Regression Analysis

Regression analysis is useful in understanding why an even occurs by analysing the factors contributing to a phenomenon and determining the extent to which these factors contribute to the phenomenon. Regression mathematically models the relationship between variables to provide more insight to the understanding of why events occur.

Ordinary Least Squares (OLS) Regression

Enables you to see how well the dependent variable is explained by the combination of the explanatory variables that are provided by assigning one equation to all the feature you are analysing. The results from the regression must undergo six statistical checks to check if the model can be trusted. The first check calculates a coefficient for each variable together with a statistically significant Koenker statistic to check if you have nonstationary relationships in the model. The second check indicates the type of relationship the variable has with the dependent variable by giving the coefficient a positive or negative sign. The third check is for multicollinearity in the variables. The fourth is to check whether the model is biased or not. This is done by analysing the significance of the p-value for the Jarque-Bera statistic. The fifth check is carried out by running spatial autocorrelation which checks for clustering in the residuals of the model. The sixth check refers to the performance of your model by analysing the adjusted R-Squared value.

In ArcGIS the Exploratory Regression tool runs ordinary least squares on all possible combinations of explanatory variables and shows the variables that pass all six checks. The tool also outputs diagnostics such from each of the checks that further assist in choosing the best passing model.

Geographically weighted regression (GWR)

Geographically weighted regression is a local regression model that calculates an equation for each feature using nearby features rather than all features in a dataset. This allows for relationships to change over space. This type of regression is useful in predicting values by analysing how the relationship between explanatory variables and dependent variables change over space. The tool does this y creating coefficients for each feature and then using this to determine the relationship strength between the variables. These coefficients are able to be viewed in the table of contents. This regression model is calibrated using known values for dependent variable and the explanatory variables.

Point Pattern Analysis

The spatial distribution of data allows for visual analysis of clusters, dispersion, the direction of the data and where the centre of the data is. Further exploration of the data in ArcGIS can be done using the search window to explore spatial statistic tools such as mean centre and median centre. The Spatial Statistics toolbox also has the Measuring Geographic Distributions toolset which generates error ellipses, to show the directional distribution of the data, among many other useful outputs. Exploration of the spatial correlation in the data is done in the semivariogram cloud. The Semivariogram cloud is created using the Geostatistical Analyst menu. This menu also generates a histogram and the normalQQPlot which are also very useful in analysing the distribution of the data. The cloud works by allowing you to specify parameters for its extent. Trends in the spatial data describe patterns of variation which can reveal important information about data. The software fits a trend line to the data which is a mathematical function that describes variation in the data. Point pattern analysis can be approached using various tools from the Spatial Statistics toolbox and the Geostatistical Analyst toolbar. These tools can also be used in combination to compare and conduct more detailed analysis on data distribution and validation.