Chapter 6
Exploratory Data Analysis

ABSTRACT

Exploratory data analysis (EDA) tries to summarize datasets main characteristics such as nearest neighborhood indexes, standard deviation, scatterplots or quadrat analysis. This EDA chapter is divided into several sections to cover myGeoffice© options not forgetting the graphical mode when facing outputs: file data input (after all, any analysis demands data); Descriptive study of the variable (mean, kurtosis, distribution plot, etc.); 2D-3D data posting (spatial location of the data samples); Cutoff layout map (a spatial colorful plot according to the data samples values that are higher and lower against any particular threshold); G and Kipley’s K Index (to disclose clustered, uniform and random space sampling); Kernel Gaussian density (a non-parametric way to estimate the probability space density function of a variable); T-Student and F-tests (a parametric approach to check statistical differences between two sub-regions), including a brief section regarding the two-way ANOVA technique; Quadrat analysis (comparison of the statistically expected and actual counts of objects within spatial sampling areas to test randomness and clustering); XX profile scatterplot (silhouette view of the data along XX axis); and YY profile scatterplot (silhouette view of the data along YY axis).
Exploratory Data Analysis

INTRODUCTION

Exploratory data analysis (EDA) tries to summarize datasets main characteristics such as nearest neighborhood indexes, standard deviation, scatterplots or quadrat analysis. This EDA covers myGeoffice© ten options, not forgetting the graphical mode when facing outputs:

- File data input (after all, any analysis demands data)
- Descriptive study of the variable (mean, kurtosis, distribution plot, etc.)
- 2D-3D data posting (spatial location of the data samples)
- Cutoff layout map (a spatial colorful plot according to the data samples values that are higher and lower against any particular threshold)
- G & Kipley’s K Index (to disclose clustered, uniform and random space sampling)
- Kernel Gaussian density (a non-parametric way to estimate the probability space density function of a variable)
- T-Student and F-test (a parametric approach to check statistical differences between two sub-regions), including a brief sub-section regarding the two-way ANOVA technique.
- Quadrat analysis (comparison of the statistically expected and actual counts of objects within spatial sampling areas to test randomness and clustering)
- XX profile scatterplot (silhouette view of the data along XX axis)
- YY profile scatterplot (silhouette view of the data along YY axis)

Be reminded that this second menu of myGeoffice© deals with variables that are continuous in nature such as groundwater or air pollution levels. For discrete distributions (the number of diamonds in different sub-regions, for example) should be investigated and evaluated with discrete statistics such as the compound Poisson (check Clark & Harper (2000) bibliography for further information), Poisson, Bernoulli and negative binomial distributions.
Big Data, 3D Printing Technology, and Industry of the Future
[www.igi-global.com/article/big-data-3d-printing-technology-and-industry-of-the-future/204445?camid=4v1a](www.igi-global.com/article/big-data-3d-printing-technology-and-industry-of-the-future/204445?camid=4v1a)

Prediction Length of Stay with Neural Network Trained by Particle Swarm Optimization