



Contribution ID: 26

Type: **not specified**

Shiny-for-Python Multivariate Outlier Detection App

In pharmaceutical statistics, traditional outlier detection often focuses on univariate methods. However, a multivariate approach is essential for analysing complex datasets representing critical quality attributes, such as assay, dissolution, and disintegration time.

The Shiny-for-Python application described here employs advanced machine learning techniques, specifically Principal Components Analysis (PCA) and k-Nearest Neighbours (k-NN), to classify outliers in multivariate datasets relevant to the pharmaceutical industry. PCA facilitates dimensionality reduction, enhancing the visualization of data structures and identifying abnormal patterns. To strengthen our outlier detection approach, the medcouple statistic is integrated to determine an upper limit in the distribution of k-NN distances. This enables a detection mechanism tailored for skewed data.

The application brings accessibility to the methodology, making it easier for practitioners to implement advanced statistical techniques in their work. In many cases, scientists encountering outliers need to act immediately as conditions can change rapidly. Real-time analysis is critical, as immediate insights support informed decision-making and improve operational efficiency.

While traditional methods like Univariate Statistical Process Control (USP) may struggle with multivariate data, the combination of PCA, k-NN, and the medcouple enhances outlier detection. Additionally, the ongoing challenge of integrating real-time outlier detection with data analytics in pharmaceutical processes promotes better risk management and compliance.

Special/ Invited session

Classification

Both methodology and application

Keywords

Outlier Detection, Multivariate Analysis, Shiny-for-Python

Primary author: TOTHILL, Charles (Johnson & Johnson)

Presenter: TOTHILL, Charles (Johnson & Johnson)

Track Classification: Statistics in Pharma and Health Care