Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression

Duration: 21.0 hours

This course is for SAS software users who perform statistical analyses using SAS/STAT software. The focus is on *t* tests, ANOVA, linear regression, and logistic regression. This course (or equivalent knowledge) is a prerequisite to many of the courses in the statistical analysis curriculum.

Learn how to

- generate descriptive statistics and explore data with graphs
- perform analysis of variance and apply multiple comparison techniques
- perform linear regression and assess the assumptions
- use regression model selection techniques to aid in the choice of predictor variables in multiple regression
- use diagnostic statistics to assess statistical assumptions and identify potential outliers in multiple regression
- use chi-square statistics to detect associations among categorical variables
- fit a multiple logistic regression model.

Who should attend: Statisticians, researchers, and business analysts who use SAS programming to generate analyses using either continuous or categorical response (dependent) variables

Prerequisites

Before attending this course, you should

- have completed the equivalent of an undergraduate course in statistics covering *p*-values, hypothesis testing, analysis of variance, and regression
- be able to execute SAS programs and create SAS data sets. You can gain this experience by completing the *SAS Programming 1: Essentials* course.

Course Contents

Prerequisite Basic Concepts

- descriptive statistics
- inferential statistics
- steps for conducting a hypothesis test
- basics of using your SAS software

Introduction to Statistics

- examining data distributions
- obtaining and interpreting sample statistics using the UNIVARIATE and MEANS procedures
- examining data distributions graphically in the UNIVARIATE and SGPLOT procedures
- constructing confidence intervals
- performing simple tests of hypothesis

t Tests and Analysis of Variance

- performing tests of differences between two group means using PROC TTEST
- performing one-way ANOVA with the GLM procedure
- performing post-hoc multiple comparisons tests in PROC GLM
- performing two-way ANOVA with and without interactions

Linear Regression

- producing correlations with the CORR procedure
- fitting a simple linear regression model with the REG procedure
- understanding the concepts of multiple regression
- using automated model selection techniques in PROC REG to choose from among several candidate models
- interpreting models

Linear Regression Diagnostics

- examining residuals
- investigating influential observations
- assessing collinearity

Categorical Data Analysis

- producing frequency tables with the FREQ procedure
- examining tests for general and linear association using the FREQ procedure
- understanding exact tests
- understanding the concepts of logistic regression
- fitting univariate and multivariate logistic regression models using the LOGISTIC procedure

Predictive Modeling Using Logistic Regression

Duration: 14.0 hours

This course covers predictive modeling using SAS/STAT software with emphasis on the LOGISTIC procedure. This course also discusses selecting variables, assessing models, treating missing values and using efficiency techniques for massive data sets.

Learn how to

- use logistic regression to model an individual's behavior as a function of known inputs
- create effect plots and odds ratio plots using ODS Statistical Graphics
- handle missing data values
- tackle multicollinearity in your predictors
- assess model performance and compare models.

Who should attend: Modelers, analysts and statisticians who need to build predictive models, particularly models from the banking, financial services, direct marketing, insurance and telecommunications industries

Prerequisites
Before attending this course, you should

- have experience executing SAS programs and creating SAS data sets, which you can gain from the *SAS Programming 1: Essentials* course
- have experience building statistical models using SAS software
- have completed a statistics course that covers linear regression and logistic regression, such as the *Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression* course.

Course Contents

Predictive Modeling

- business applications
- analytical challenges

Fitting the Model

- parameter estimation
- adjustments for oversampling

Preparing the Input Variables

- missing values
- categorical inputs
- variable clustering
- variable screening
- subset selection

Classifier Performance

- ROC curves and Lift charts
- optimal cutoffs
- K-S statistic
- c statistic
- profit
- evaluating a series of models