SAS - Applied Clustering Techniques

The course looks at the theoretical and practical implications of a wide array of clustering techniques currently available in SAS. The techniques considered include cluster preprocessing, variable clustering, -means clustering, and hierarchical clustering.

Skills Gained
- prepare and explore data for a cluster analysis
- distinguish among many different clustering techniques, making informed choices about which to use
- evaluate the results of a cluster analysis
- determine the appropriate number of clusters to retain
- profile and describe clustered observations
- score observations into clusters.

Who Can Benefit
- Intermediate or senior level statisticians, data analysts, and data miners

Prerequisites
- Before attending this course, you should
- be able to execute SAS programs and create SAS data sets. You can gain this experience by completing the SAS(R) Programming I: Essentials course.
- have completed a graduate-level course in statistics or the Statistics I: Introduction to ANOVA, Regression, and Logistic Regression course.
- have an understanding of matrix algebra.

Course Details

Introduction to Clustering
- identifying types of clustering
- measuring similarity
- classification performance

Preparation for Clustering
- preparing data for cluster analysis
- using variable clustering for variable selection
- using graphical clustering aids
- making elongated clusters more spherical
- viewing the impact of input standardization

Partitive Clustering
- -means clustering using PROC FASTCLUS
- outline the advantages of nonparametric clustering
- introducing PROC MODECLUS

Hierarchical Clustering
- comparing hierarchical clustering methods

Assessing Clustering Results
- determining the number of clusters in hierarchical and -means clustering
- profiling a cluster solution
- scoring new observations

Cluster Analysis Case Study
- variable selection
- graphical exploration of selected variables
- hierarchical clustering and determining the number of clusters
- profiling the seven-cluster solution
- modeling cluster membership
- scoring the customer database

Canonical Discriminant Analysis (CDA)Plots
- using canonical discriminant analysis to summarize multivariate data
- interpret CANDISC procedure output

Fuzzy Clustering
- performing fuzzy clustering using the (Q-technique) FACTOR procedure
- interpreting PROC FACTOR output

Assessing Multivariate Normality
- assessing multivariate normality

Schedule (as of 1 )

<table>
<thead>
<tr>
<th>Date</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Date Range</td>
<td>Type</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Jun 11, 2019 – Jun 14, 2019</td>
<td>Virtual</td>
</tr>
<tr>
<td>Oct 1, 2019 – Oct 4, 2019</td>
<td>Virtual</td>
</tr>
<tr>
<td>Dec 3, 2019 – Dec 6, 2019</td>
<td>Virtual</td>
</tr>
</tbody>
</table>