

Predictive Modeling Using SAS(R) High-Performance Analytics Procedures

Code:	PMHP41
Length:	3 days
URL:	View Online

SAS high-performance procedures provide predictive modeling tools that have been specially developed to take advantage of parallel processing in both multithreaded single-machine mode and distributed multiple-machine mode to solve big data problems. This course gives overview of all SAS High-Performance solutions and specifically introduces the functionality in the SAS High-Performance Statistics and Data Mining procedures for predictive modeling. The course shows examples of applying advanced statistics to huge volumes of data and quickly retrain many predictive models using all available processing power in a single-machine mode and in distributed mode.

Skills Gained

- set session options to specify the high-performance architecture for a SAS session
- explain how High-Performance procedures are designed and intended to be used
- identify similarities and differences between traditional SAS procedures and their SAS High-Performance Analytics counterparts
- use Base SAS High-Performance utility procedures
- perform model selection, dimension reduction, and identification of important variables
- use SAS High-Performance procedures to build and assess predictive models for a binary target as well as an interval target
- perform model selection for generalized linear models
- fit zero-inflated models with variable selection.

Who Can Benefit

- Experienced statisticians and predictive modelers who need to learn the functionality and use of SAS High-Performance Analytics procedures to build and assess predictive models

Prerequisites

- Before attending this course, you should have
- experience in statistical analysis and predictive modeling using SAS/STAT
- experience using SAS programming.

Course Details

Introduction to SAS High-Performance Analytics

- introduction to machine learning
- overview of SAS High-Performance Analytics procedures
- shared concepts and topics (self-study)

Exploratory Analysis and Descriptive Statistics

- exploratory analysis with the HPCORR, HPDMDDB, and HPSUMMARY procedures
- recoding variables with the HPDS2 procedure

Data Preparation and Transformation for Supervised Learning

- partitioning data with the HPSAMPLE procedure
- imputing missing values with the HPIMPUTE procedure
- creating new inputs with the HPBIN procedure
- selecting variables using the HPREDUCE procedure
- supervised interval variable selection using the HPPLS procedure (self-study)

Supervised Models with Binary Target

- logistic regression with the HPLOGISTIC procedure
- random forests with the HPFOREST and HP4SCORE procedures
- neural networks with the HPNEURAL procedure

Assessing Binary Predictive Models

- model scoring and assessment procedure

Supervised Models with Interval Target

- fitting a continuous response with the HPREG procedure
- fitting generalized linear models with the HPGENSELECT procedure

Unsupervised Learning Methods

- unsupervised dimension reduction using the HPPRINCOMP procedure
- clustering using the HPCLUS procedure

Appendices

- econometric modeling (self-study)
- nonlinear modeling with SAS High-Performance Analytics (self-study)