

Predictive Modelling by using SAS/STAT Course Details

- **Introduction to Data Analytics**
 - Definition of Data
 - Difference between data management and data analytics
 - Overview of SAS
 - Levels of Measurement and Variable types
 - Descriptive Statistics and Picturing Distributions
 - Confidence Interval for the Mean
- **Hypothesis Testing and ANOVA**
 - One Sample t-test of comparing means by using proc ttest
 - Two Sample t-test of comparing means by using proc ttest
 - One tail t-test using proc ttest
 - Paired Sample t-test using proc ttest
 - One Way ANOVA by using proc glm
 - Assumptions of ANOVA Modelling by using ODS GRAPHICS
 - n-way ANOVA by using proc glm
 - Two Way ANOVA with interactions
 - ANOVA Post Hoc Studies by using proc glm
- **Exploratory Data Analysis**
 - Data Exploration by using Scatter Plots
 - Pearson and Spearman Correlations by using proc corr
 - Analyse the output from proc corr
- **Linear Regression**
 - Fit Simple Linear Regression Model by using proc reg
 - Assumptions of Linear Regression Model
 - Analyse the output of the REG procedure for simple linear regression models
 - Producing Predicted Values by using proc score
 - Difference between Simple Linear Regression and Multiple Linear Regression Models
 - Fit Multiple Linear Regression Model by using proc reg
 - Stepwise Regression/Model Selection Techniques
- **Regression Diagnostics**
 - Residual Plots by using ODS GRAPHICS
 - Influential Observation
 - Difference between Influential Observation and Outliers
 - Collinearity Diagnostics
- **Model Building Process**
- **Categorical Data Analysis**
 - Examining Distributions
 - Test of Associations by using chi-square test
 - Fisher's Exact p-values for Pearson Chi-square test
- **Logistic Regression**
 - Odds and Odds Ratio
 - Simple Logistic Regression
 - Multiple Logistic Regression with categorical predictors
 - Analyse the output from Proc logistic

- **Measure Model Performance**

- Apply the principles of honest assessment to model performance measurement
- Assess classifier performance using the confusion matrix
- Model selection and validation using training and validation data
- Create and interpret graphs (ROC, lift, and gains charts) for model comparison and selection
- Establish effective decision cut-off values for scoring