Exploratory Regression Basic Tool Help

Summary

The Exploratory Regression tool evaluates all of the possible combinations of the input candidate explanatory variables, looking for models that best explain the dependent variable within the context of user-specified criteria. It produces an output report file and optional tables. A full explanation of each output is provided in Interpreting Exploratory Regression Results document. This tool uses Ordinary Least Squares (OLS) and Spatial Autocorrelation (Global Moran's I), both from the Spatial Statistics toolbox.

Usage

- Using Exploratory Regression to find a properly specified OLS model increases your chances of overfitting your model and diminishes your ability to make inferences. Please see "The Wiggle Clause" in Learn more about the Exploratory Regression tool before using this tool.

- The diagnostics in the output report and tables are obtained from OLS and Spatial Autocorrelation analyses. For more information about interpreting the diagnostics and finding a properly specified OLS model, see What they don't tell you about Regression Analysis. Additional resources:
    - Ordinary Least Squares (OLS)
    - Interpreting OLS results
    - Regression Analysis Basics
    - Spatial Autocorrelation (Global Moran's I)

- Only those models that pass the Min Adj. R-Squared, Max Coefficient p-value, Max VIF, and Min Jarque-Bera p-value criteria, as well as the models with the highest Jarque-Bera and the highest R-Squared, will have their residuals tested for Spatial Autocorrelation. As a result, there will be fewer tests of Spatial Autocorrelation than the other criteria.

- We recommend that explanatory variables should only be included in the analysis if there is appropriate justification for their inclusion.

- The Exploratory Regression tool tests all combinations of the included candidate explanatory variables. As a result, it is combinatorially complex and each additional explanatory variable will increase the execution time. In fact, if you include many candidate explanatory variables (more than 50) and enter a large number (like 6) for the Max Number of Explanatory Variables, the tool will likely run out of memory. The best

thing you can do to improve tool performance is reduce the number of candidate explanatory variables.

  - Start with a small Max Number of Explanatory Variables parameter (like 3) and run the Exploratory Regression tool. Look at the Summary of Statistical Significance output and remove any candidate explanatory variables that are never associated with statistically significant models.

  - Gradually increase the Max Number of Explanatory Variables parameter and re-run Exploratory Regression. After each run look at the Summary of Multicollinearity to identify redundancy among the candidate explanatory variables. Remove redundancies where possible. You may want to consult the Summary of Statistical Significance to decide which variables to remove.

- Models are considered passing if they meet all of the user-specified search criteria.

- If there are no models where all of the explanatory variables pass the Max VIF value and Max Coefficient p-value criteria, the resulting table will be empty.

- **Note:** A passing model is not necessarily a properly specified model. If all of the defaults are accepted, however, the resulting passing models will be properly-specified OLS models.