponíveis diferentes ferramentas agrupadas nas opções: tipo de agregação, variável de ponderação e ano da variável. Por último, clique no botão “Executar”.



São geradas três novas abas contendo o resultado da compatibilização: Tabelas Concatenadas; Associação (depara) AMC's; e Relatório da Agregação.



codmun10	nomemun10	popmun10	CODAMC97	UF	NOMEAMC97
110001	Alta Floresta D'Oeste	24422	110001	11	Alta Floresta D'Oeste
110002	Ariquemes	90354	110002	11	Ariquemes
110003	Cabixi	6309	110003	11	Cabixi
110004	Cacoal	78601	110004	11	Cacoal
110005	Cerejeiras	17030	110005	11	Cerejeiras
110006	Colorado do Oeste	18602	110006	11	Colorado do Oeste
110007	Corumbiara	8802	110007	11	Corumbiara
110008	Costa Marques	13700	110008	11	Costa Marques
110009	Espigão D'Oeste	28741	110009	11	Espigão D'Oeste
110010	Guajará-Mirim	41646	110010	11	Guajará-Mirim
110011	Jaru	52043	110011	11	Jaru
110012	Ji-Paraná	116587	110012	11	Ji-Paraná
110013	Machadinho D'Oeste	31107	110013	11	Machadinho D'Oeste
110014	Nova Brasilândia D'Oeste	19845	110014	11	Nova Brasilândia D'Oeste
110015	Ouro Preto do Oeste	37941	110015	11	Ouro Preto do Oeste
110018	Pimenta Bueno	33754	110018	11	Pimenta Bueno
110020	Porto Velho	426558	110020	11	Porto Velho
110025	Presidente Médici	22319	110025	11	Presidente Médici
110026	Rio Crespo	3316	110026	11	Rio Crespo
110028	Rolim de Moura	50672	110028	11	Rolim de Moura
110029	Santa Luzia D'Oeste	8886	110029	11	Santa Luzia D'Oeste
110030	Vilhena	76187	110030	11	Vilhena
110032	São Miguel do Guaporé	21824	110032	11	São Miguel do Guaporé
110033	Nova Mamoré	22552	110033	11	Nova Mamoré
110034	Alvorada D'Oeste	16864	110034	11	Alvorada D'Oeste

Conglomerados Espaciais

Hierárquico

Visão Geral

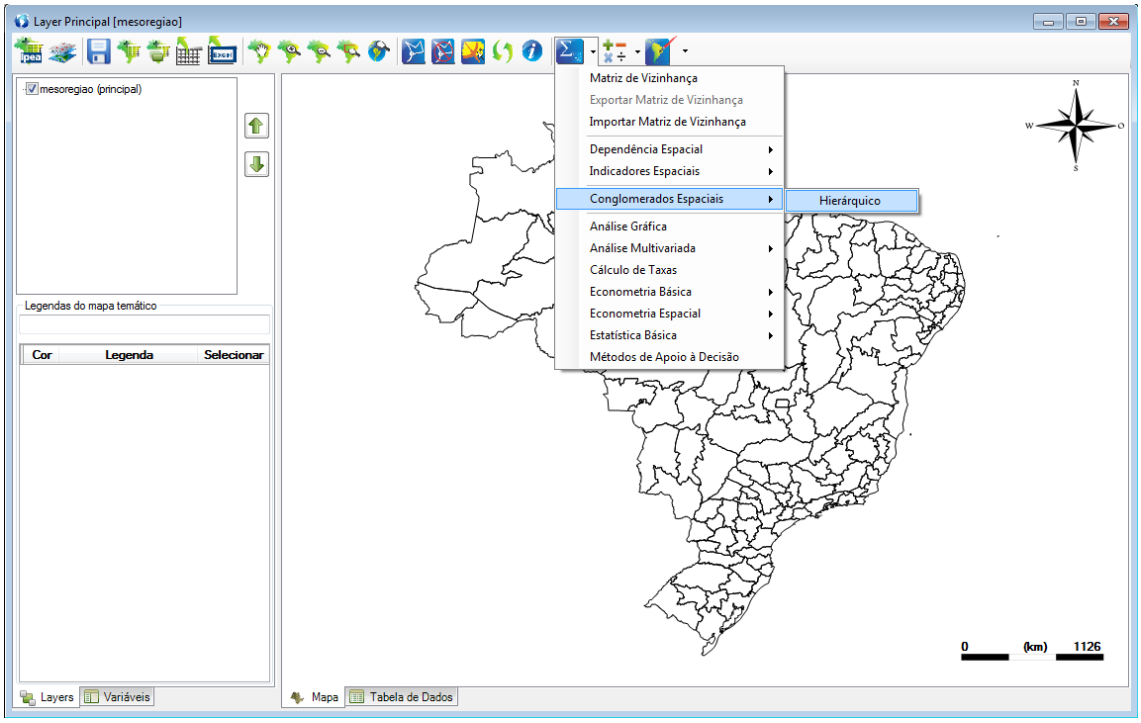
Em mineração de dados, a análise de agrupamentos hierárquicos é um método de análise de clusters que visa construir uma hierarquia de conglomerados. As estratégias de agrupamento hierárquico geralmente se enquadram em dois tipos: aglomerativo e divisionista.

Nos algoritmos tradicionais de clusterização (hierárquica ou não), quando são agrupadas unidades geográficas do tipo municípios ou setores censitários, não necessariamente os grupos homogêneos são formados por municípios ou setores censitários estritamente vizinhos. Pode acontecer que, em um mesmo agrupamento (cluster), existam municípios geograficamente separados. A formação de agrupamentos homogêneos de municípios, com componentes não necessariamente contíguos, pode não ser um problema em muitas das aplicações. De fato, pode acontecer que o analista ou pesquisador esteja interessado justamente em identificar se existem regiões (setores censitários, áreas de ponderação) na periferia de São Paulo, por exemplo, que são semelhantes, em termos de atributos socioeconômicos, a regiões no centro da cidade.

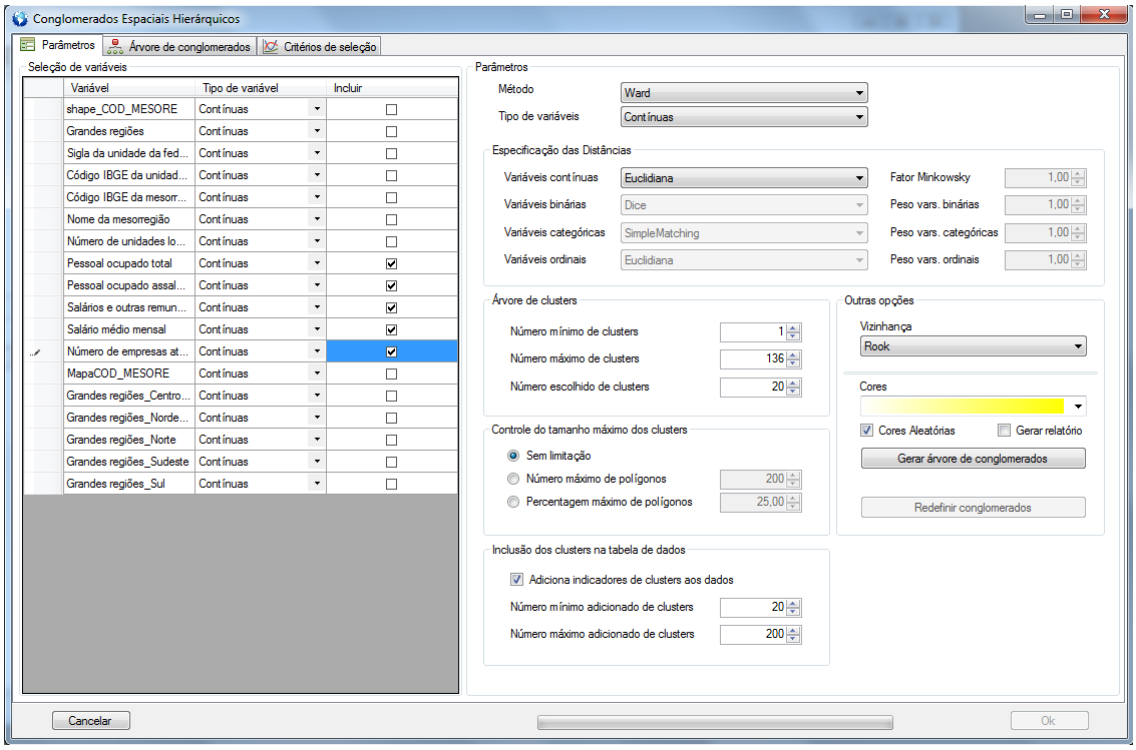
Os algoritmos de agrupamento hierárquico espacial são criados por meio de modificações no método de agrupamento tradicional, de forma a incorporar a restrição de unidades geográficas (por exemplo, municípios, setores censitários, Unidades da Federação, áreas de ponderação) contíguas.

Exemplo

Após concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , opção “Conglomerados Espaciais” e, em seguida, na opção “Hierárquico”.

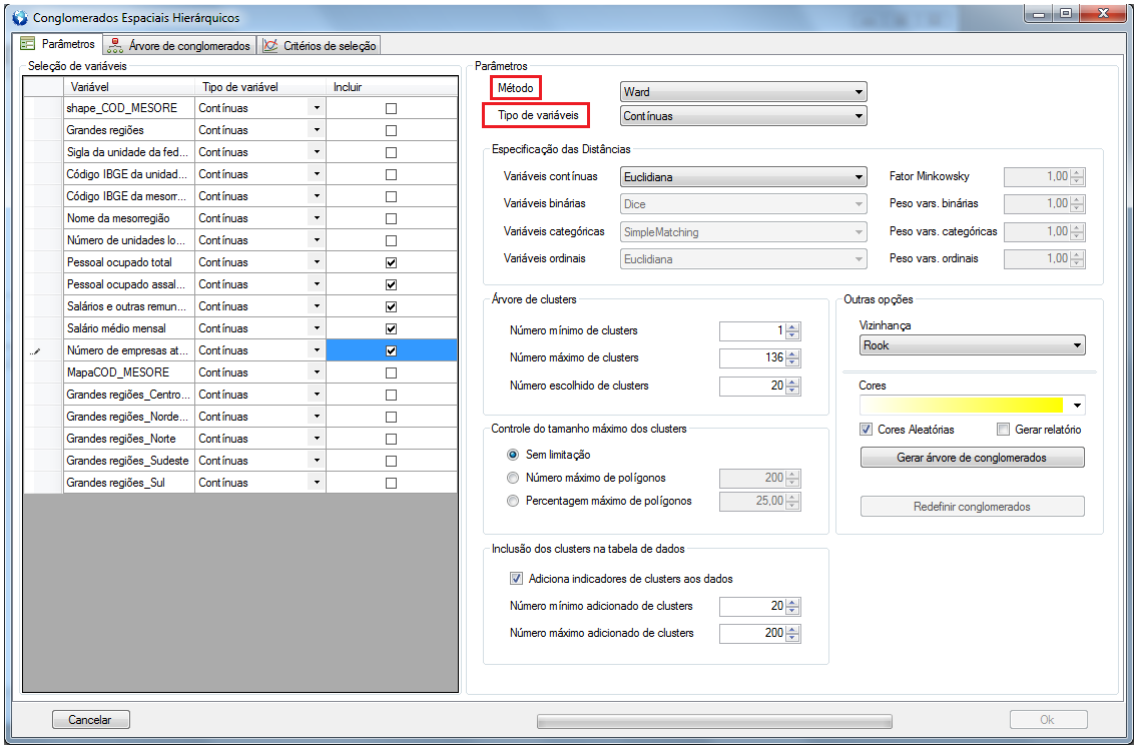


O seguinte formulário irá se abrir:

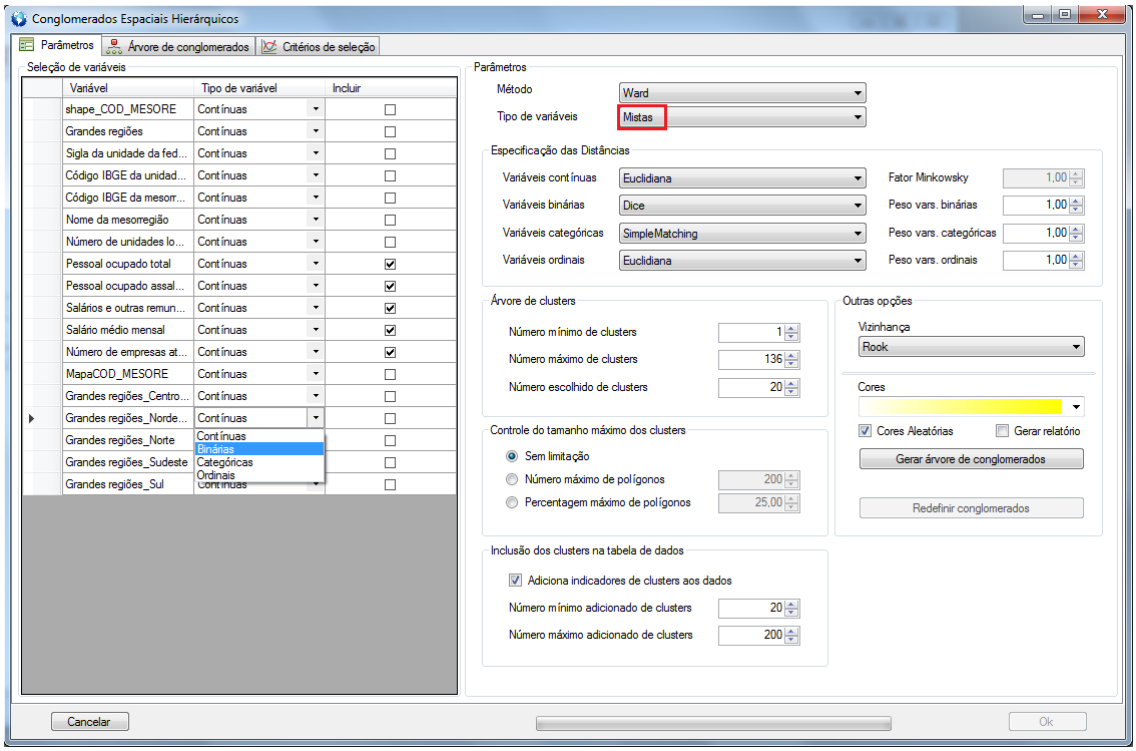


Na aba “Parâmetros” em “Seleção de variáveis”, selecione as variáveis desejadas para a análise clicando em “Incluir”.

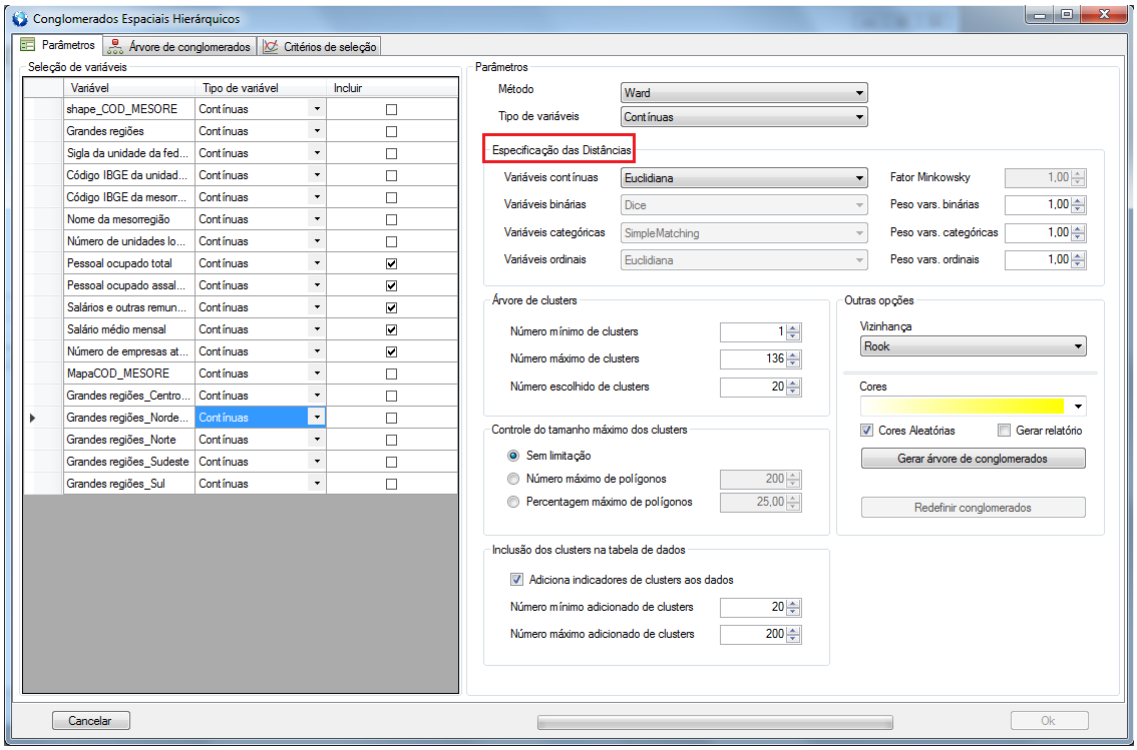
Depois de selecionadas as varáveis desejadas, selecione o método de análise, que pode ser: “Ward”, ligação média (Average Linkage), centróide (Centroid), ligação completa (Complete Linkage) e ligação simples (Single Linkage) e o tipo de variável em questão.



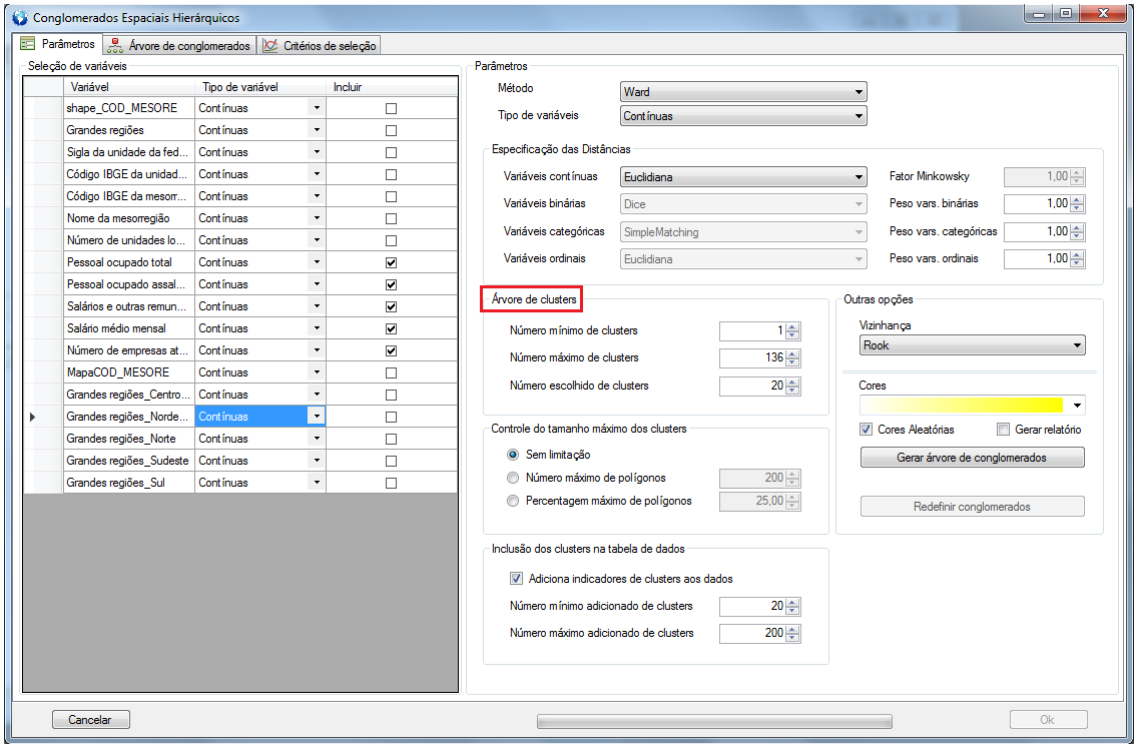
Para utilizar mais de um tipo de variável para o agrupamento, selecione o tipo de variável “mista” e, em seguida, altere os tipos de acordo com o interesse na seção “Tipo de variável” em “Seleção de variáveis”.



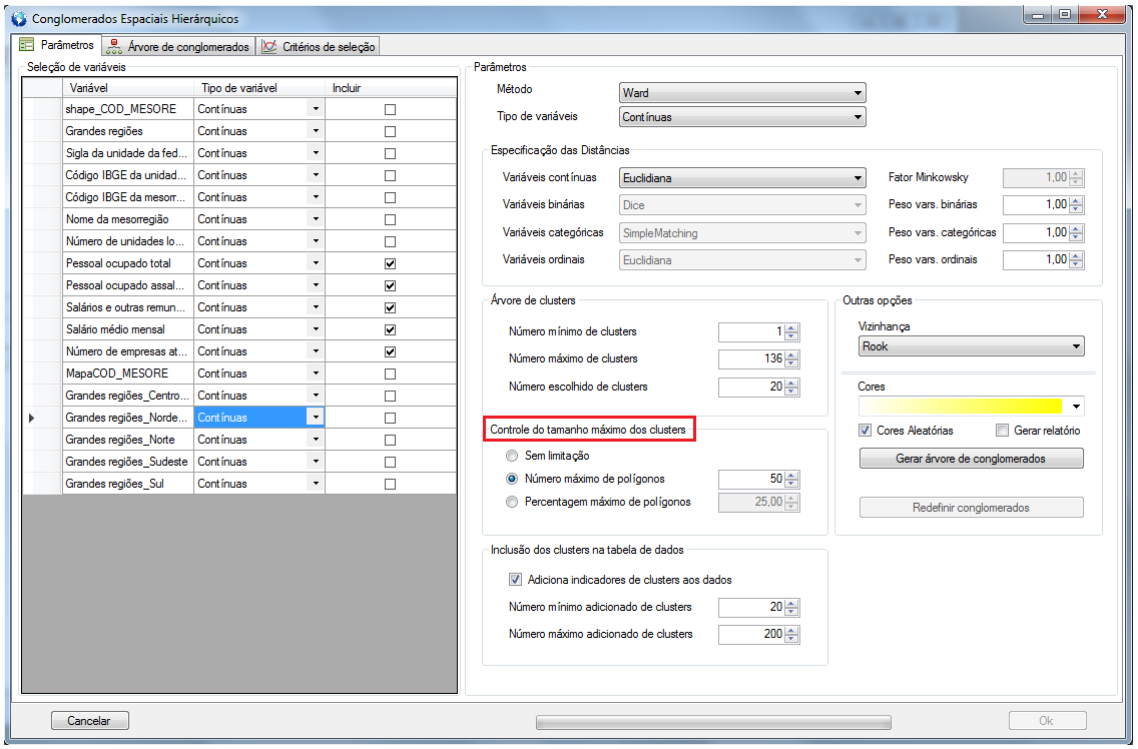
Em “Especificação das Distâncias” escolha qual a opção desejada para calcular a distância entre as variáveis selecionadas, de acordo com o seu tipo (contínua, binária, categórica e ordinal), bem como o peso que cada variável terá na análise e o fator Minkowsky (caso seja uma variável contínua com distância Minkowsky).



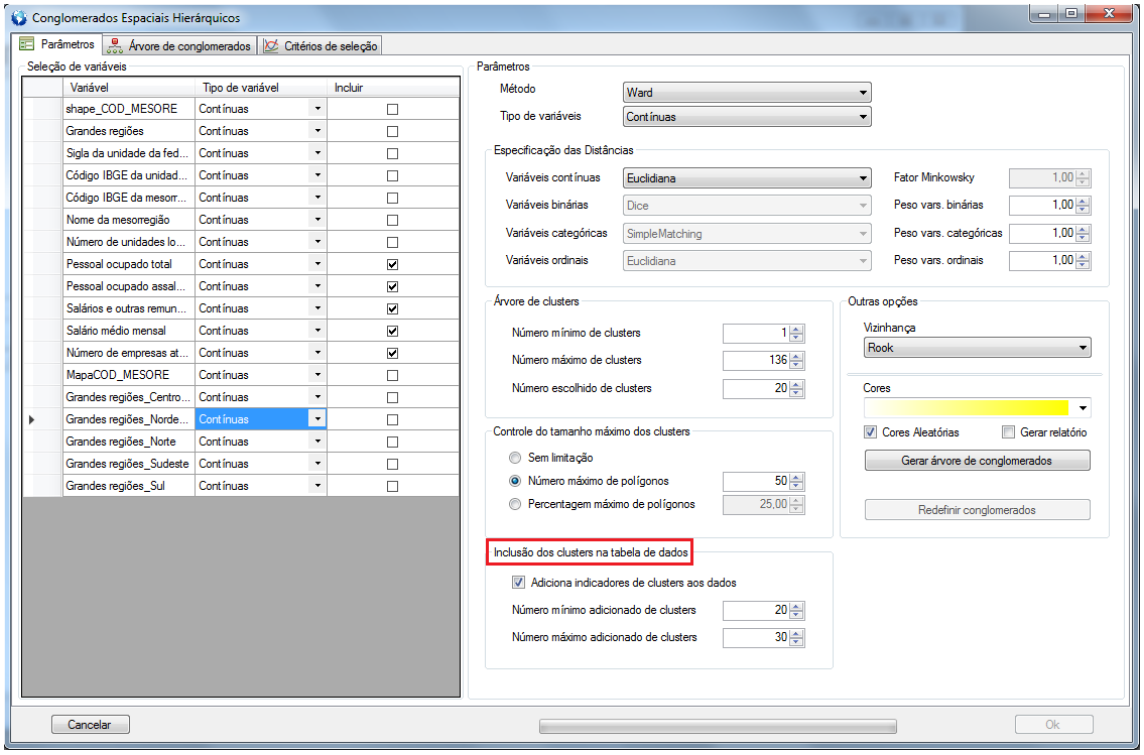
Em “Árvore de clusters”, selecione as seguintes opções: número mínimo de clusters, número máximo de clusters e número escolhido de clusters.



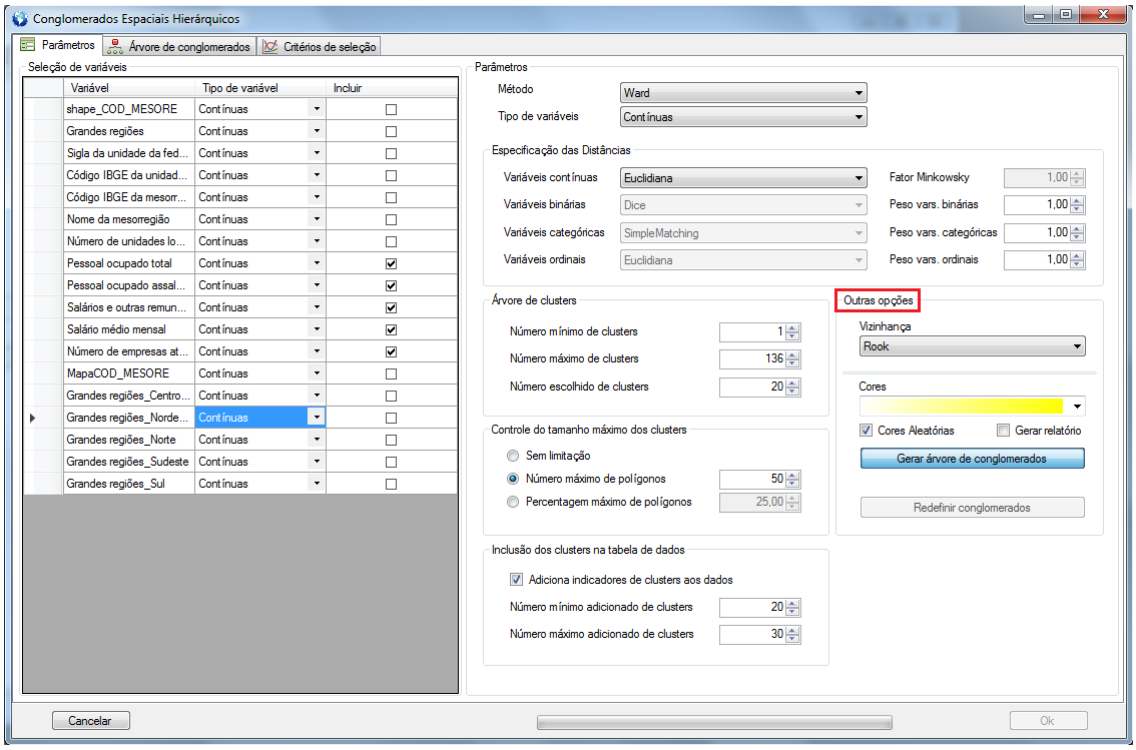
Em “Controle do tamanho máximo dos clusters” existem três opções a disposição do usuário: sem limitação, número máximo de polígonos e percentagem máximo de polígonos.



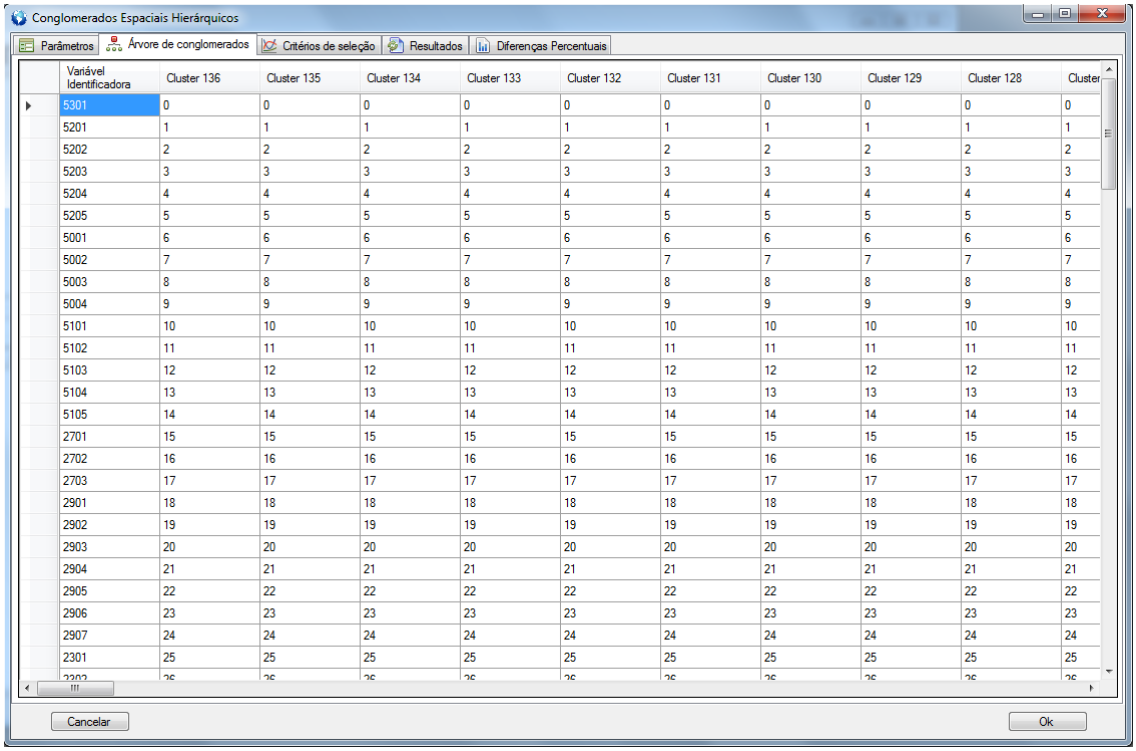
Em “Inclusão dos clusters na tabela de dados” existe a opção de adicionar na tabela de dados os clusters gerados. Para isso, o usuário pode escolher o número mínimo e máximo de colunas a serem adicionadas, assim como adicionar os indicadores de clusters aos dados.



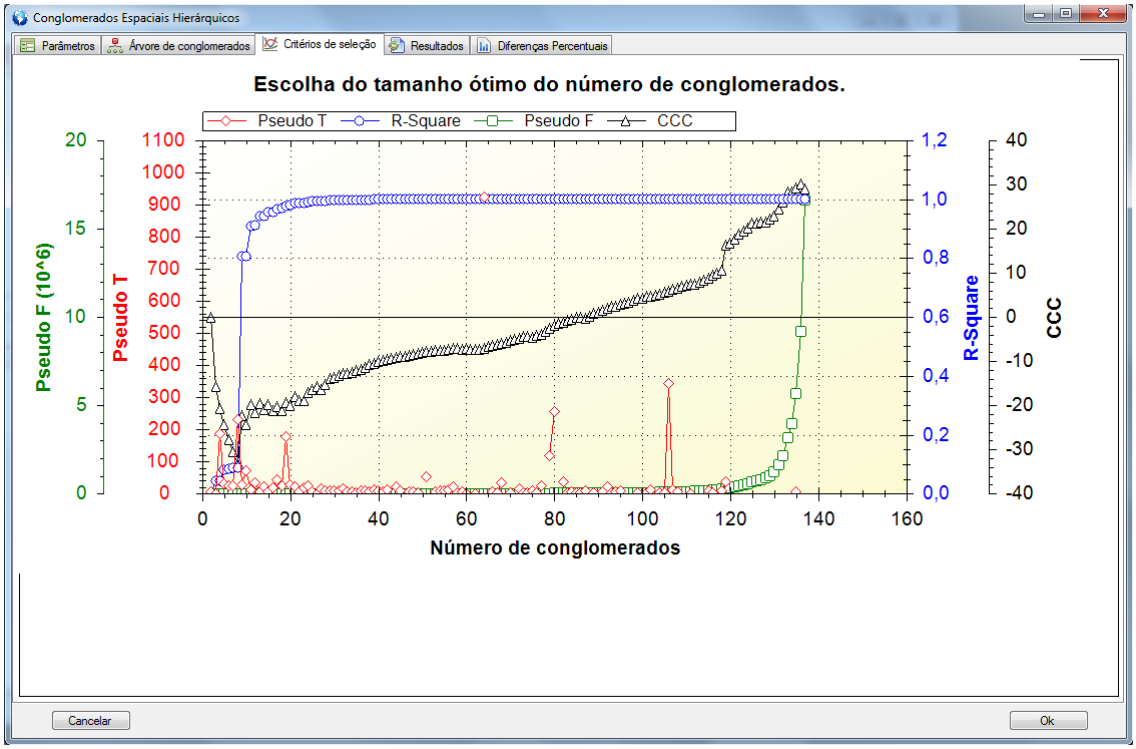
Na parte “Outras opções” escolha qual o tipo de método a ser aplicado para a definição da matriz de vizinhança e quais as cores a serem utilizadas na construção do mapa com os resultados da análise. Caso seja de interesse, pode-se optar pelas opções “Cores Aleatórias” e “Gerar relatório”. Definidos os parâmetros, clique no botão “Gerar árvore de conglomerados”.

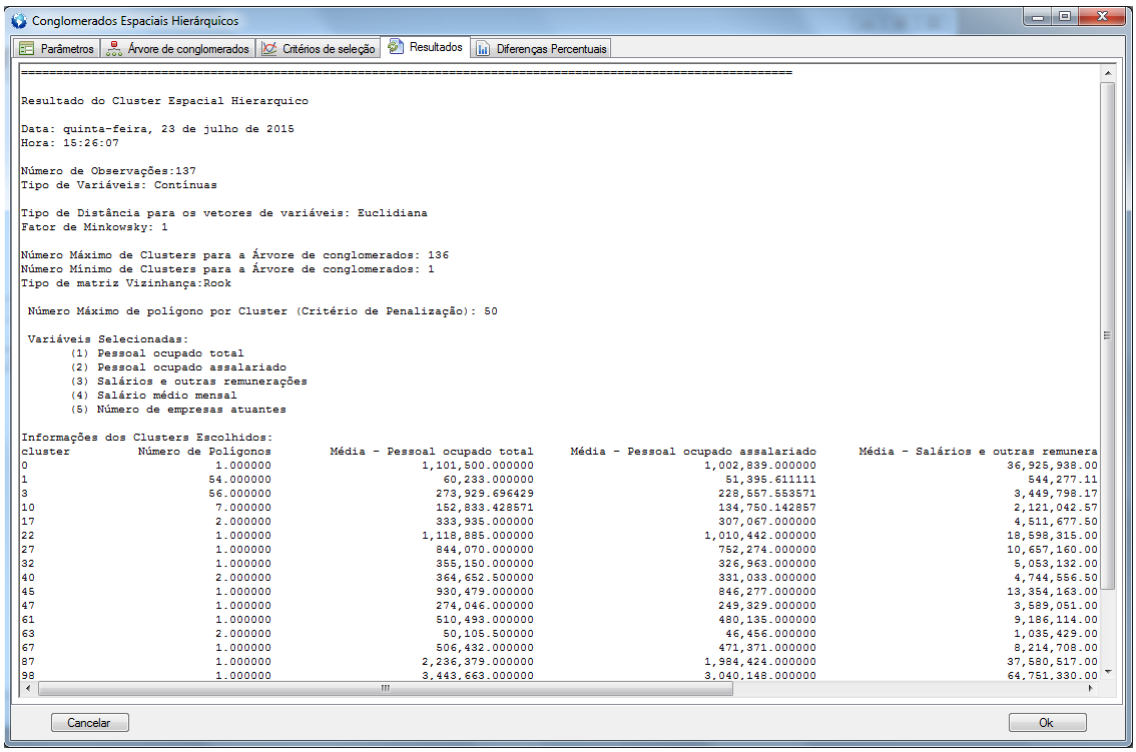


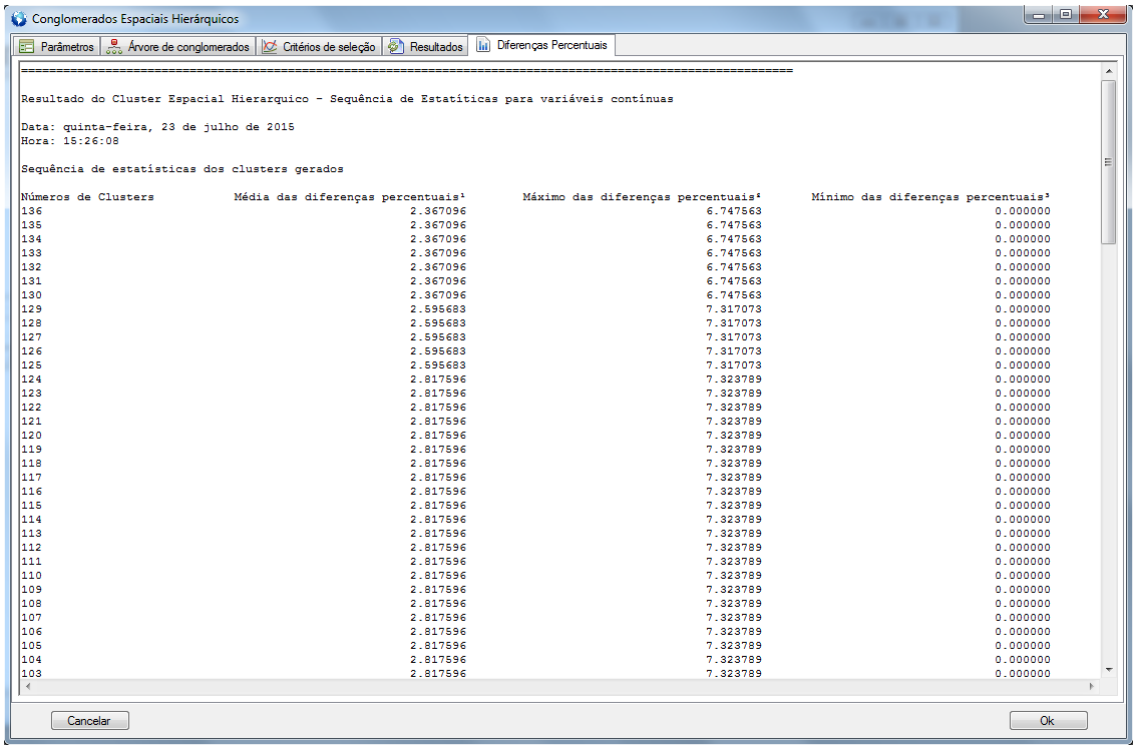
Quatro novas abas irão se abrir. A primeira contendo a “Árvore de conglomerado”; a segunda os “Critérios de seleção”, que contém o gráfico para definição do número ótimo de clusters a serem formados; a terceira os “Resultados” obtidos e; a quarta as “Diferenças Percentuais” contendo as diferenças percentuais entre as médias de cada variável para cada par de clusters (disponível para o caso de variáveis contínuas). Seguindo o exemplo acima, as imagens de cada uma destas abas estão evidenciadas abaixo.



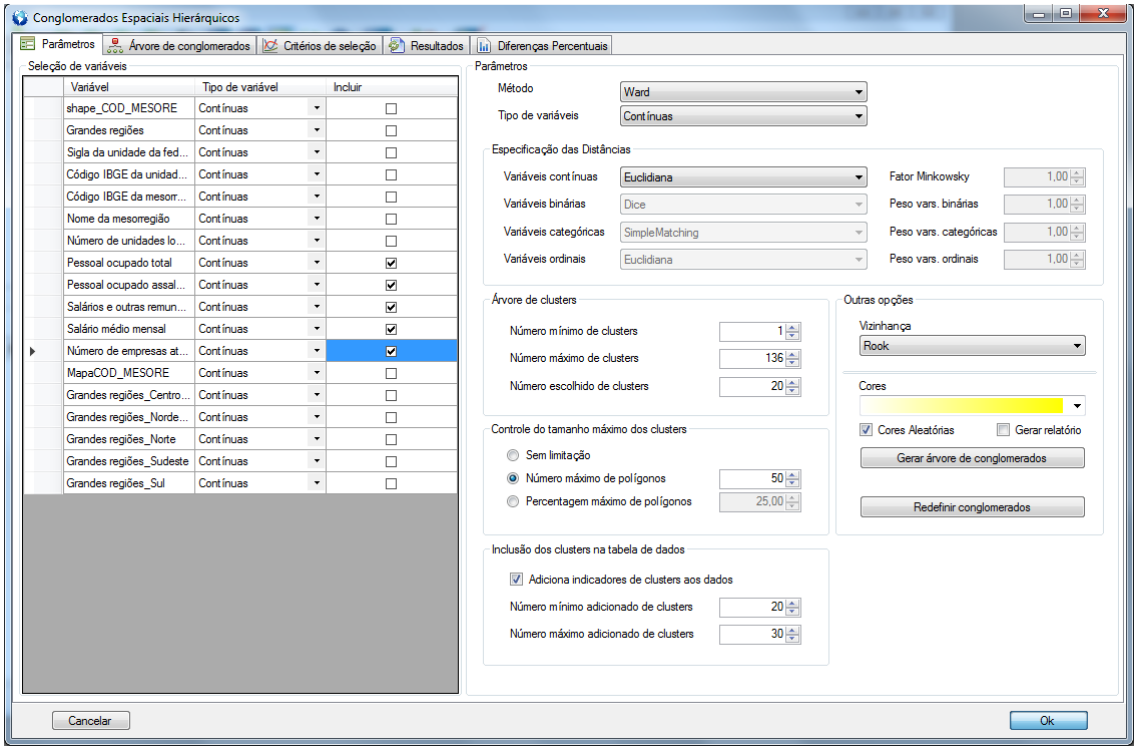
No gráfico é possível dar zoom ou selecionar apenas uma parte do gráfico para melhor visualização. Essas manipulações são feitas com o mouse. Ao se utilizar o zoom os eixos do gráfico são ajustados automaticamente.







Por último, clicando no botão “Ok” será visualizado o novo mapa gerado.



Mapa gerado:

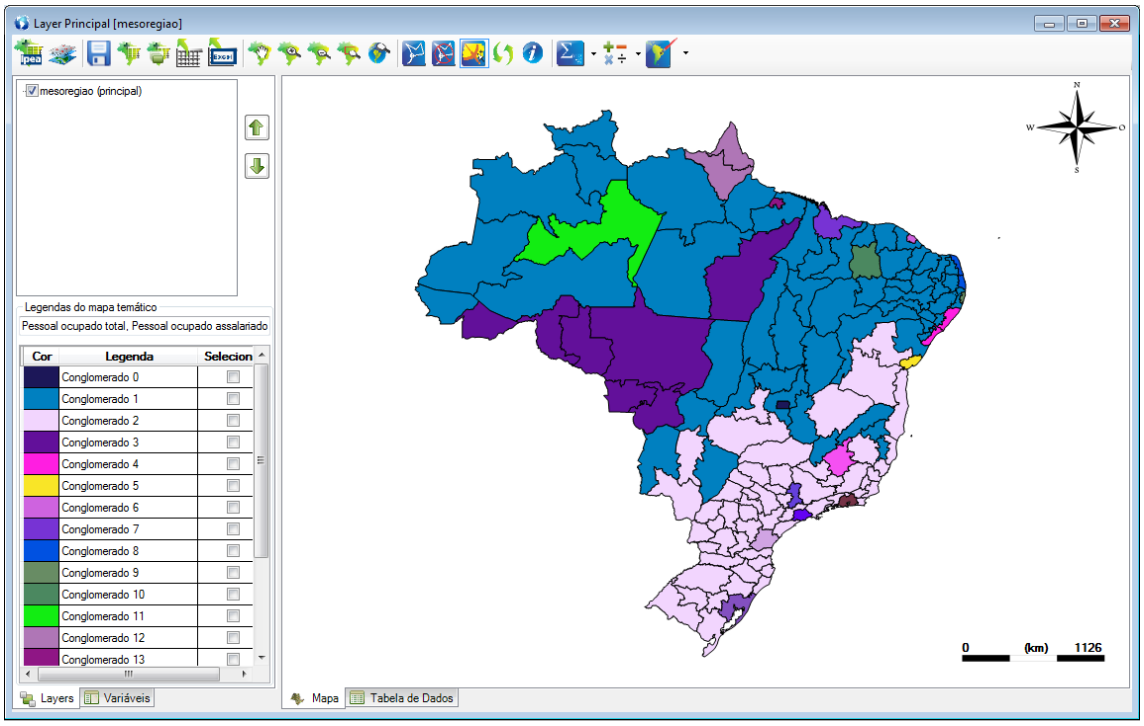
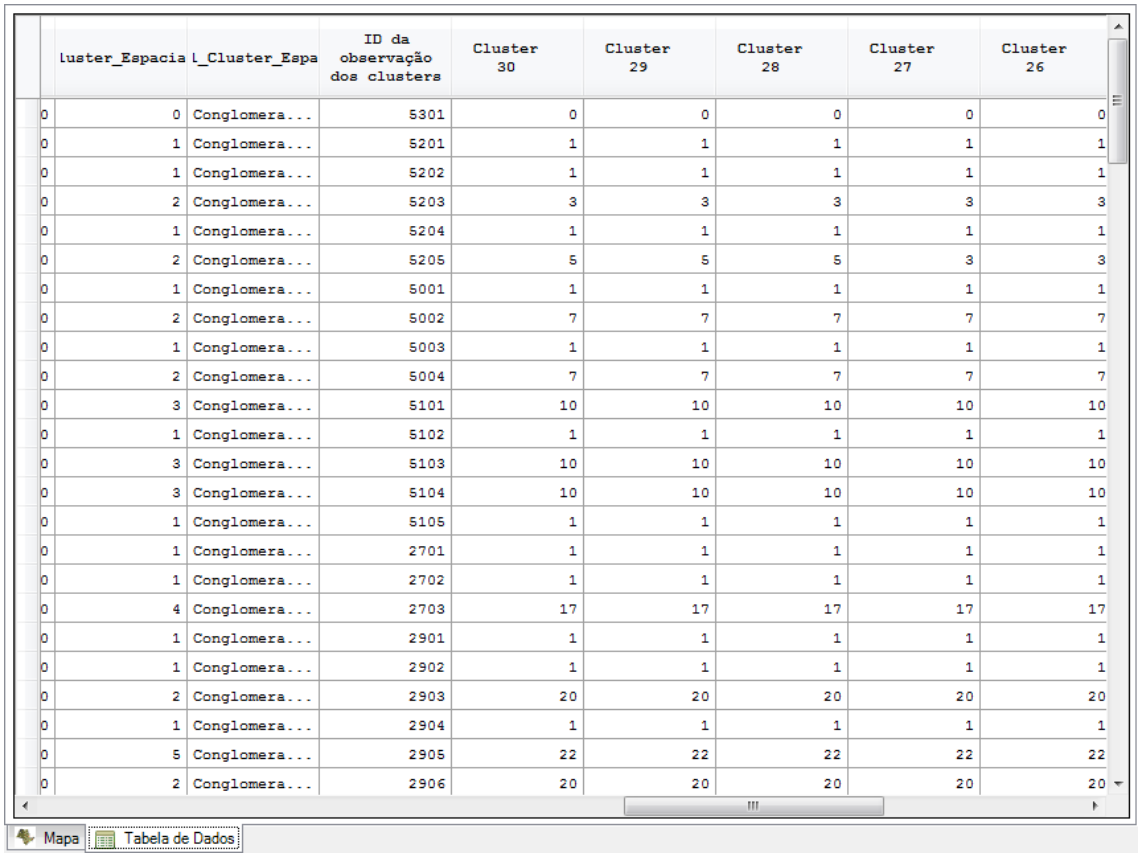


Tabela de dados atualizada com os clusters gerados:



	luster_Espacia	l_Cluster_Espa	ID da observação dos clusters	Cluster 30	Cluster 29	Cluster 28	Cluster 27	Cluster 26
0	0	Conglomera...	5301	0	0	0	0	0
0	1	Conglomera...	5201	1	1	1	1	1
0	1	Conglomera...	5202	1	1	1	1	1
0	2	Conglomera...	5203	3	3	3	3	3
0	1	Conglomera...	5204	1	1	1	1	1
0	2	Conglomera...	5205	5	5	5	3	3
0	1	Conglomera...	5001	1	1	1	1	1
0	2	Conglomera...	5002	7	7	7	7	7
0	1	Conglomera...	5003	1	1	1	1	1
0	2	Conglomera...	5004	7	7	7	7	7
0	3	Conglomera...	5101	10	10	10	10	10
0	1	Conglomera...	5102	1	1	1	1	1
0	3	Conglomera...	5103	10	10	10	10	10
0	3	Conglomera...	5104	10	10	10	10	10
0	1	Conglomera...	5105	1	1	1	1	1
0	1	Conglomera...	2701	1	1	1	1	1
0	1	Conglomera...	2702	1	1	1	1	1
0	4	Conglomera...	2703	17	17	17	17	17
0	1	Conglomera...	2901	1	1	1	1	1
0	1	Conglomera...	2902	1	1	1	1	1
0	2	Conglomera...	2903	20	20	20	20	20
0	1	Conglomera...	2904	1	1	1	1	1
0	5	Conglomera...	2905	22	22	22	22	22
0	2	Conglomera...	2906	20	20	20	20	20

Mapa | Tabela de Dados

Dependência Espacial

Visão Geral

Uma etapa fundamental na análise espacial é a mensuração da dependência espacial entre as observações, ou seja, testar se há ou não dependência geográfica entre os eventos de estudo. Para isso, se utiliza duas abordagens diferentes: (1) é possível estimar o quanto um determinado elemento está relacionado com seus vizinhos de fronteira, de acordo com uma variável de interesse – essa relação pode ser avaliada pelos índices de autocorrelação espacial, tais como os índices de Moran, Geary e Getis-Ord Gi; (2) em alguns casos é importante verificar se existem agrupamentos de unidades geográficas similares quanto às variáveis em análise – isso pode ser feito por meio dos indicadores de conglomerados espaciais de Tango e Rogerson.

Índices Globais

Visão Geral

Os indicadores globais são úteis por apresentarem uma única medida de tendência espacial para toda a região em estudo. Assim, são as técnicas mais utilizadas na fase inicial de análises exploratórias de dados de área. Além do teor exploratório, permitem ainda efetuar testes de hipótese, cuja hipótese nula é que não há dependência espacial entre as unidades geográficas, de acordo com as variáveis originais.

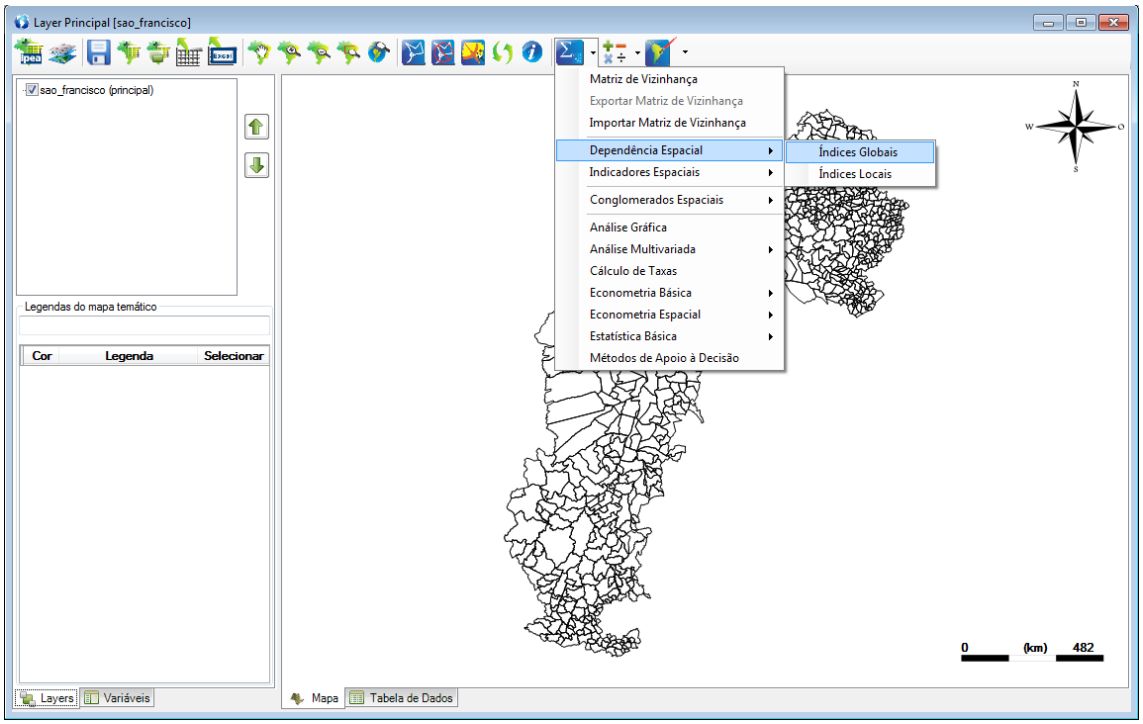
As opções de indicadores disponíveis no IpeaGEO são:

- Indicadores de Dependência: Moran Geral; Getis-Ord Gi; e Geary.
- Indicadores de Conglomerados: Tango; e Rogerson.

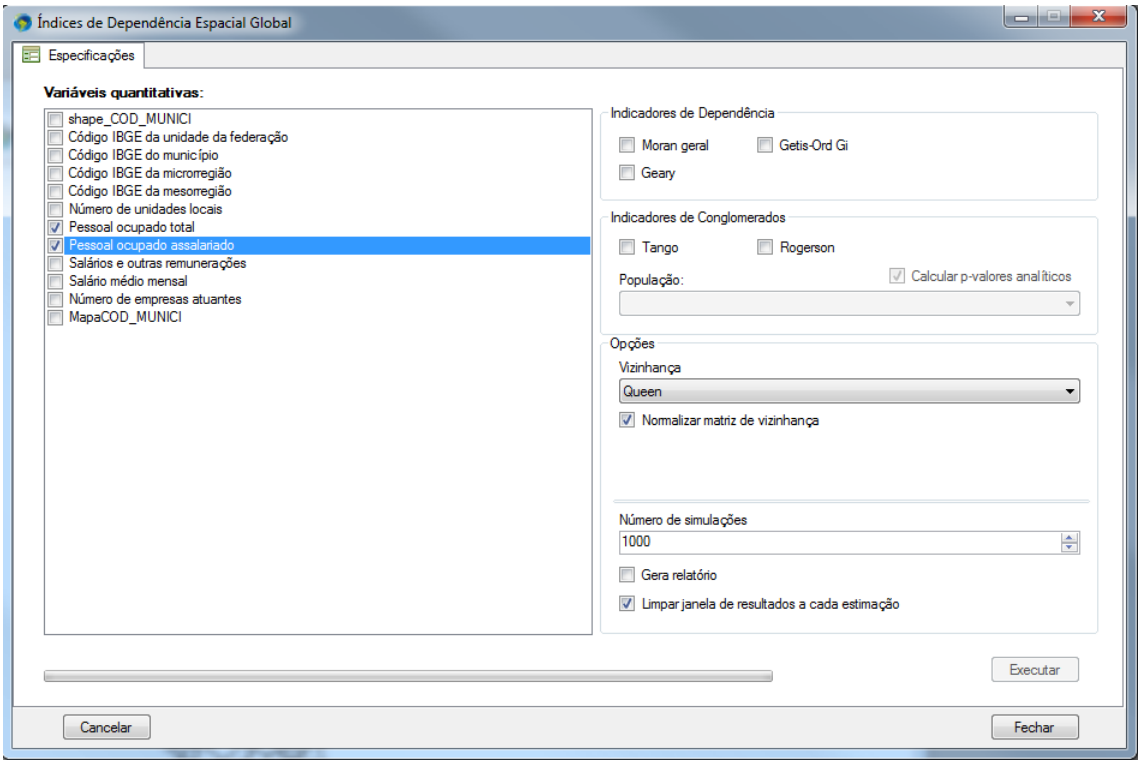
Também é possível escolher o número de simulações de Monte Carlo.

Exemplo

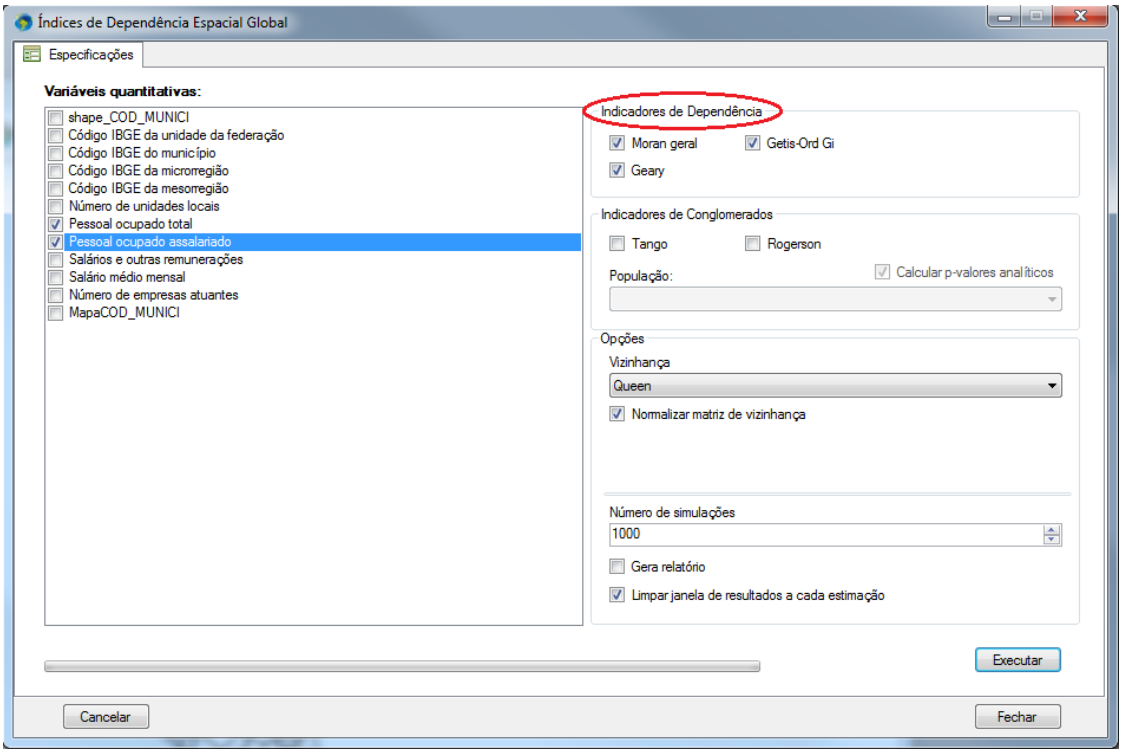
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , opção “Dependência Espacial” e, em seguida, na opção “Índices Globais”.



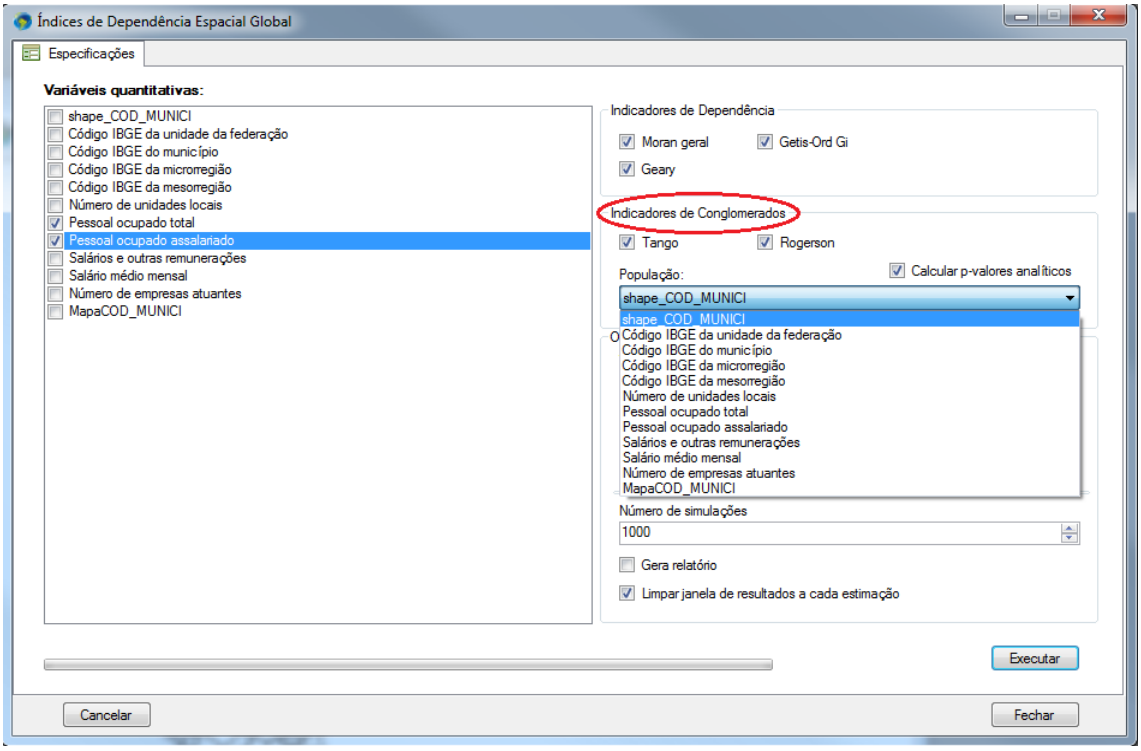
O seguinte formulário irá se abrir:



Na aba “Especificações” em “Variáveis Quantitativas”, selecione as variáveis desejadas para a análise. Após selecionar as varáveis escolha os “Indicadores de Dependência” que podem ser: “Moran geral”, “Getis-Ord Gi” ou “Geary".

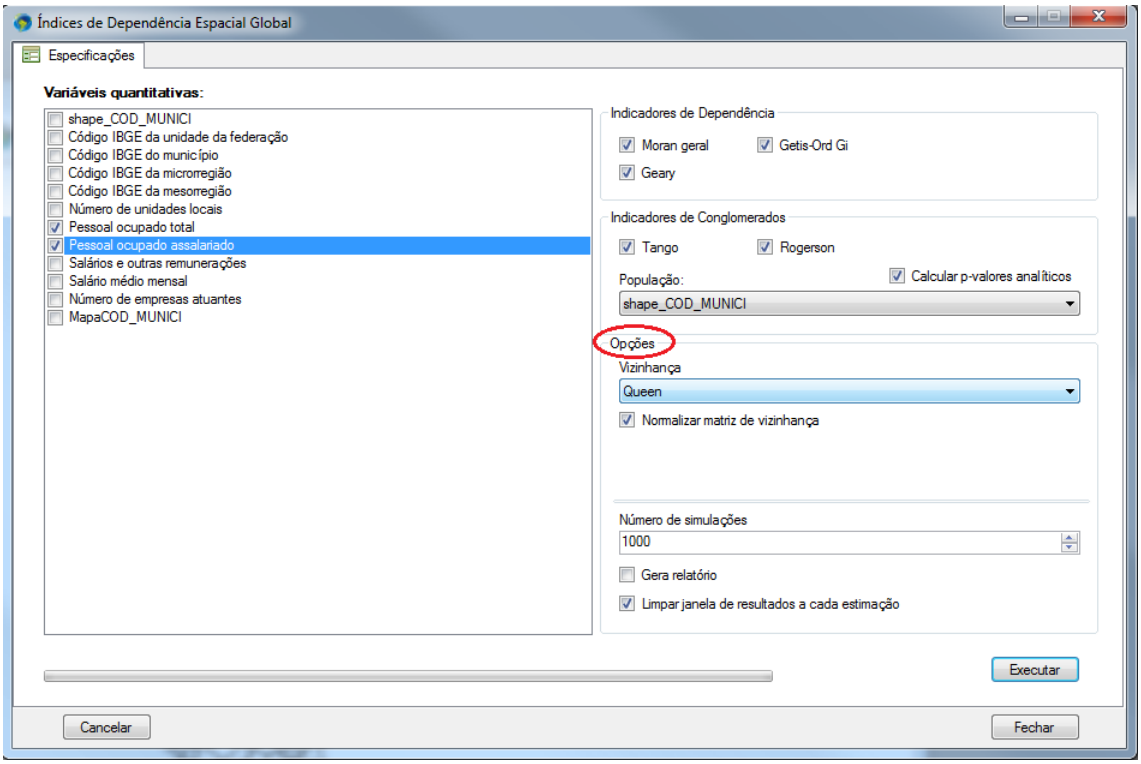


Em “Indicadores de Conglomerados” existem as opções “Tango” e “Rogerson”. Após escolher os indicadores desejados, deve ser selecionada a população de interesse.

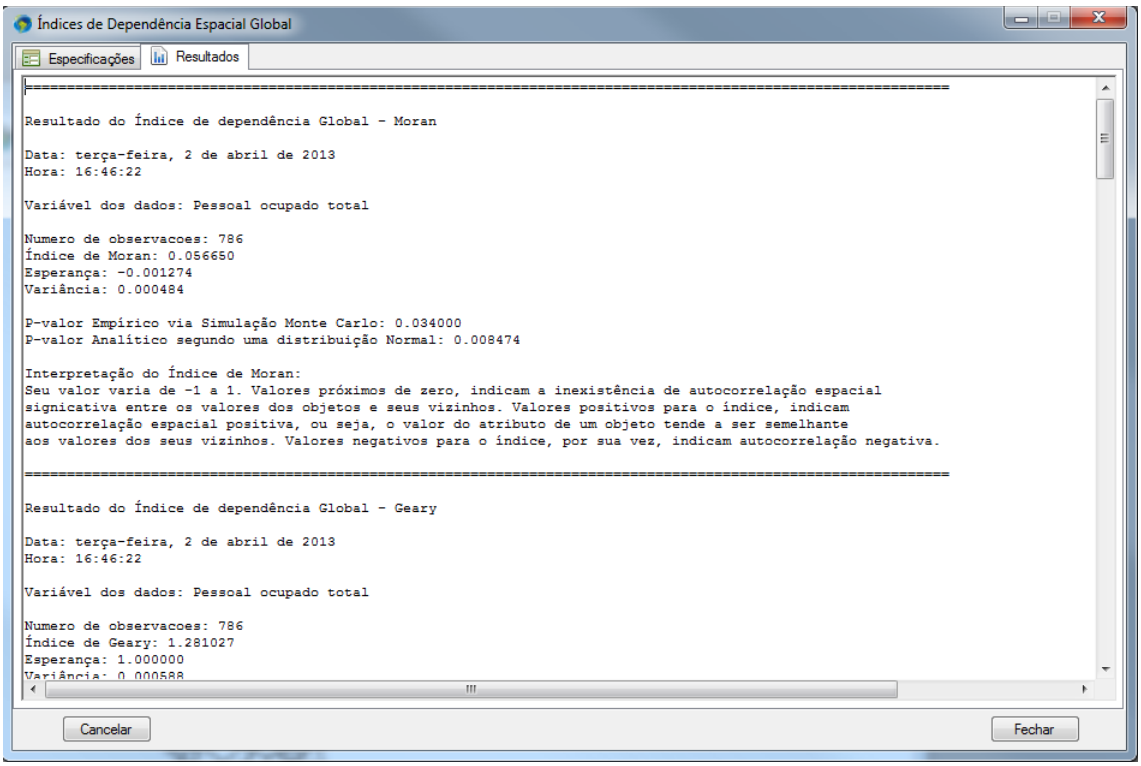


Em “Opções”, selecione as opções para cálculo da matriz de vizinhança “Queen” ou “Rook”. Também é possível escolher o número de simulações de Monte Carlo

desejadas.



Depois de selecionadas as opções de interesse, clique no botão “Executar”. Uma nova aba chamada “Resultados” aparecerá contendo os resultados de acordo com as opções selecionadas anteriormente.



Índices Locais

Visão Geral

Quando muitos polígonos são analisados é provável que ocorra diferentes regimes de autocorrelação espacial na região em estudo. Nesses casos, apenas os índices globais não seriam suficientes para entender a estrutura dos dados analisados. Portanto, é importante apresentar também, para cada polígono, uma medida de autocorrelação espacial. Essas estatísticas locais, em geral, derivam de algum dos índices de dependência global apresentados no módulo de Índices Globais.

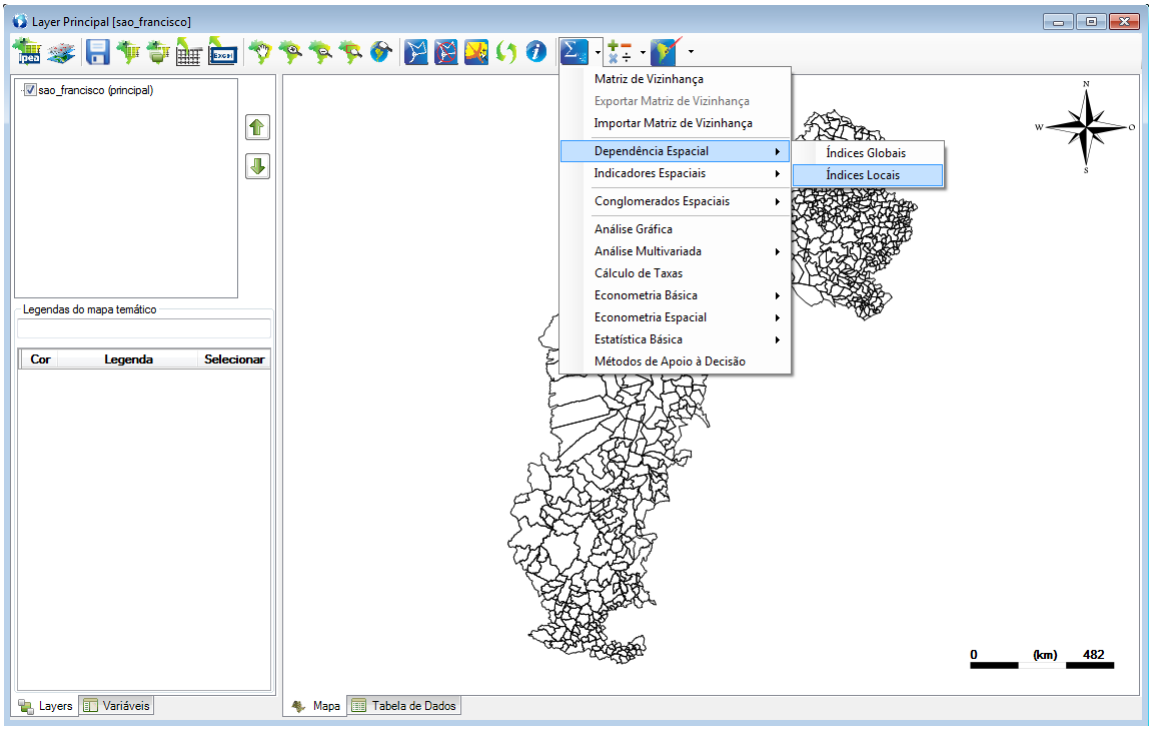
As opções de Indicadores para os usuários do IpeaGEO são:

- Indicadores de Dependência: LISA; Getis-Ord Gi*; e Getis-Ord Gi.
- Indicadores de Conglomerados: Escore.

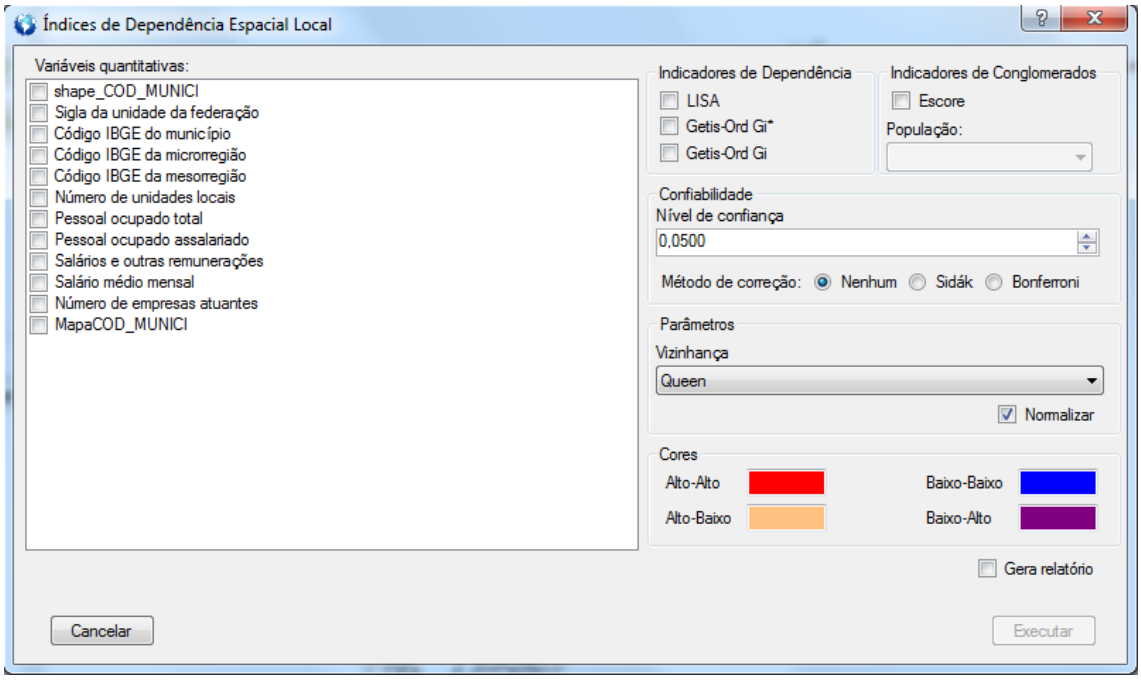
Também é possível escolher o nivel de confiança desejado, as cores a serem plotadas no mapa, o tipo de vizinhança e os seguintes métodos de correção: nenhum; Sidák; ou Bonferroni.

Exemplo

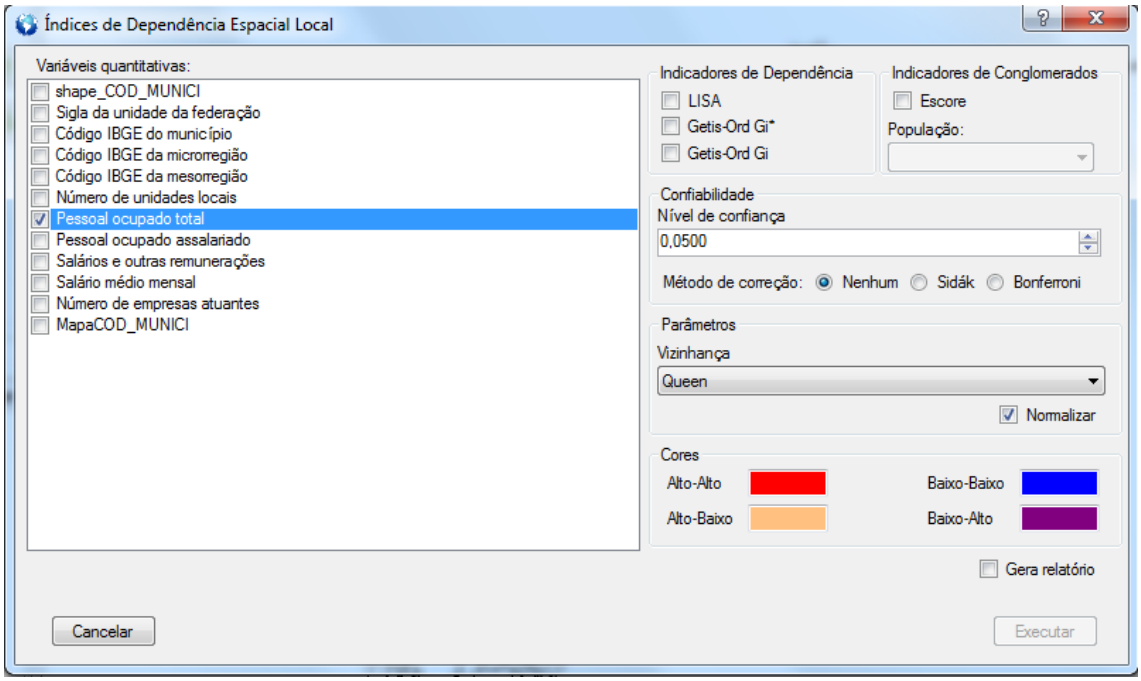
Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique no ícone “Ferramentas” , opção “Dependência Espacial” e, em seguida, na opção “Índices Locais”.



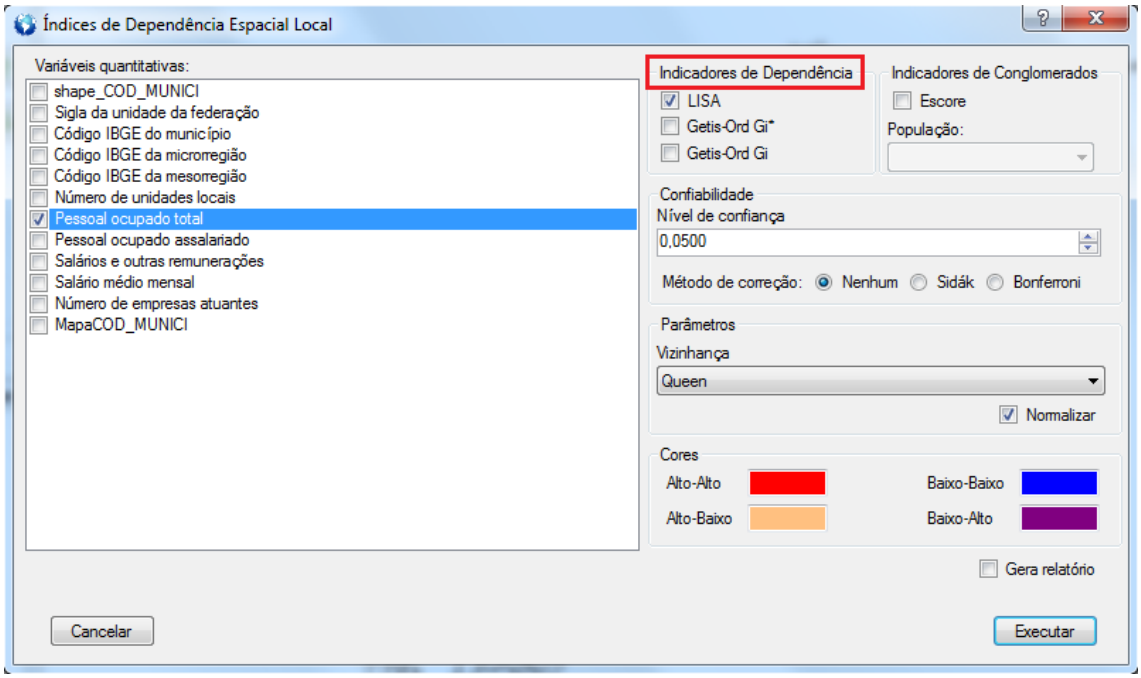
O seguinte formulário irá se abrir:



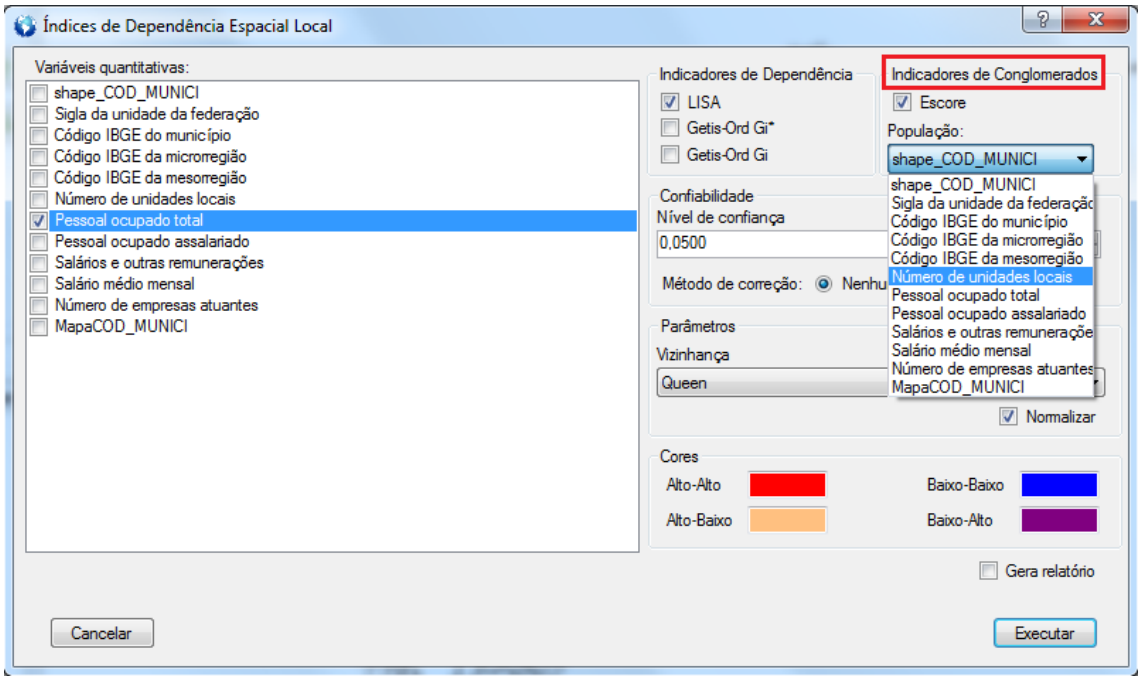
Em “Variáveis Quantitativas”, selecione as variáveis desejadas para a análise.



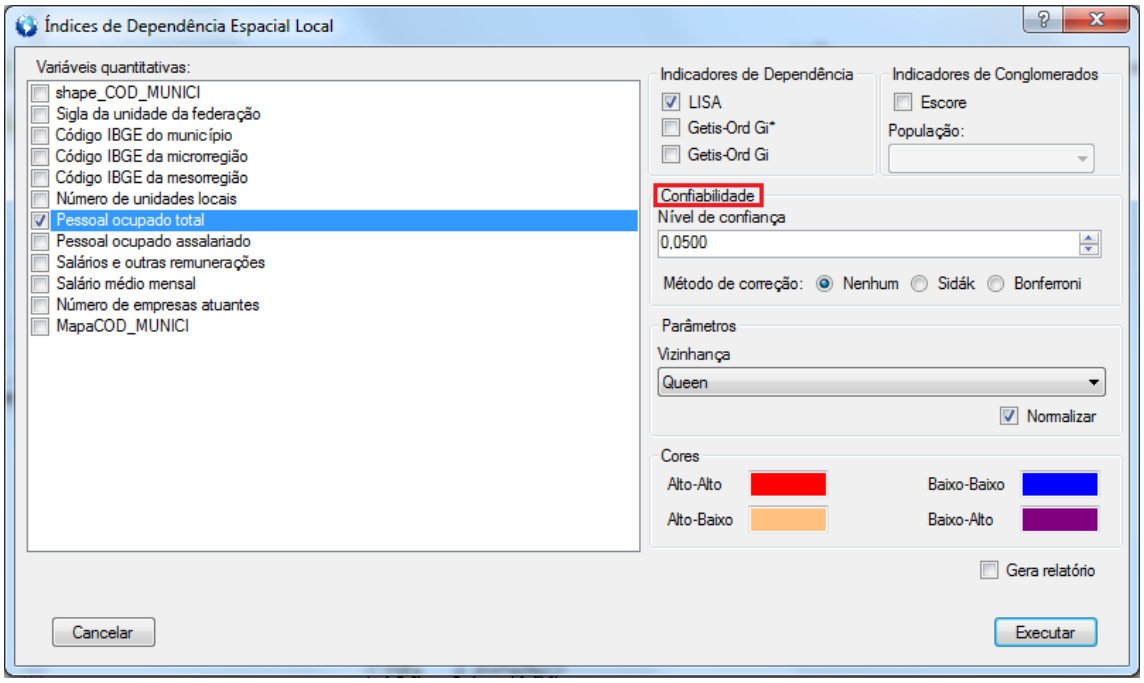
Depois de selecionadas as varáveis desejadas, selecione os “Indicadores de Dependência” que podem ser: “LISA”, “Getis-Ord Gi*” ou “Getis-Ord Gi”.



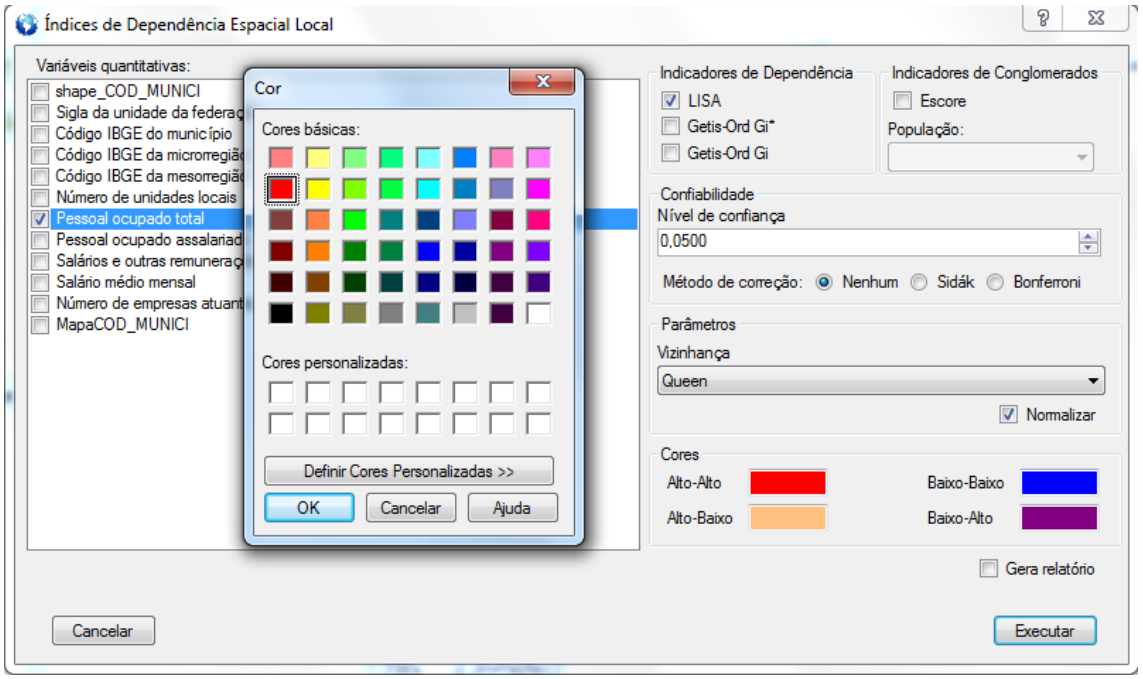
Em “Indicadores de Conglomerados” existe a opção “Escore”. Após escolher os indicadores desejados, deve ser escolhida a variável correspondente à “População” da variável de interesse.



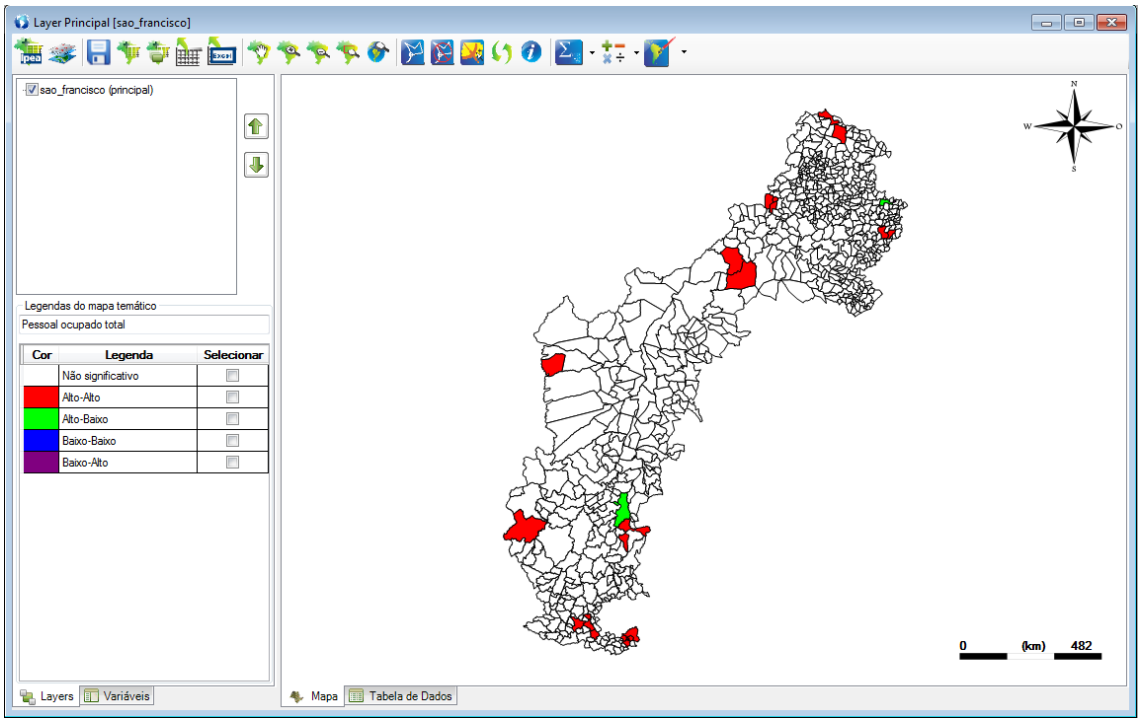
Em “Confiabilidade”, selecione o “Nível de confiança” e o “Método de correção” (Nenhum, Sidák ou Bonferroni) de interesse.



Também estão disponíveis opções de cores a serem utilizadas no mapa temático, clicando nos retângulos coloridos é possível alterar as cores escolhidas.



Depois de selecionadas as opções de interesse, clique no botão “Executar”. O novo mapa gerado aparecerá contendo os resultados de acordo com as opções selecionadas anteriormente.



Econometria Espacial

Cross-Section

Visão Geral

O formulário de dados *cross-section* espaciais tratam modelos de regressão para observações, em um momento do tempo, referentes a um conjunto de unidades geográficas. O objetivo principal desses modelos é incorporar a possível presença de correlações espaciais entre as unidades geográficas. Para isso, existem diversas opções disponíveis para representação da dependência espacial entre as unidades.

Além das diferentes opções para determinação da estrutura de dependência espacial entre os vizinhos, é possível optar por diferentes métodos de estimação para os modelos de regressão disponíveis. Essas opções são aplicáveis para diferentes situações, dependendo, por exemplo, do tipo de informações disponíveis. As opções de modelagem disponíveis no formulário de regressão para dados *cross-section* espaciais são:

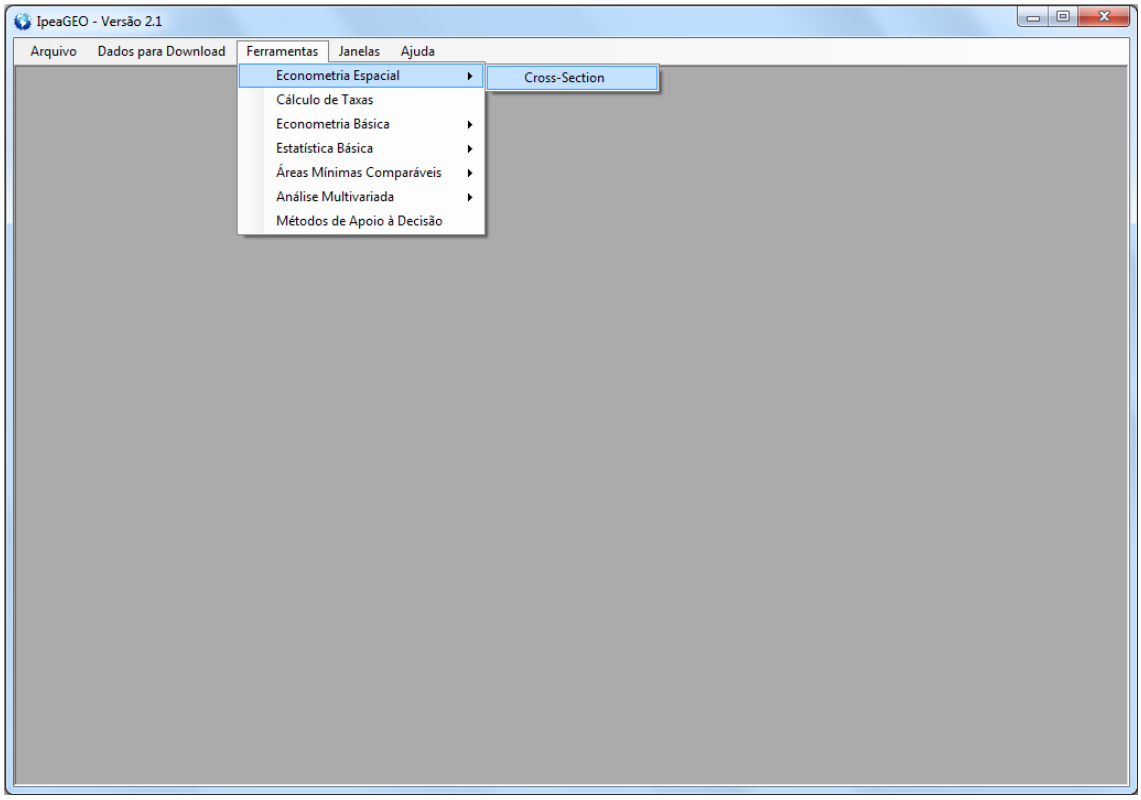
- Estimador de mínimos quadrados ordinários (OLS);
- Estimador de mínimos quadrados de dois estágios (2SLS);
- Modelos de regressão com dependência espacial estimados via máxima verossimilhança;
- Método de momentos generalizado com correção para dependência espacial (GMM Espacial).

Exemplo

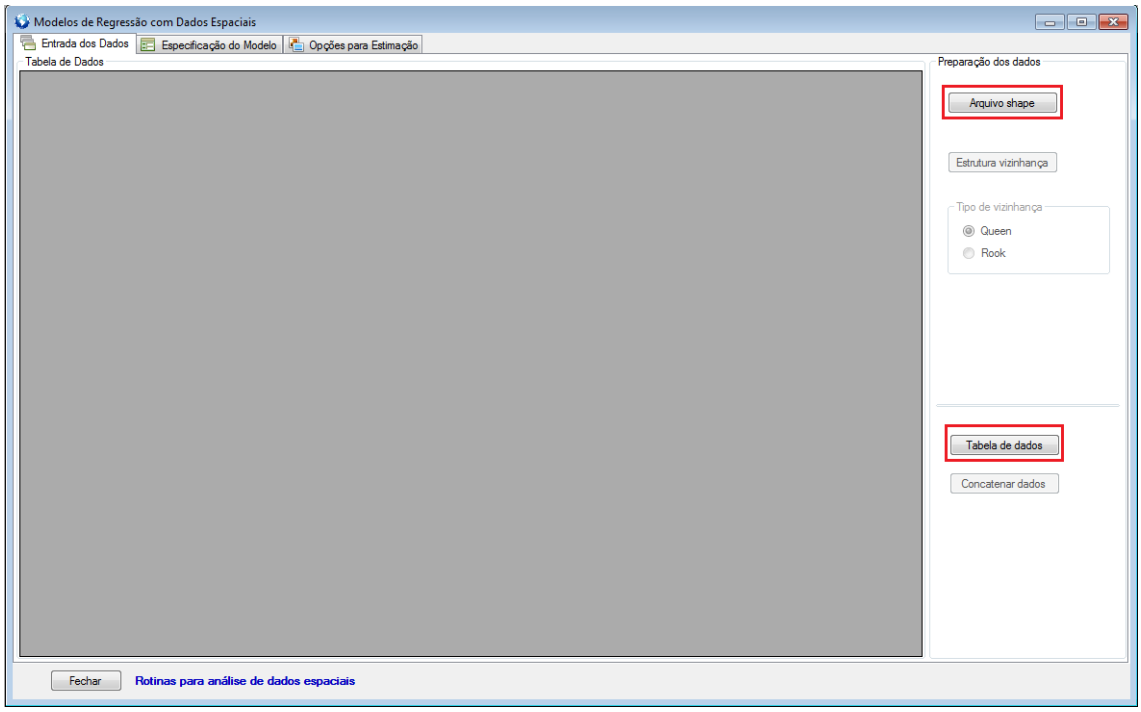
Existem dois modos para realizar essa análise no IpeaGEO:

- Direto do menu principal:

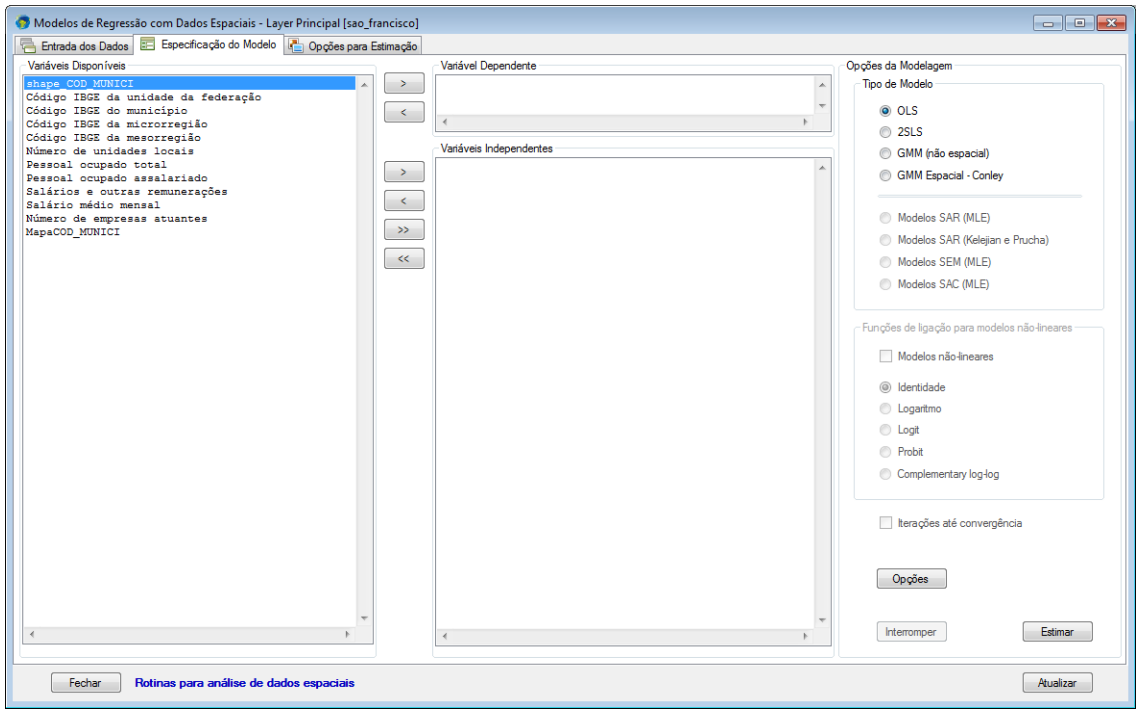
No ícone "Ferramentas", clique na opção “Econometria Espacial” e, em seguida, na opção “Cross-Section”.



Abrir-se-á uma aba chamada “Entrada de Dados”, na qual será necessário importar um arquivo de dados e um arquivo shape. Para isto, clique em "Tabela de dados" e em “Arquivo shape” respectivamente.



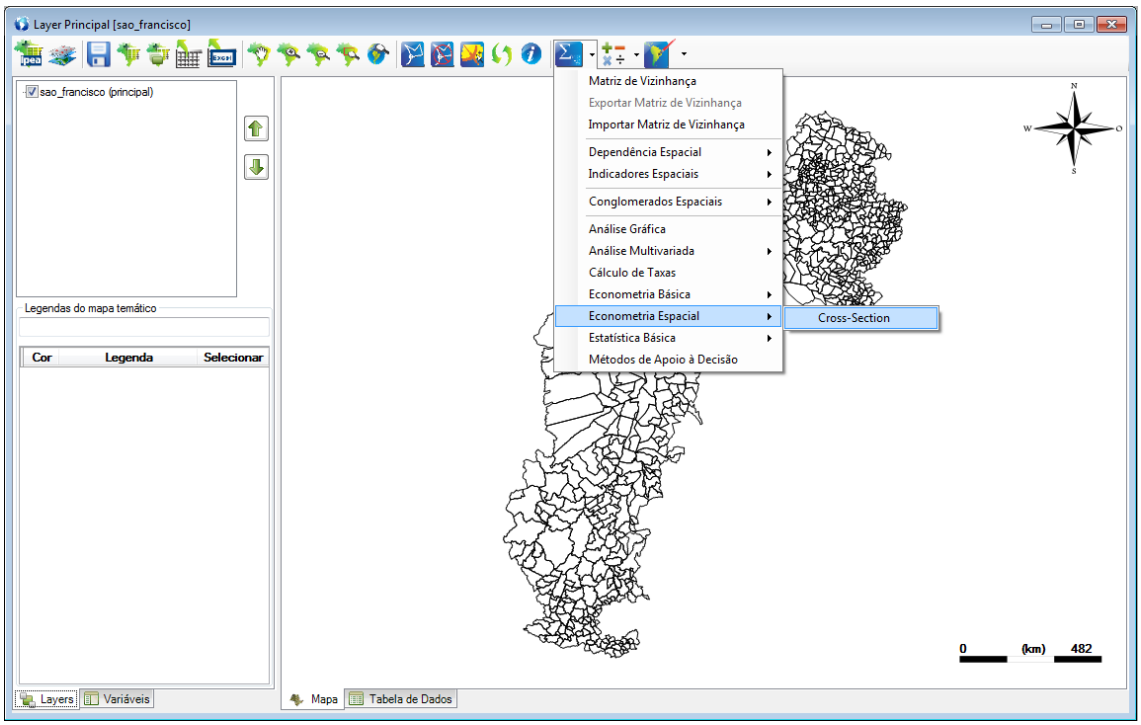
Depois de concatenados a tabela e a malha digital o seguinte formulário, localizado na aba “Especificação do Modelo”, se abrirá:



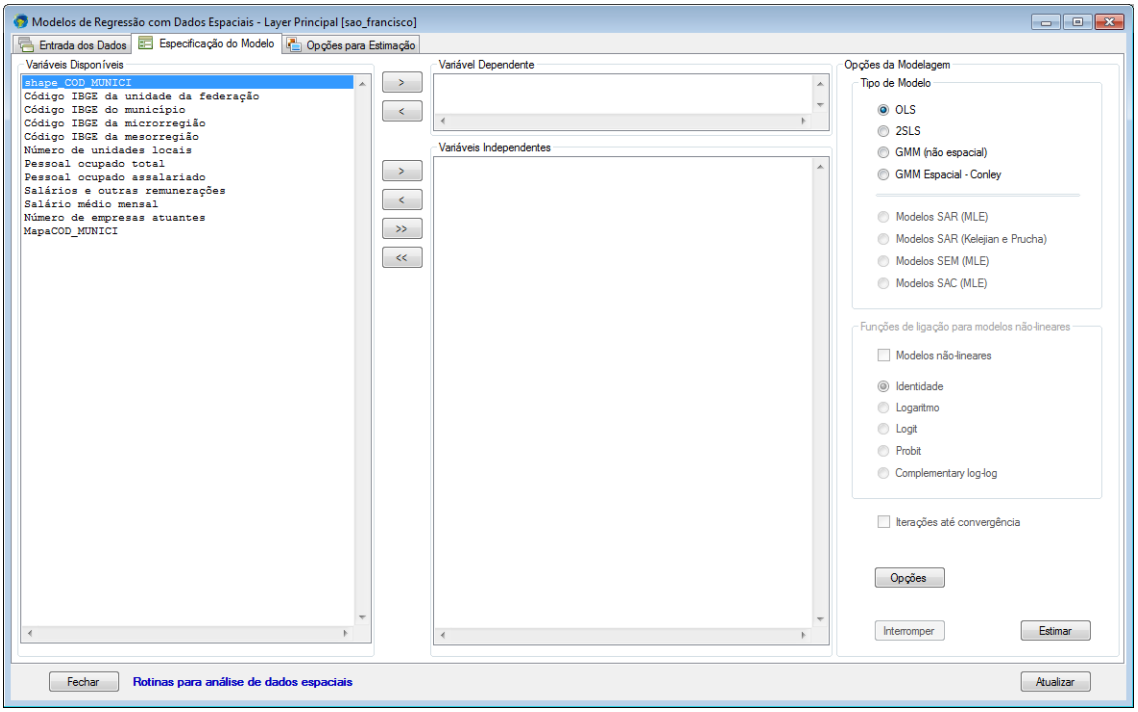
- A partir da importação de uma malha digital (shape):

Após importar a tabela de dados no IpeaGEO (ver tópico Importação de dados) ou concatenar o arquivo de malha (shape) com a tabela de dados (ver tópico de

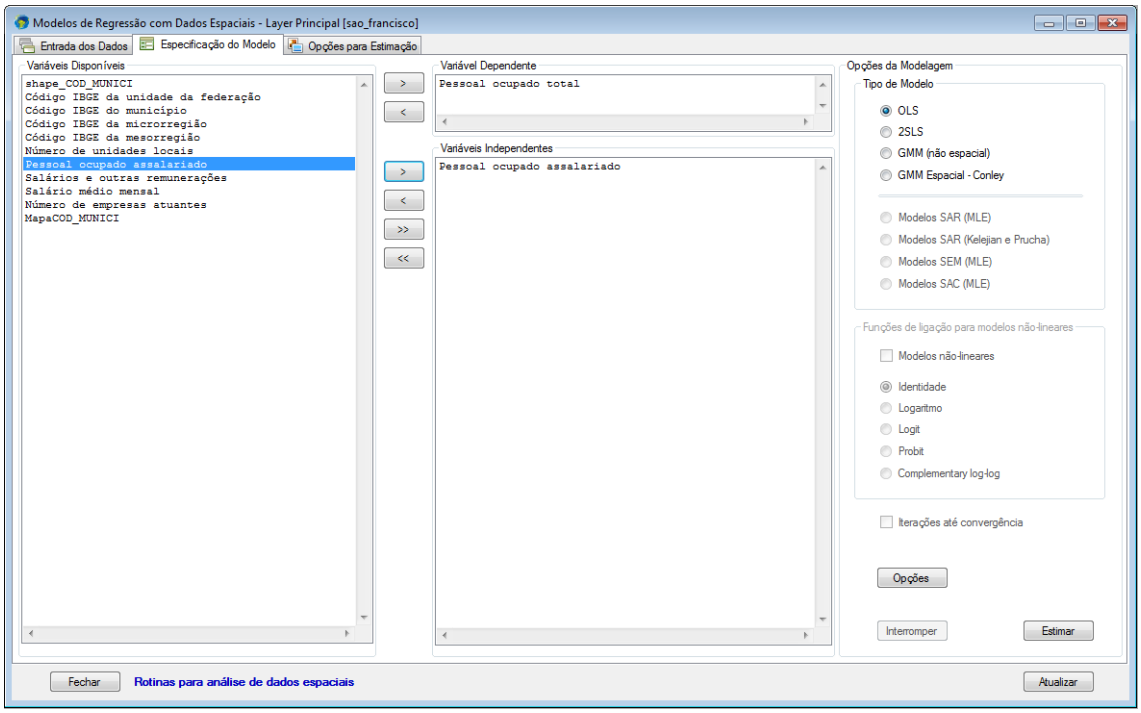
concatenação), clique no ícone “Ferramentas” , opção “Econometria Espacial” e, em seguida, na opção “Cross-Section”.



O seguinte formulário irá se abrir:

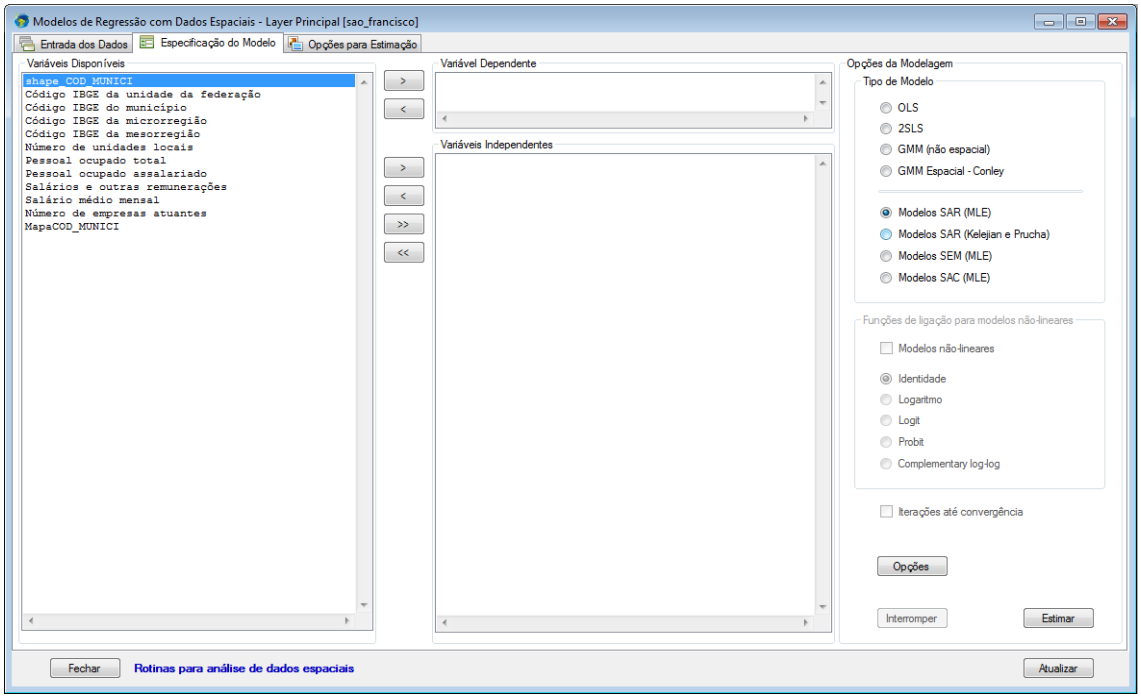


Em “Variáveis Disponíveis”, selecione as variáveis que serão utilizadas como dependentes e independentes na análise.



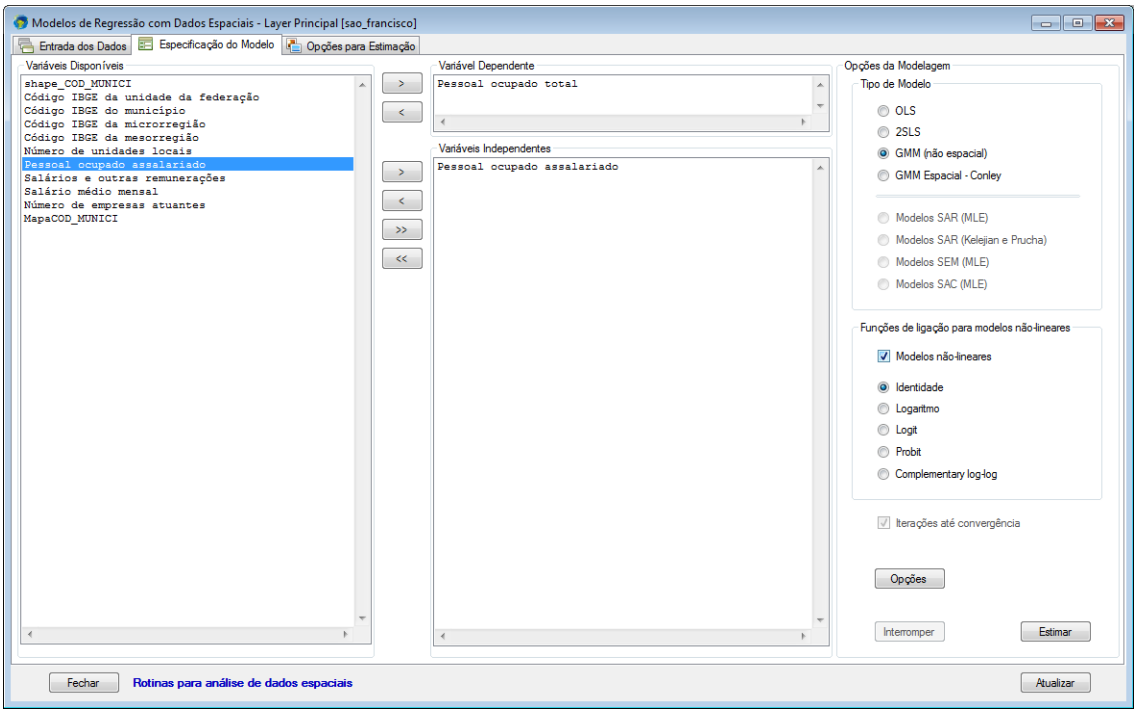
Selecionadas as variáveis de interesse, escolha um dos “Tipos de Modelo”: mínimos quadrados ordinários (OLS); 2SLS; GMM (não espacial); e GMM Espacial – Conley.

Se for realizada a atualização da matriz de vizinhança (ver tópico Matriz de Vizinhança), estarão disponíveis os seguintes “Tipos de Modelo”: SAR (MLE); SAR (Kelejian e Prucha); SEM (MLE); e SAC (MLE).

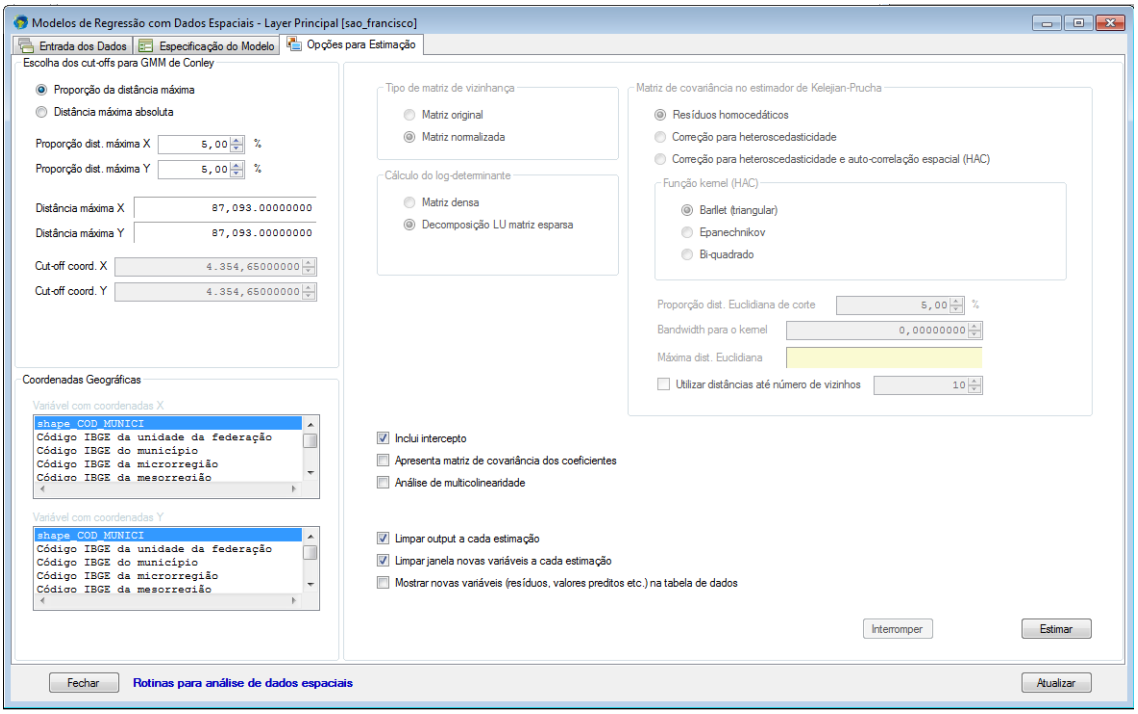


Caso seja selecionada a opção “GMM (não espacial)” ou “GMM Espacial –

Conley", a parte de "Funções de ligação para modelos não lineares" se torna disponível para que o usuário selecione uma das seguintes opções: Identidade; Logaritmo; Logit; Probit; ou Complemento log-log (Complementary log-log).



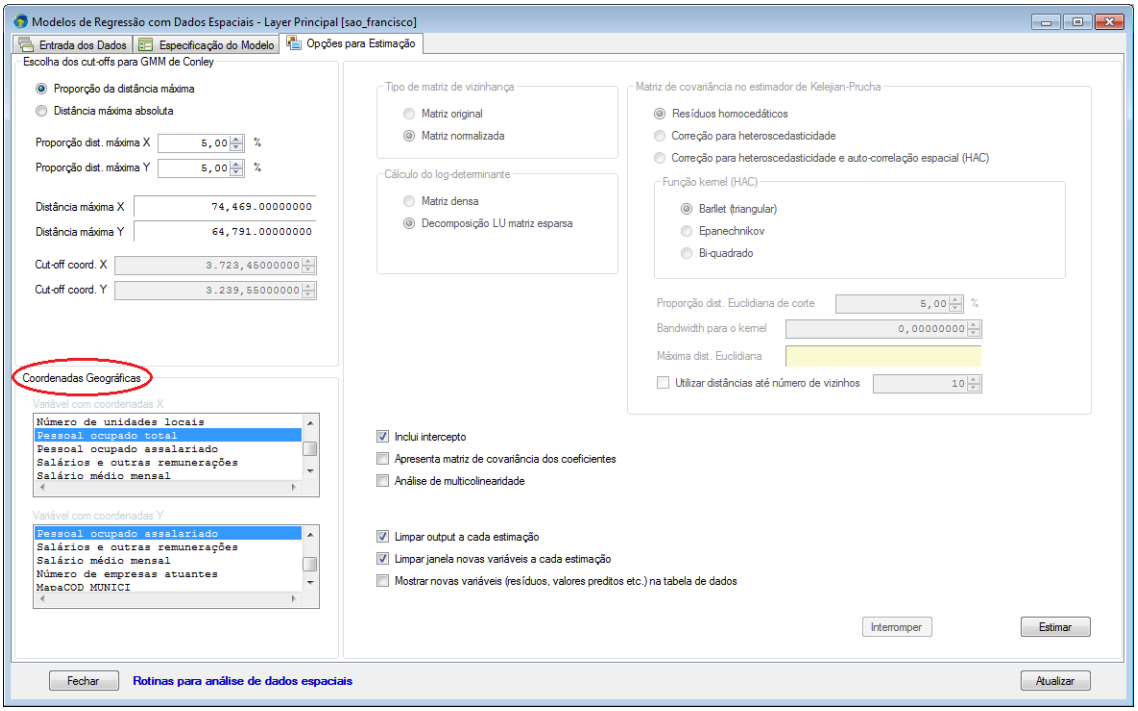
Ao selecionar a opção "GMM Espacial – Conley" se abrirá a aba "Opções para Estimação". Esta aba também pode ser acessada clicando no botão "Opções".



As opções para a escolha dos pontos de corte (cut-offs) são:

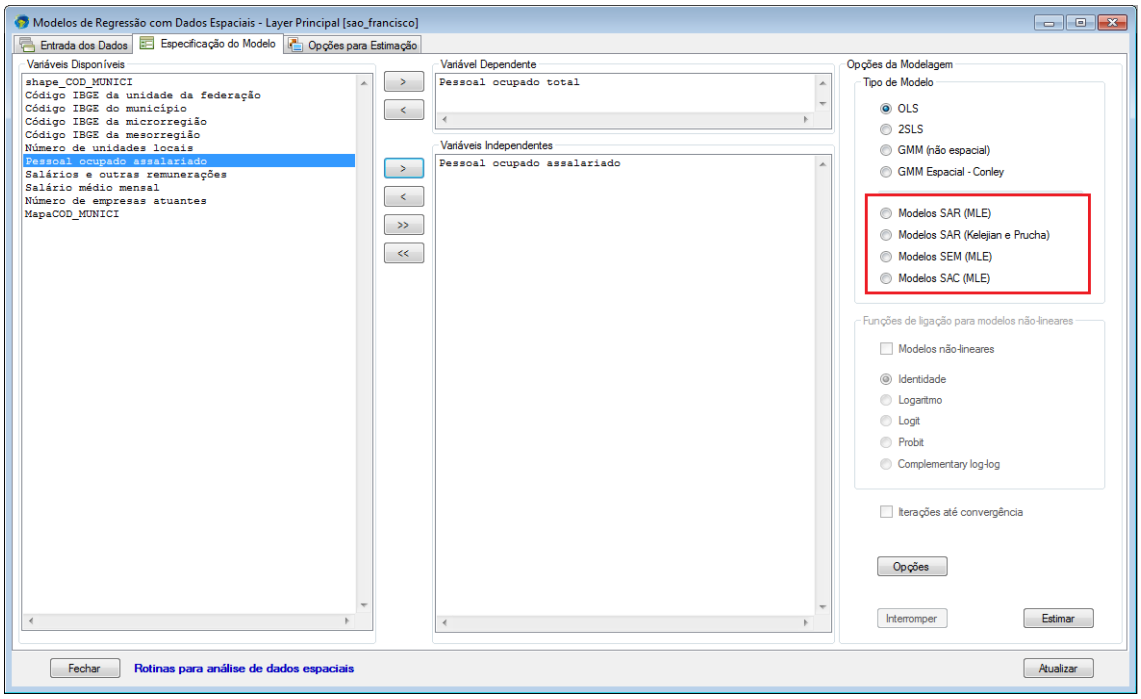
- Proporção da distância máxima: X e Y.
- Distância máxima absoluta: cut-off para as coordenadas X e Y.

Em “Coordenadas Geográficas” é possível escolher as variáveis para as coordenadas X e Y.

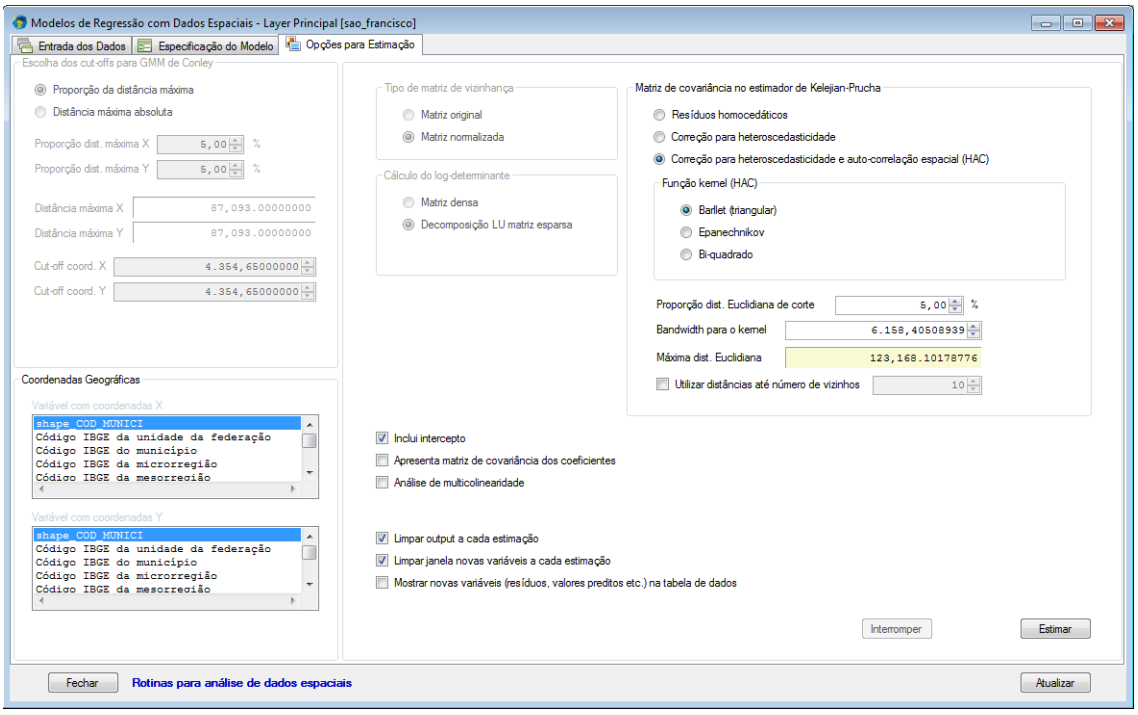


Para tornar as opções “Tipo de matriz de vizinhança”, “Cálculo do log-determinante”, “Matriz de covariância no estimador de Kelejian-Prucha” e “Função kernel (HAC)” habilitadas, é necessário, inicialmente, determinar a matriz de vizinhança desejada (ver tópico Matriz de Vizinhança).

Após escolher a matriz de vizinhança desejada, o usuário tem a disposição novas opções de tipos de modelos, são elas: SAR (MLE); SAR (Kelejian e Prucha); SEM (MLE); e SAC (MLE).

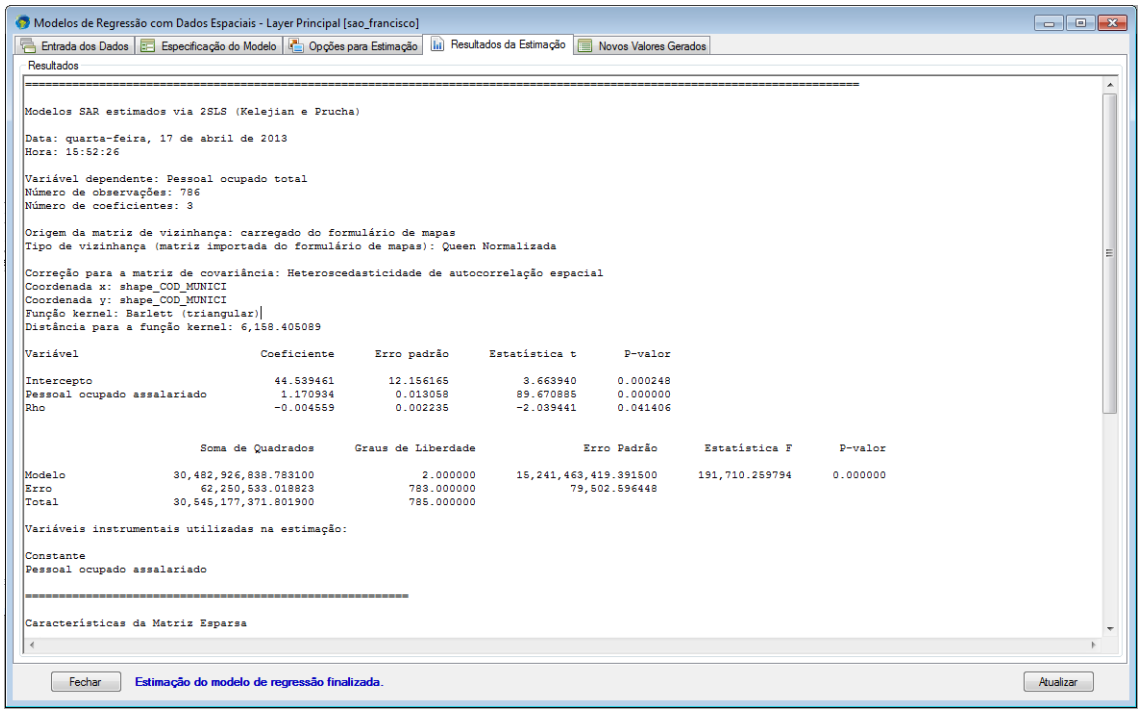


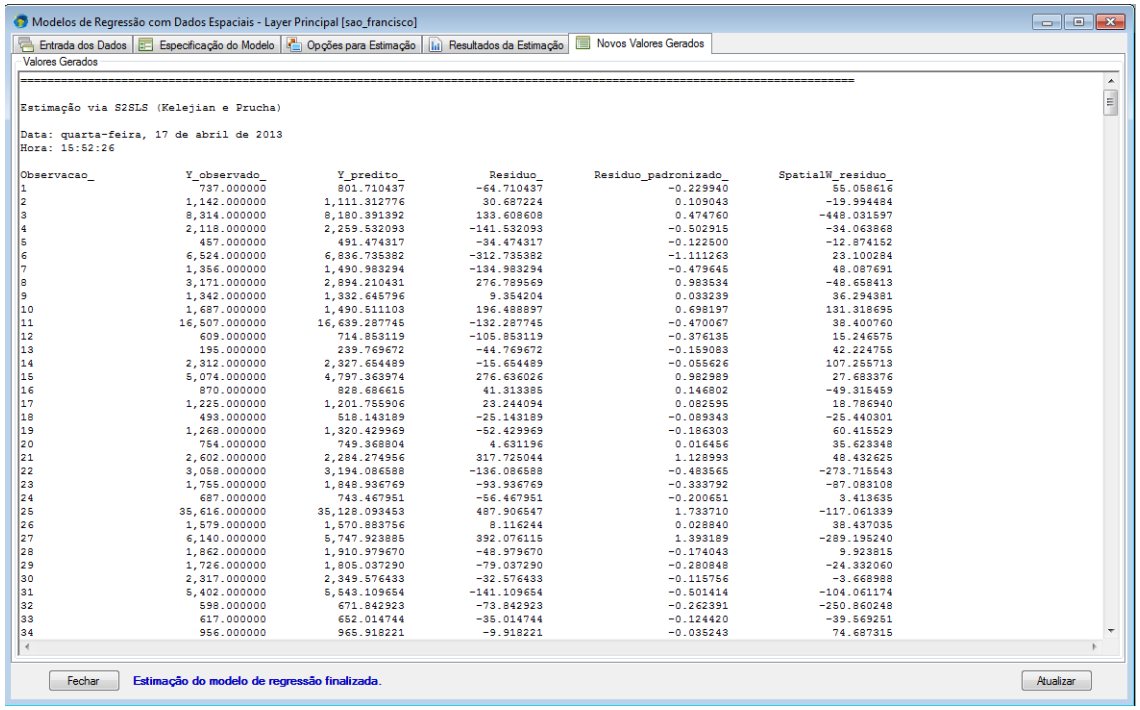
Ao escolher o modelo de interesse clique no botão “Opções” e, de acordo com o modelo selecionado, novas opções correspondentes ao modelo selecionado estarão disponíveis. Por exemplo, para os modelos SAR (Kelejian e Prucha) se observa:



Após selecionar o modelo desejado existe a possibilidade de escolher uma das opções: incluir intercepto; apresentar matriz de covariância dos coeficientes; análise de multicolinearidade; limpar saída de resultados (output) a cada estimação; limpar janela novas variáveis a cada estimação; e manter novas variáveis (resíduos, valores etc.) na tabela de dados.

Para estimar os modelos clique em “Estimar”. Duas novas abas se abrirão: a primeira contendo os “Resultados da Estimação” e a segunda os “Novos Valores Gerados”.





O botão “Atualizar” atualiza a tabela de dados no formulário de mapas.

Indicadores de Segregação Espacial

Visão Geral

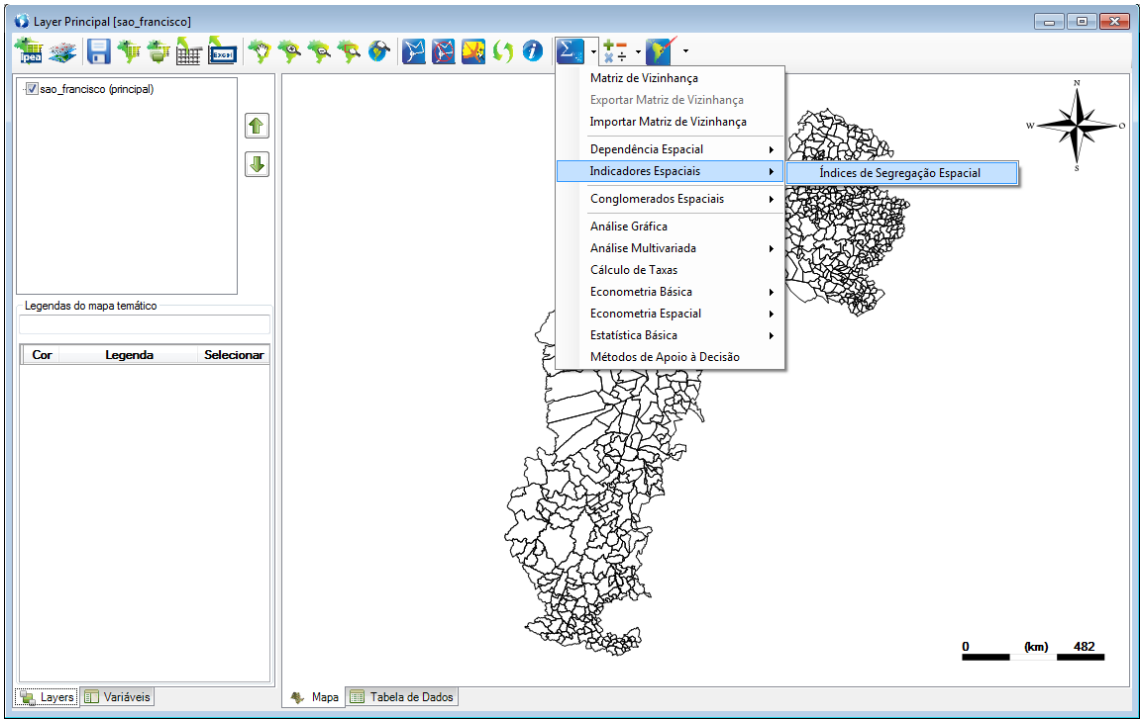
Os índices de segregação espacial medem o quão irregular é a distribuição de certo grupo em uma determinada área geográfica. Os primeiros estudos dos índices de segregação surgiram na década de 1950 com o estudo de Ducan e Ducan (1955a). Segundo eles, o problema das medidas de segregação foi a construção delas a partir de noções simples e pobres ao invés de serem formuladas por meio de articulações conceituais claras de seu processo. Ducan e Ducan (1955a, p.217) caracterizavam essa simplicidade conceitual como uma “ingênua” noção de segregação.

[Segregação] é um conceito rico em sugestividade teórica e de um valor heurístico inquestionável. Claramente nós não desejamos sacrificar o capital da teorização e observação já investidas no conceito. Até o momento, isso é o que está envolvido na solução oferecida pelo operacionalismo “ingênuo”, mais ou menos uma combinação arbitrária de alguns convenientes procedimentos numéricos de um conceito verbal de segregação.

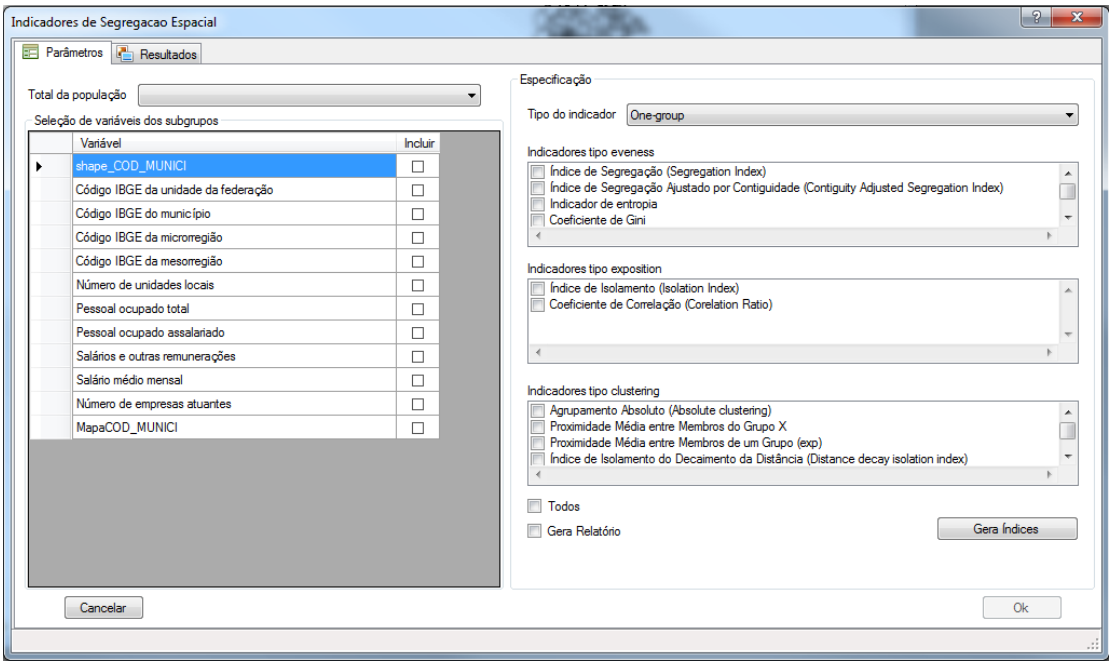
Segundo Wong (1999), a medida segregação de Duncan e Duncan (1955a) é ainda a melhor medida de segregação, pois ela capta efetivamente a dimensão *evenness*, a mais importante dimensão de segregação.

Exemplo

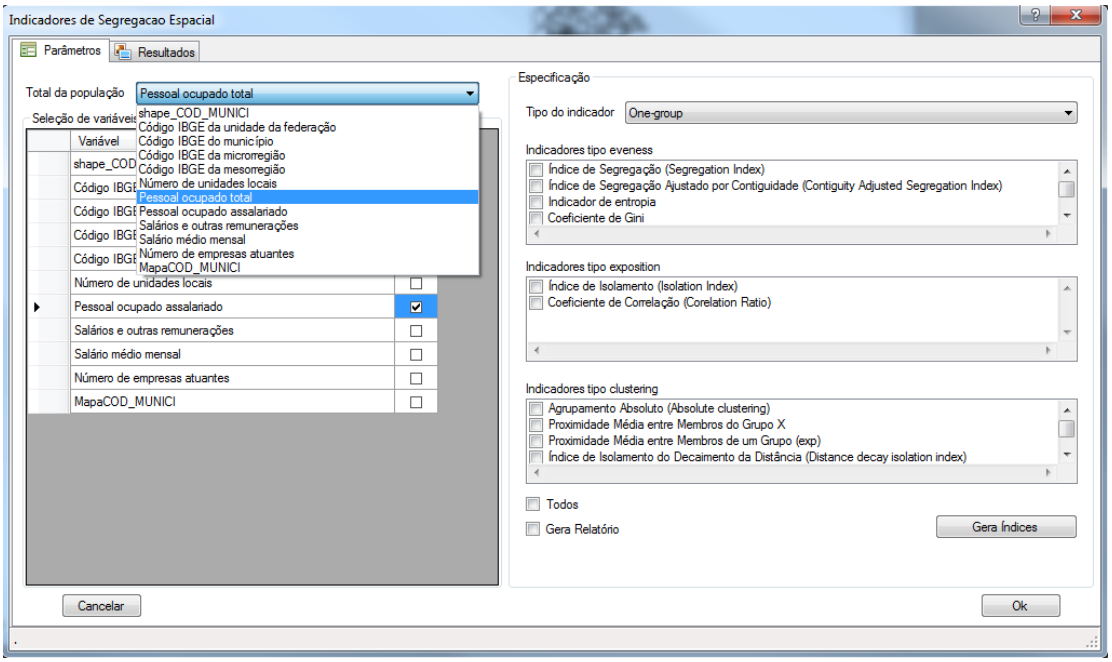
Após concatenar o arquivo de malha digital (shape) com a tabela de dados (ver tópico de concatenação), clique no botão “Ferramentas” , opção “Indicadores Espaciais” e, em seguida, na opção “Índices de Segregação Espacial”.



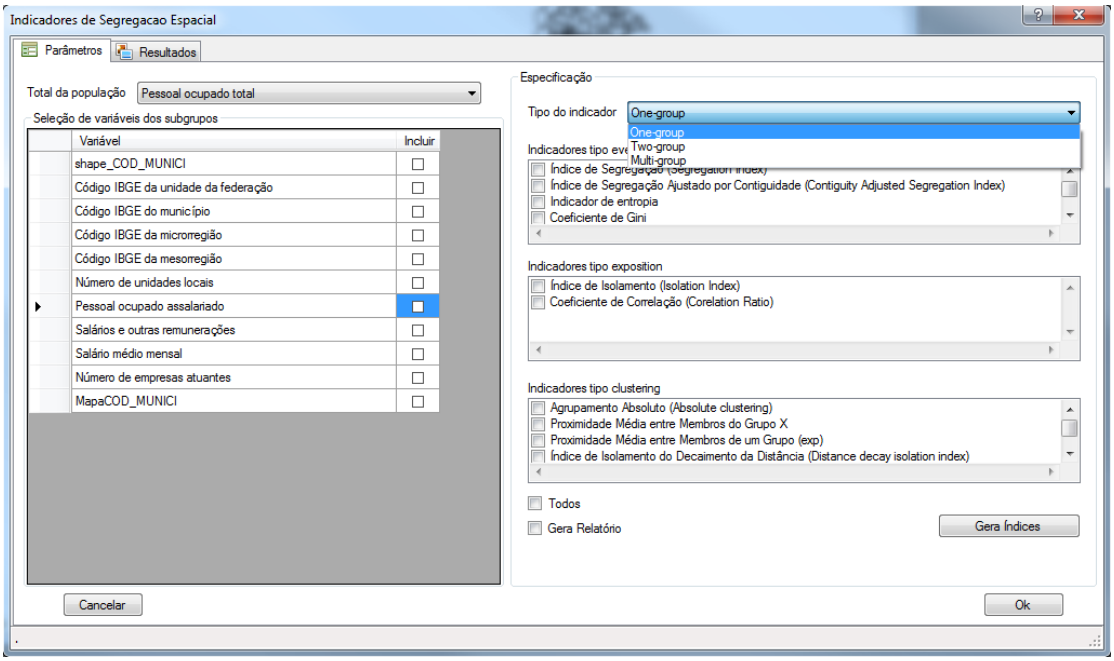
O seguinte formulário irá se abrir:



No IpeaGEO, é necessário escolher inicialmente o “Total da população” – que será o total da variável de interesse:



Escolhido o total da população e selecionadas as variáveis dos subgrupos de interesse, escolha um dentre os tipos de indicadores espaciais: único grupo (One-group); dois grupos (Two-group); ou múltiplos grupos (Multi-group).



Escolhido o um único grupo (One-Group) são disponibilizados os seguintes indicadores:

Tipo eveness:

- Índice de Segregação (Segreration índex);
- Índice de Segregação Ajustado por Contiguidade (Segregation índex adjusted for contiguity);
- Indicador de Entropia;
- Coeficiente de Gini;
- Índice de Atkinson.

Tipo exposition:

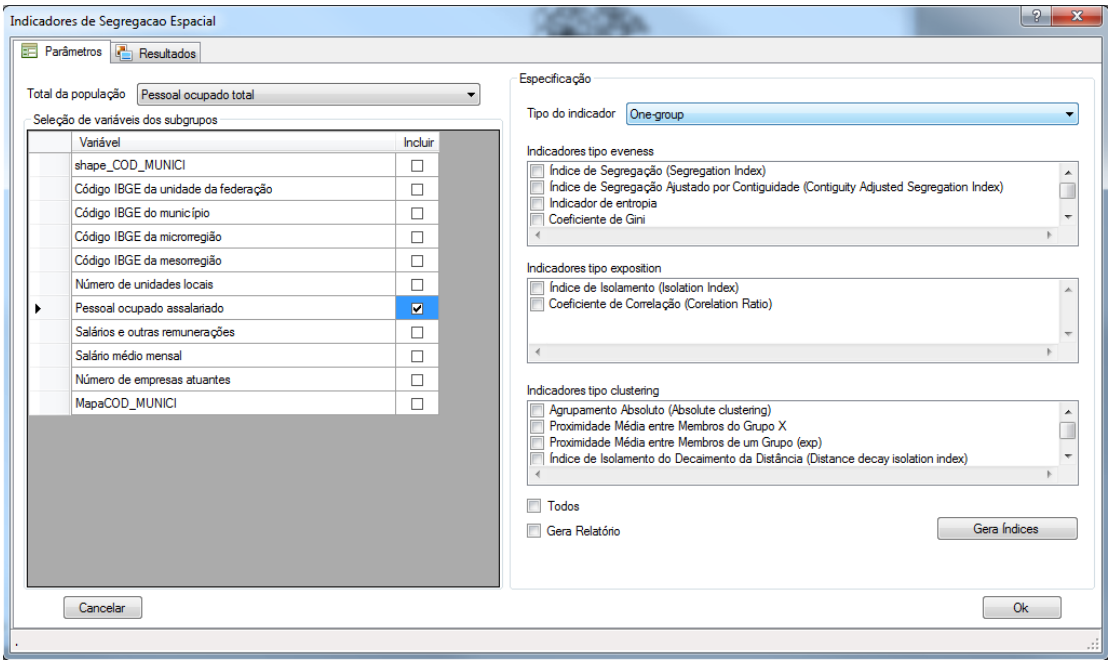
- Índice de Isolamento (Isolation índex);
- Coeficiente de Correlação (Correlation ratio).

Tipo clustering:

- Agrupamento Absoluto (Absolute clustering);
- Proximidade Média entre Membros do Grupo X (Mean proximity between members of group X);

- Proximidade Média entre Membros de um Grupo (exp) (Mean proximity between members of one group (exp));

- Índice de Isolamento do Decaimento da Distância (Distance decay isolation index).



Escolhido o tipo dois grupos (Two-Group) são disponibilizados os seguintes indicadores:

Tipo eveness:

- Índice de Dissimilaridade entre XY (Índex of dissimitarity between XY);

- Dissimilaridade Ajustada para Contiguidade entre XY (Dissimitarity adjusted for contiguity between XY);

Tipo exposition:

- Índice de Interação (Interation índex).

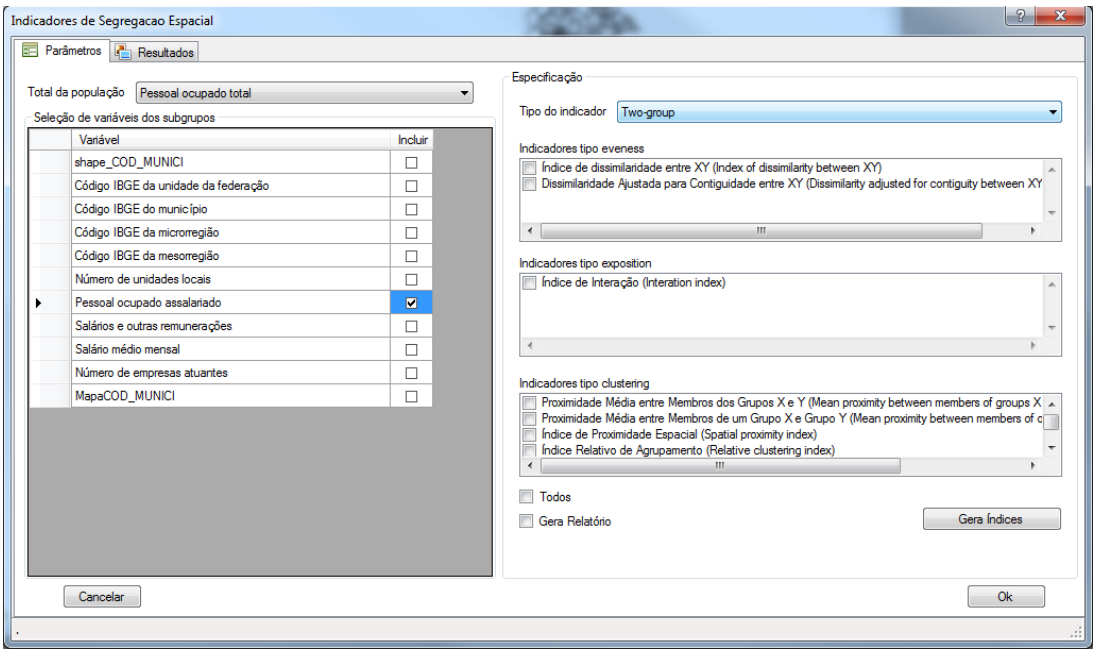
Tipo clustering:

- Proximidade Média entre Membros dos Grupos X e Y (Mean proximity between members of groups X and Y);

- Proximidade Média entre Membros de um Grupo X e Grupo Y (Mean proximity between members of one group X and group Y);

- Índice de Proximidade Espacial (Spatial proximity índex);

- Índice Relativo de Agrupamento (Relative clustering índex);
- Índice de Interação de Decaimento de Interação (Distance decay interation índex).



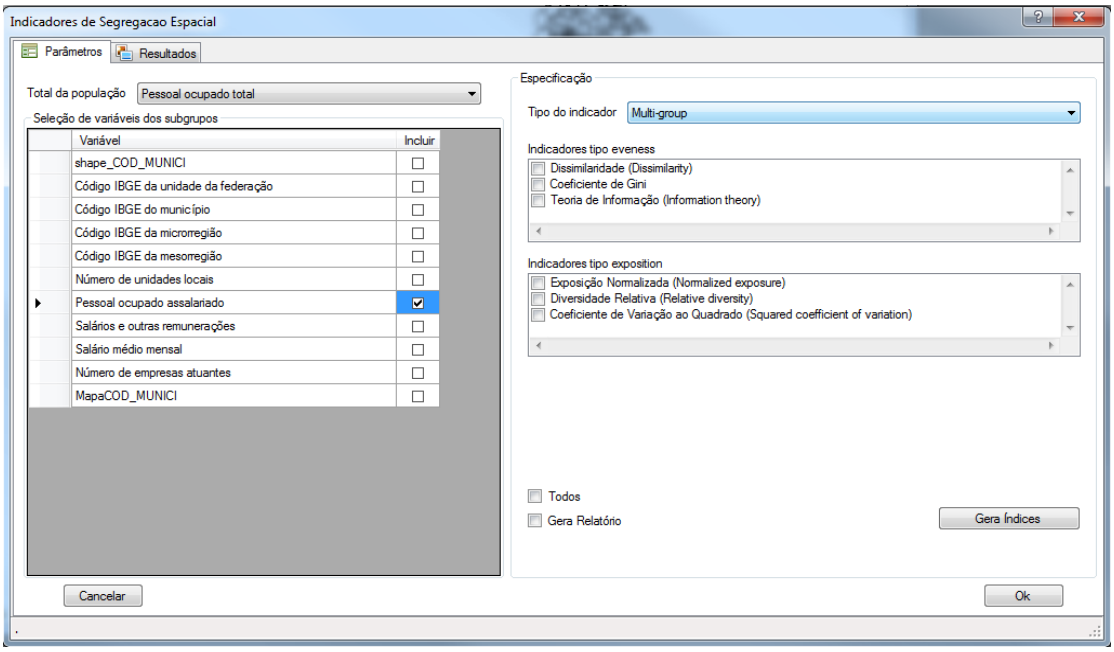
Escolhido o tipo múltiplos grupos (Multi-Group) são disponibilizados os seguintes indicadores:

Tipo eveness:

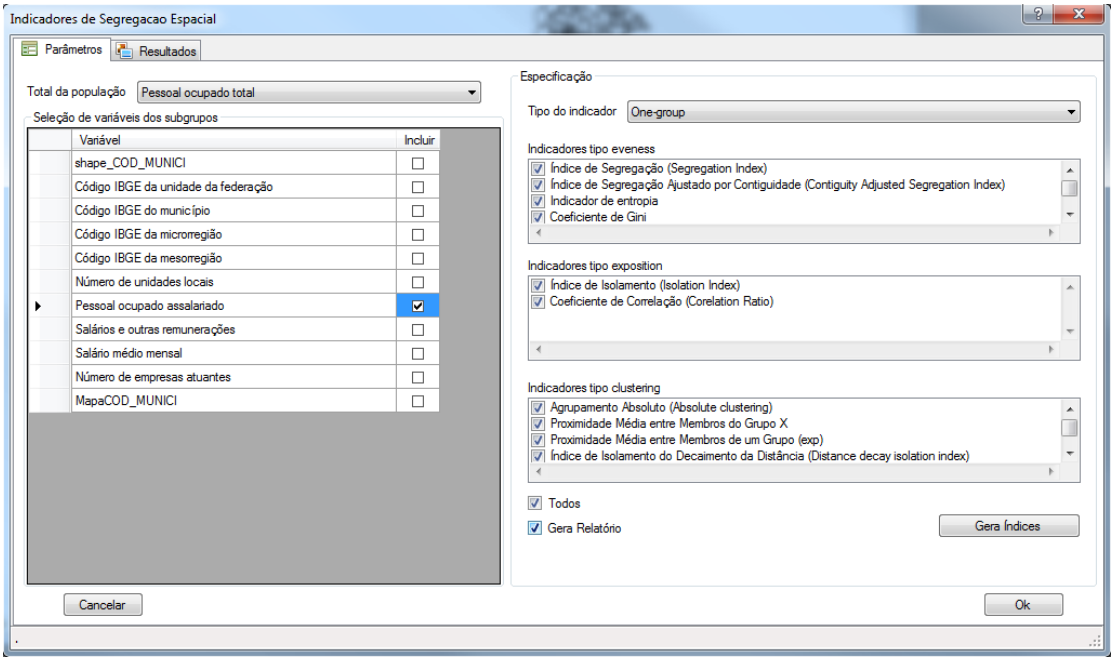
- Dissimilaridade (Dissimirarity);
- Coeficiente de Gini;
- Teoria da Informação (Information Theory).

Tipo exposition:

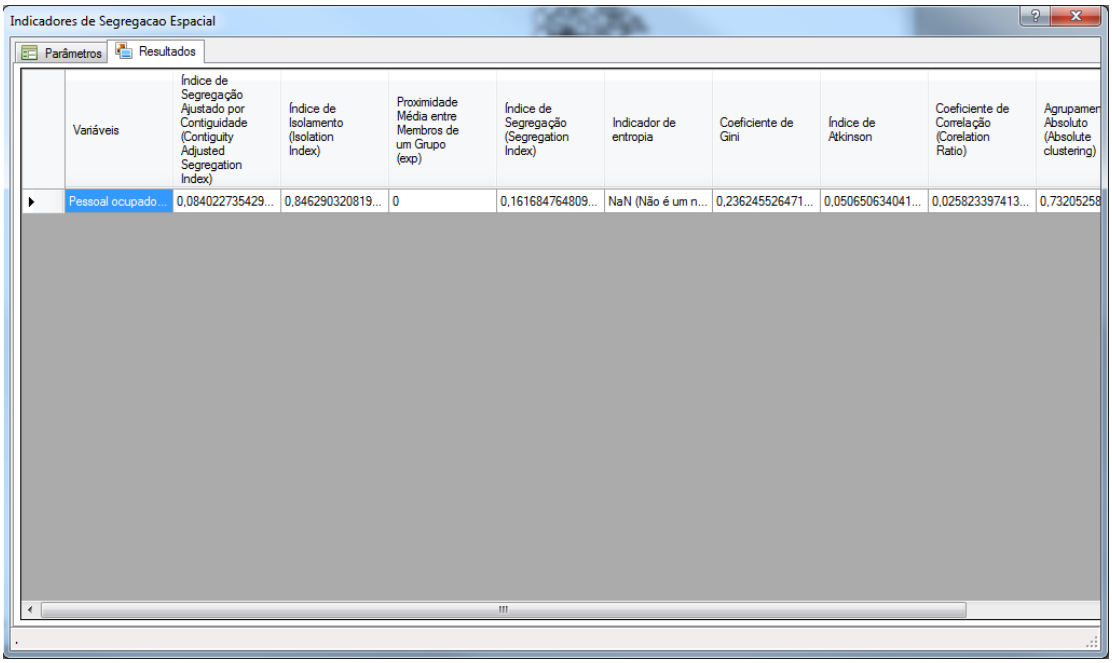
- Exposição Normalizada (Normalized Exposure);
- Diversidade Relativa (Relative Diversity);
- Coeficiente de Variação ao Quadrado (Squared coefficient of variation).



O IpeaGEO possibilita também a seleção de todos os índices, bem como a opção de gerar um relatório contendo os resultados.



Após finalizar a seleção das variáveis e das especificações desejadas, clique no botão “Gerar Índices”. A aba “Resultados” é atualizada contendo os índices solicitados.


Indicadores de Segregacao Espacial
Parâmetros
Resultados
Variáveis
Índice de Segregação Ajustado por Contiguidade (Contiguity Adjusted Segregation Index)
Índice de Isolamento (Isolation Index)
Proximidade Média entre Membros de um Grupo (exp)
Índice de Segregação (Segregation Index)
Indicador de entropia
Coeficiente de Gini
Índice de Atkinson
Coeficiente de Correlação (Corelation Ratio)
Agrupamen Absoluto (Absolute clustering)
Pessoal ocupado...
0,084022735429...
0,846290320819...
0
0,161684764809...
NaN (Não é um n...
0,236245526471...
0,050650634041...
0,025823397413...
0,73205258

Referências Bibliográficas

AGRESTI, A. Categorical data analysis. 2ª edição. New York: Wiley, 2002.

AGRESTI, A. Introduction to the Analysis of Categorical Data. New York: Wiley, 1995.

ANSELIN, L. Local indicators of spatial association – LISA. Geographical Analysis, 27, 93-115. 1995.

ANSELIN, L.; LE GALLO, J.; JAYET, H. Spatial panel econometrics. The Econometrics of Panel Data. In Mátyás, L., Sevestre, P. (eds.). Springer-Verlag, 2008.

ANSELIN L., Luc Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, 1988.

BAILEY, T.; GATRELL, A. Interactive Spatial Data Analysis. Longman, Har-low 1995.

BALTAGI, B.; EGGER, P.; PFAFFERMAYR, M. A generalized spatial panel data model with random effects. Syracuse University, Working Paper, 2007a.

BALTAGI, B.; SONG, S. H.; JUNG, B. C.; KOH, W. Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics, v. 140, 2007b.

BUSSAB, W. O.; MORETTIN, P. A. Estatística Básica. 5ª Ed. São Paulo: Saraiva, 2002. 536p.

CARON, R. Regressão de dados binários: distribuição weibull. São Carlos: UFSCar, 2010.

CARVALHO, A. X. Y, ALBUQUERQUE, P. H. M., ALMEIDA JÚNIOR, G. R., GUIMARÃES, R. D., LAURETO, C. R. Clusterização Hierárquica Espacial com Atributos Binários. Texto para Discussão no. 1428. Ipea. 2009. http://www.ipea.gov.br/sites/000/2/publicacoes/tds/td_1428.pdf.

CLAYTON, D.; KALDOR, J. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics, v. 43, n. 3, p. 671-681, 1987.

CONNOVER, W. J. Practical Nonparametric Statistics. John Wiley & Sons, 2 ed., 1980.

CRESSIE, N. A.C. Spatial Statistics. John Wiley, New York, 1993.

CÂMARA, G.; DAVIS C.; MONTEIRO, A.M.; D'Alge, J. C. Introdução à Ciência da Geoinformação. São José dos Campos, INPE, 2001.

DEMÉTRIO, C. G. B. Modelos Lineares Generalizados em Experimentação Agronômica. ESALQ/USP, 2002.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood for incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, v. 39, n. 1, p. 1-38, 1977 (Series B).

DOANE, D.P. Aesthetic frequency classifications. Amer. Statist. 30-4, 181- 183. 1976.

DRUCK, S. CÂMARA, G. e MONTEIRO, A.M. Análise espacial de dados geográficos. Embrapa, Planaltina, 2004.

DUNCAN, O. D., DUNCAN, B. A methodological analysis of segregation indexes. American Sociological Review, 20, 1955a

HIRAI, Y. Some Remarks on Class Interval of Histograms. 1989.

JOHNSON, R. A., WICHERN, D. W. Applied Multivariate Statistical Analysis. 6ª ed. Pearson International Edition, 2007.

KAPOOR, M.; KELEJIAN, H. H.; PRUCHA, I. R. Panel data models with spatially correlated error components. Journal of Econometrics, v. 140, 2007.

KHANDKER S. R.; KOOLWAL G. B.; SAMAD H. A. Handbook on impact evaluation: quantitative methods and practices. World Bank, 2010.

LEE, L. F. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory, v. 26, pp. 564-597, 2010a.

LEE, L. F. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics, v. 152(2), pp. 165-185, 2010b

MANN, W. B; WHITNEY, D. R. On a test of whether one of two random variables is stochasically larger than the other. Annals of Mathematical Statistics, v. 18, n. 1, p. 50 – 60, 1947.

MARSHALL, R. J. Mapping disease and mortality rates using empirical bayes estimators. Journal of the Royal Statistical Society, v. 40, p. 2, p. 283-294, 1991 (Series C - Applied Statistics).

MINGOTI, S. A. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Editora UFMG, 2005.

MORAN, P. A. P. Notes on Continuous Stochastic Phenomena. Biometrika, 37, 17-33. 1950.

MORI,T. An optimal length of class interval for histogram. Japan.J.Applied Statistics , 4-1, 17-24. 1974.

MUTL, J. Dynamic panel data models with spatially correlated disturbances. Ph.D. thesis, University of Maryland, College Park, 2006.

MUTL, J.; PFAFFERMAYR, M. The spatial random effects and the spatial fixed effects model: the Hausman test in a Cliff and Ord panel model. Manuscript, Institute of Advanced Studies, Vienna, 2008.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General), v. 135, n. 3, 1972.

PAULA, G. A. Modelos de Regressão com apoio computacional. São Paulo: IME/USP, 2010.

PRINGLE, D. G. Mapping disease risk estimates based on small numbers: An assessment of empirical bayes techniques. Economic and Social Review, v. 27, p. 341-363, 1996.

RIPLEY, B. D. Spatial Statistics. John Wiley, New York, 1981.

ROGERSON, P. The detection of clusters using a spatial version of the chi squared goodness-of-fit test. Geographical Analysis, 31, 130-147, 1999.

ROGERSON, P.; YAMADA, I. Statistical Detection and Surveillance of Geo-graphic Clusters. Chapman and Hall, 2008.

SATTERTHWAITE, F. E. An approximative distribution of estimates of variance componentes. Biometrics Bulletin, v. 2, p. 110 – 114, 1947.

STEVENSON, W.J. Estatística: Aplicada à Administração. São Paulo: Harbra, 2001. 511p.

STURGES, H.A. The choice of a class interval. J.Amer.Statist.Assoc. 21.65-66. 1926.

TANGO, T. A class of tests for detecting ”general”and ”fo-cused”clustering of rare diseases. Statistics in Medicine, 7, 649-660. 1955.

WALLER, L. A.; TURNBULL, B. W.; CLARK, L. C.; NASCA, P. Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. Environmetrics 3, 281-300, 1992.

WELCH, B. L. The generalization of “student’s” problem when several different population variances are involved. Biometrika, v. 34, p. 28 – 35, 1947.

WONG, D. W. S. Geostatistics as measures of spatial segregation. Urban Geography, 20 (7), 635-647, 1999.

YU, J., LEE, L. F. (2010). Some recent developments in spatial panel data models. Regional Science and Urban Economics, v. 40, pp. 255-271.