A178 Multivariate Statistical Procedures for the Classification of Controlled Substances in Simulated Street Samples

Ruth Waddell Smith, PhD, Michigan State University, School of Criminal Justice, 560 Baker Hall, East Lansing, MI 48824; and Andrew T. DeJarnette, 2331 Knob Hill Drive, Apt 10, Okemos, MI 48864*

After attending this presentation, attendees will be more familiar with the application of statistical procedures for the identification and classification of controlled substances.

This presentation will impact the forensic science community by further demonstrating the potential of statistical procedures for the evaluation of forensic evidence. While the focus will be on controlled substance data collected using infrared spectroscopy, the statistical procedures demonstrated are applicable to a wider range of evidence types and analytical techniques. Recommendations will be discussed for the application of such procedures in the analysis of spectral data.

As with many other forensic disciplines, the identification of controlled substances in submitted samples often involves comparison of analytical data to known reference standards analyzed under the same conditions. For the analysis of controlled substances, Infrared (IR) spectroscopy is among the analytical techniques considered to have the highest discriminating power (SWGDRUG Category A technique). However, as street samples often contain numerous other components, comparison to spectra of reference standards can be challenging without prior extraction and isolation of the substance.

In this research, the potential of multivariate statistical procedures for the identification of controlled substances in street samples analyzed by IR spectroscopy was investigated. First, a set of controlled substance reference standards was analyzed by IR spectroscopy with an attenuated total reflectance sampling accessory. Selected reference standards were also mixed with caffeine to generate a set of simulated street samples that were analyzed in a similar manner. All spectra were smoothed and normalized prior to further statistical analysis, which involved both exploratory and classification procedures.

Exploratory data analysis procedures are unsupervised in nature as no *a priori* knowledge of the data under investigation is necessary. Instead, natural groupings of the data are identified, which can be beneficial as the first step in the analysis of complex data. In this research, Hierarchical Cluster Analysis (HCA) and Principal Components Analysis (PCA) were investigated as the exploratory procedures. HCA assesses similarity among samples, displaying results as a dendrogram from which the similarity level at which such sample groups form can be determined. In contrast, PCA identifies sources of variance in the data set to reduce dimensionality without losing discriminatory information. Results are displayed in the form of a scatter plot in which chemically similar samples are positioned closely and distinctly from chemically different samples.

While exploratory procedures have potential and utility in a forensic setting, the ability to classify samples with statistical confidence is also desirable. To that end, two classification procedures, *k*-Nearest Neighbors (*k*-NN) and Soft Independent Modeling of Class Analogy (SIMCA) were also investigated in this research. In *k*-NN, new samples are considered members of the group to which the nearest *k* neighbors belong. For SIMCA, previously defined groups within the data set are firstly modeled using PCA. New samples are introduced to each model and the significance of the fit is evaluated statistically, with samples considered members of the groups to which there is the highest probability of membership. This highlights one major difference between *k*-NN and SIMCA: the former is a hard classification procedure, meaning that every sample will be classified to one group. In contrast, SIMCA is a soft classification procedure, meaning that the classification is not forced. As such, new samples may be classified into one group, more than one group, or no groups at all.

This presentation will demonstrate application of each exploratory and classification procedure to the spectral data set to investigate association and subsequent classification of the simulated samples to the appropriate reference standard. Results of each procedure will be discussed and compared, highlighting the importance of appropriate data pretreatment and representative statistical models for classification. Distinct advantages and disadvantages of the procedures for this application will also be presented, along with recommendations to facilitate future implementation in forensic laboratories.

Multivariate Statistics, Controlled Substances, IR Spectroscopy

 ** Presenting Author*