Mixed Model Regression – Part 3: Aspatial Regression

A SpaceStat Software Tutorial

AllCancerFemale.SPT SpaceStat project

ESTIMATED TIME

15 minutes

OBJECTIVE

This tutorial will use the Mixed Model Regression method in SpaceStat software to explore the patterns in all cancer mortality rates among women, 1969-2002. We will look for changing spatial patterns over time and see if some basic sociodemographic covariates can help to explain them. The mixed model approach will allow specified covariates to vary across broad U.S. regions.

In this step we will apply Aspatial Regression models to examine the significance of the covariates and to calculate how much of the variation of the rates we can explain by them.

GETTING STARTED

Continue with your project from Mixed Model Regression – Part 2: Data Exploration or open the project file AllCancerFemale.SPT, which contains all the data needed for this tutorial.

BACKGROUND BEHIND ASPATIAL REGRESSION

We will first apply Aspatial Regression models to examine the significance of the covariates and to calculate how much of the variation of the rates we can explain by them (as measured by the adjusted R-square statistic).

STEP 1: START THE ASPATIAL REGRESSION METHOD

Go to the Methods menu and select the Aspatial Regression method from the Regression submenu. This should start the Aspatial Regression method. If you do not have the Task Manager open, it will be opened for you. This window is located in a pane on the right side of the SpaceStat interface by default.
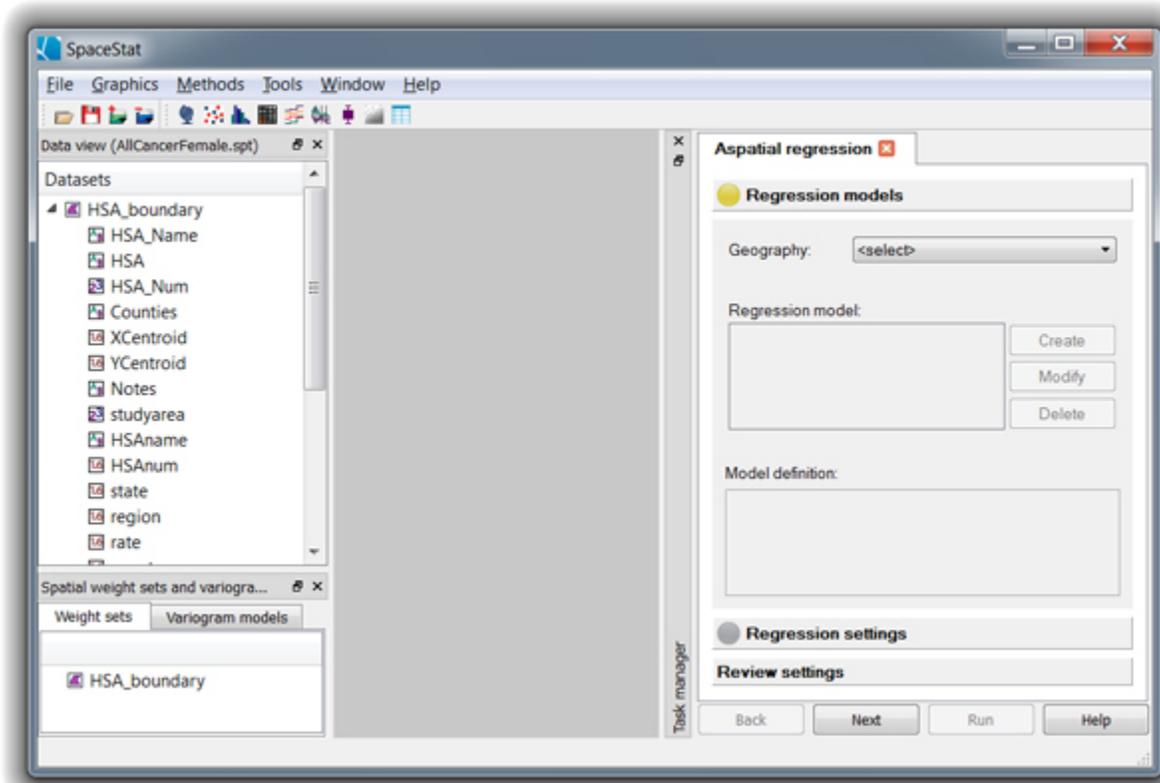
The SpaceStat interface after starting the Aspatial Regression method. The Task Manager is the right pane. Note: all windows created during Mixed Model Regression – Part 2: Data Exploration have been closed.

STEP 2: SELECT THE GEOGRAPHY

From the dropdown at the top of the Task Manager, **choose HAS_boundary as the geography**. This should enable the Create button in the Regression model settings.
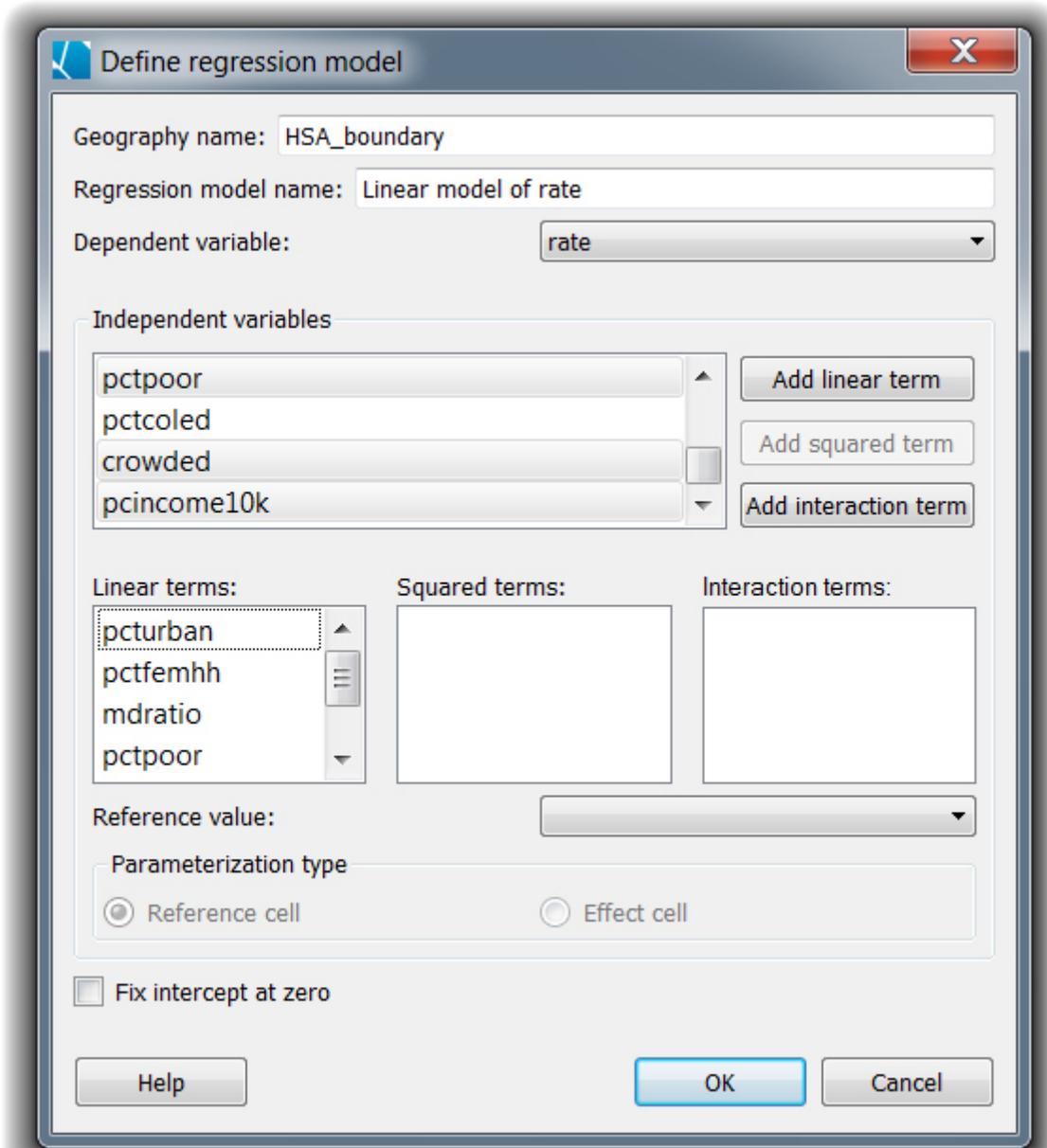
STEP 3: SETUP THE REGRESSION MODEL

Press the 'Create' button in the Regression model section to begin the process of setting up the regression model.

We include as fixed main effects all of the covariates except pctcoled (we'll choose the scaled pcincome10k instead of the original pcincome). This will use the independent variables pcturban, pctfemhh, mdratio, pctpoor, crowded, pcincome10k. Because the rates appeared to be normally distributed, we can select rate as the dependent variable and choose a linear regression model.

Replace the default name for your regression model with 'Linear model of rate'. Next **select rate from the dropdown of possible dependent variables. Select each of the independent variables**
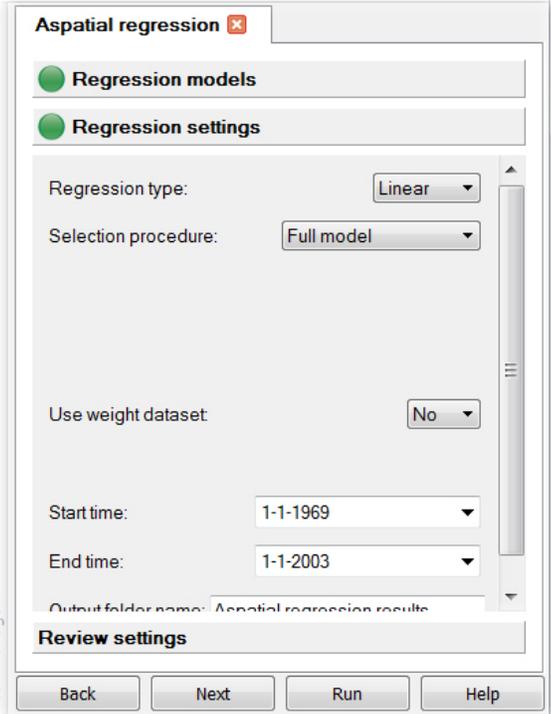
listed above (pcturban, pctfemhh, mdratio, pctpoor, crowded, and pcincome10k) from the
independent variables list and **press the 'Add linear term' button**. Compare your dialog settings
with the screenshot below and press the 'OK' button to create the regression model when
everything looks right.

Dialog settings for creating the regression model (**NOTE:** there are two more independent variables in
the linear terms list that can be seen if you scroll down – crowded and pcincome10k)

Click the 'Next' button to proceed to the next part of the regression settings. In an aspatial regression, data for each time period are modeled separately by Ordinary Least Squares (OLS). Note that normally we would include a weight for each rate to reflect the dependence of its underlying variance on its population size – the weight could be the inverse of its standard error or the square root of population. However, for simplicity we are ignoring weights for this tutorial. **You can accept the defaults** for the rest of the settings (linear regression, full model selection, no weight dataset, complete start and end times, and the default output folder name).


Default settings for the remainder of the regression options.

You can **click the 'Next' button** to go the Review settings step of the Task Manager. If everything looks correct, **click the Run button** to start the analysis. A map of the residuals and a scatterplot of the predicted versus the actual dependent variable will be created for you as well as an extensive entry into the Log View.

Examine the Log View to see a summary of the results of the aspatial regression. If the Log View is not visible, you can make it visible by **selecting 'Log View' in the Window menu**. As you scroll back in the Log View, remember that the data for each time period are modeled separately and the

results are consequently shown for each time period. We have summarized the results in the following Microsoft Excel graph:
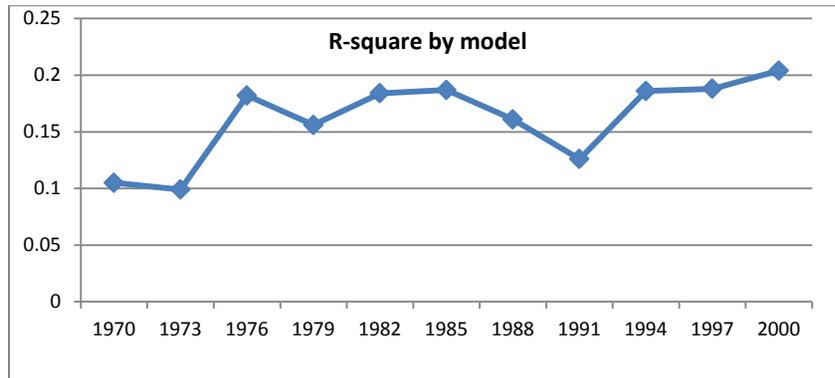
R-square values for each time period modeled by the aspatial regression.

Considering that we only have 6 sociodemographic covariates in this ecologic model with no interactions, we have a reasonably good fit. The R-square measure doubles from the first to last time period, suggesting that the impact of at least one of the covariates on rate prediction is increasing over time. The following 2 plots of the regression coefficients show that this is true for most of the covariates. Notable is the increase in the pctpoor coefficient in the later periods at the same time that the crowded coefficient is decreasing. The coefficient for mdratio is so high for the 1969-1971 period that we should probably check to see if there is an input data error.
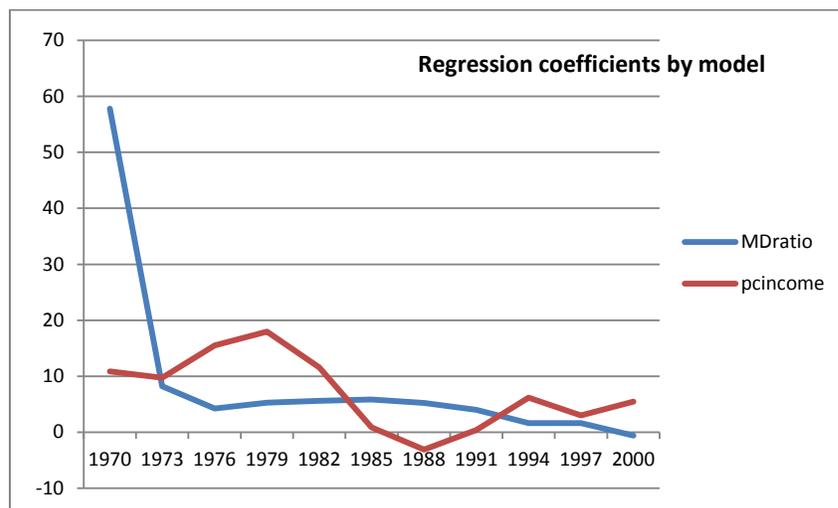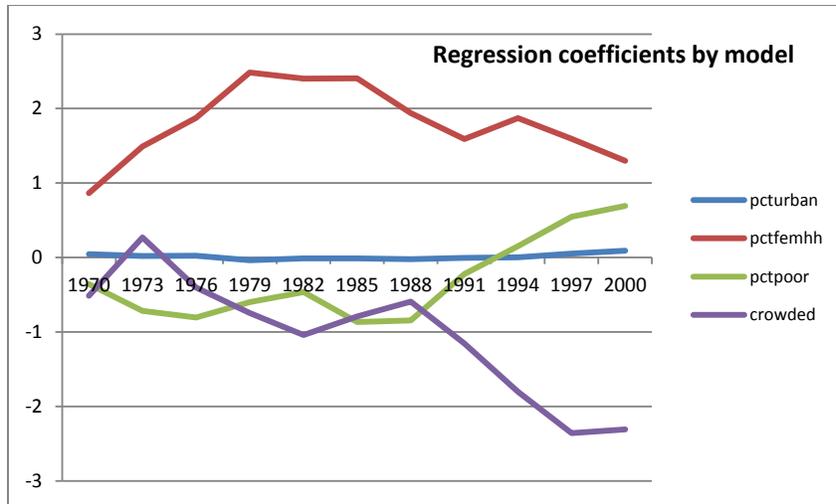
Further summary plots of the aspatial regression