

SEMINARIO

Valentin Todorov

*United Nations Industrial Development Organization -UNIDO-
Vienna*

Introduction to Robust Regression and Outlier Detection (with R)

Abstract: It is often the case that in real-world data sets some observations behave differently from the majority of data. Such data points are called outliers (in statistics) or anomalies (in machine learning). Sometimes outliers are caused by errors, but they could also have been recorded under exceptional circumstances, or belong to another population. Therefore it is very important to be able to detect anomalous cases, which may have a harmful effect on the conclusions drawn from the data, or, on the other hand, may contain valuable information. Robust statistical methods are designed to provide fitted models which are not sensitive to outliers, i.e. the effect of the outliers is suppressed or reduced. They also, ideally, should provide information about departures from the assumed model. Robust regression is an alternative to least squares regression when data are contaminated with outliers or influential observations, and it can also be used for the purpose of detecting influential observations. Robust regression can be used in any situation in which you would use least squares regression.

We start by introducing the robust methods for multiple linear regression. In the second part robust methods for regression in high dimensions will be considered. The third topic is about compositional data (real-valued vectors strictly positive components describing the parts of a whole and carry only relative information) and orthogonal regression. The presentation is accompanied by examples which are computed in R. The output of the different functions and their diagnostic methods available in R are discussed.

Sala de Grados I, Facultad de Ciencias

Lunes 26 de Marzo de 2018 (13:00)

Organiza: G.I.R. Probabilidad y Estadística Matemática

