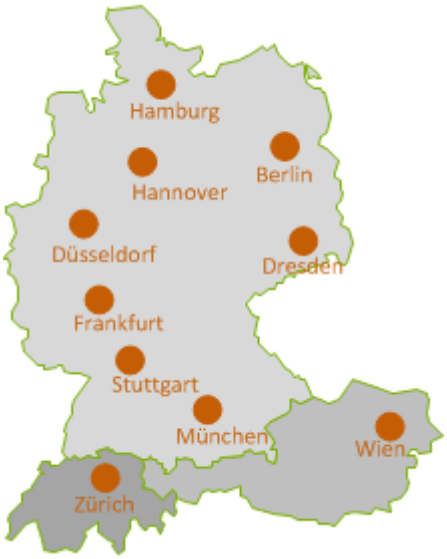
Course Catalog

Table Of Contents

a. Locations



Our trainings take place at various locations in the German-speaking countries.

Public trainings:

You can enroll for public trainings at our training centers across Germany like in Berlin, Dresden, Hamburg, München / Munich, Düsseldorf, Frankfurt, and Stuttgart. Not all public trainings will be organized in all cities but you can still book a particular training for your team in one of our training and conference centers.

In Austria you can attend seminars and trainings in Wien / Vienna while we offer training dates in Switzerland in Zürich / Zurich.

On-site trainings:

We have mobile and flexible trainers / lecturers who like to visit you and your team for an on-site training or a training in a conference center or hotel near you.

USA

Chicago	Tel: Fax:
Miami	Tel:+1.305.395.7962 Fax:+1.305.395.7964
New York	Tel:+1.212.380.1181 Fax:+1.305.395.7964

1. Mathematics

A. Data Mining

(i) Concepts and Techniques

Overview

Course ID	2020753
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Information workers, IT professionals
Prerequisites	Basics in Statistics
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,400.00 USD	1,350.00 USD	1,400.00 USD
13-14 Aug 08-09 Oct 03-04 Dec	20-21 Aug 15-16 Oct 10-11 Dec	30-31 Jul 24-25 Sep 19-20 Nov

Prices plus local taxes.

Course Description

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD) is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Course Outline

A. Introduction to Data Mining

(0.5 Days) Overview: Why Data Mining? What Is Data Mining? What Kinds of Data Can Be Mined? What Kinds of Patterns Can Be Mined? Which Technologies Are Used? - Data Preparation: Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Measuring Data Similarity and Dissimilarity - Data Preprocessing: Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization - Data Warehousing and Online Analytical Processing (OLAP)

B. Data Mining for Frequent Patterns

(0.25 Days) Frequent Itemset Mining Methods - The Apriori Algorithm - Market Basket Analysis - Pattern Evaluation Method

C. Classification using Decision Trees

(0.25 Days) Decision Tree Induction - Attribute Selection Measures - Tree Pruning - Scalability and Decision Tree Induction - Rule-Based Classification

D. Classification using Probabilistic Approaches

(0.25 Days) Bayes Classification Methods - Bayes´ Theorem –Naïve Bayes Algorithm – Bayesian Networks - Model Evaluation and Selection - Techniques to Improve Classification Accuracy

E. Classification: Advanced Methods

(0.25 Days) Classification by Backpropagation and Artificial Neural Networks - Support Vector Machines - Lazy Learners

F. Cluster Analysis

(0.5 Days) Overview of Basic Clustering Methods - Measuring Data Similarity and Dissimilarity: Data Matrix versus Dissimilarity Matrix, Proximity Measures for Nominal, Ordinal, and Binary Attributes, Dissimilarity of Numeric Data - Partitioning Methods (k-Means and k-Medoids) - Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering

(ii) Using MS Excel 2010

Overview

Course ID	2020597
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Information workers, IT Professional
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,900.00 USD	1,800.00 USD	1,900.00 USD
14-16 Sep 09-11 Nov	17-19 Aug 05-07 Oct 30 Nov - 02 Dec	24-26 Aug 12-14 Oct 07-09 Dec

Prices plus local taxes.

Course Description

Microsoft SQL Server Business Intelligence delivers a comprehensive platform empowering organizations to build and deploy secure, scalable, and manageable BI solutions. The Data Mining module provides new business insights, a reliable basis for forecasting and a comprehensive data-mining development environment. The Data Mining Add-ins allow you to harness the power of SQL Server predictive analytics in Excel and Visio. Use Table Analysis Tools to get insight with a couple of clicks. Use the Data Mining tab for full-lifecycle data mining, and build models which can be exported to a production server. Visualize your models in Visio. Microsoft SQL Server Analysis Services provides multiple algorithms for use in your data mining solutions. These algorithms are implementations of some of the most popular methodologies used in data mining. This training covers both the functions of the Data Mining Add-ins and the functions of SQL Server Data Tools. While getting to know the various software modules you will also get familiar with algorithms like Decision Trees, Naive Bayes, Clustering, Neural Networks, or Linear and Logistic Regression.

Course Outline

A. Data Mining and MS SQL Server - Introduction

(0.5 Days) Business Intelligence and Data Mining - Usage Scenarios for Data Mining - Data Mining Techniques in Microsoft SQL Server and MS Excel - Server and Client Components: MS SQL Server Analysis Services and Data Mining Add-Ins for MS Excel and MS Visio - Data Mining Life Cycle and Tasks - Data Mining Techniques in MS SQL Server - Project Cycle (Data Collection, Processing and Cleaning of Data, Modeling, Model Evaluation, Reporting, Forecasting, Integration into Applications, Model Management and Maintenance)

B. Classification using Microsoft Decision Trees

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model - DMX Queries - Classification Model, Regression Model, Relationship Model

C. Classification using Microsoft Naive Bayes

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model - DMX Queries - Dependency Network, Attribute Profiles, Attribute Characteristics, Attribute Discrimination

D. Microsoft Time Series

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Auto Regression, Multiple Time Series, Seasonality, Historic Predictions, Caching Predictions - DMX Queries

E. Microsoft Clustering

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Clustering Types, Scalable Clustering, Predictions and Cluster Assignment - DMX Queries: Cluster, Probability, Histograms, CaseLikelihood

F. Microsoft Sequence Clustering

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Markov Chains, Transition Matrix, Clustering and Markov Chains, Decomposition - DMX Queries

G. Microsoft Association Rules

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Itemset, Support, Probability and Confidence, Interestingness and Importance - DMX Queries

H. Microsoft Neural Network

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Combination and Activation, Normalization and Mapping, Topology of a Neural Network , Model Training - DMX Queries

I. Table Analysis Tools for Excel

(0.25 Days) Data Cleaning and Sampling - Prediction Calculator - Shopping Basket Analysis

J. Data Mining Client for Excel

(0.25 Days) Adding and Processing Structures and Models - Testing Models - Data Mining Queries - Using Data Mining-Models in Integration Services – Using Data Mining Results in Reporting Services

(iii) Using MS SQL Server 2012

Overview

Course ID	2020995
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Business Intelligence Developer
Prerequisites	Bases MS SQL Server
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
10-12 Aug 05-07 Oct 30 Nov - 02 Dec	17-19 Aug 12-14 Oct 07-09 Dec	07-09 Sep 02-04 Nov 28-30 Dec

Prices plus local taxes.

Course Description

Microsoft SQL Server Business Intelligence delivers a comprehensive platform empowering organizations to build and deploy secure, scalable, and manageable BI solutions. The Data Mining module provides new business insights, a reliable basis for forecasting and a comprehensive data-mining development environment. The Data Mining Add-ins allow you to harness the power of SQL Server predictive analytics in Excel and Visio. Use Table Analysis Tools to get insight with a couple of clicks. Use the Data Mining tab for full-lifecycle data mining, and build models which can be exported to a production server. Visualize your models in Visio. Microsoft SQL Server Analysis Services provides multiple algorithms for use in your data mining solutions. These algorithms are implementations of some of the most popular methodologies used in data mining. This training covers both the functions of the Data Mining Add-ins and the functions of SQL Server Data Tools. While getting to know the various software modules you will also get familiar with algorithms like Decision Trees, Naive Bayes, Clustering, Neural Networks, or Linear and Logistic Regression.

Course Outline

A. Data Mining and MS SQL Server - Introduction

(0.5 Days) Business Intelligence and Data Mining - Usage Scenarios for Data Mining - Data Mining Techniques in Microsoft SQL Server and MS Excel - Server and Client Components: MS SQL Server Analysis Services and Data Mining Add-Ins for MS Excel and MS Visio - Data Mining Life Cycle and Tasks - Data Mining Techniques in MS SQL Server - Project Cycle (Data Collection, Processing and Cleaning of Data, Modeling, Model Evaluation, Reporting, Forecasting, Integration into Applications, Model Management and Maintenance)

B. Classification using Microsoft Decision Trees

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model - DMX Queries - Classification Model, Regression Model, Relationship Model

C. Classification using Microsoft Naive Bayes

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model - DMX Queries - Dependency Network, Attribute Profiles, Attribute Characteristics, Attribute Discrimination

D. Microsoft Time Series

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Auto Regression, Multiple Time Series, Seasonality, Historic Predictions, Caching Predictions - DMX Queries

E. Microsoft Clustering

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Clustering Types, Scalable Clustering, Predictions and Cluster Assignment - DMX Queries: Cluster, Probability, Histograms, CaseLikelihood

F. Microsoft Sequence Clustering

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Markov Chains, Transition Matrix, Clustering and Markov Chains, Decomposition - DMX Queries

G. Microsoft Association Rules

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Itemset, Support, Probability and Confidence, Interestingness and Importance - DMX Queries

H. Microsoft Neural Network

(0.25 Days) Introduction to the Algorithm - Parameters - Building a Model and Using the Model: Combination and Activation, Normalization and Mapping, Topology of a Neural Network , Model Training - DMX Queries

I. Scripting for Data Mining

(0.5 Days) XML/A (XML for Analysis): Generating and Using Scripts, Building, Managing and Training Data Mining Models - DMX (Data Mining Extensions): Building Data Mining Models, Managing, Training, and Querying Data Mining Models

J. Data Integration and Reporting Services

(0.25 Days) Using Data Mining-Models in Integration Services – Using Data Mining Results in Reporting Services

A. Minitab

(i) Descriptive and Inductive Statistics using Minitab

Overview

Course ID	2024696
Language	en
Duration	5 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	no
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,850.00 USD	2,650.00 USD	2,850.00 USD
24-28 Aug 19-23 Oct 14-18 Dec	31 Aug - 04 Sep 26-30 Oct 21-25 Dec	27-31 Jul 21-25 Sep 16-20 Nov

Prices plus local taxes.

Course Description

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data, or the quantitative description itself. Statistical inference (or inductive statistics) is the process of drawing conclusions from data that is subject to random variation, for example, observational errors or sampling variation. This training provides you with a substantial overview of both descriptive and inductive statistics. All topics are firstly explained in presentations with the fundamental mathematical theory and examples followed secondly by hands-on exercices.

Course Outline

A. Introduction to Statistics

(0.5 Days) Descriptive and Inductive Statistics - Uni-/Bi- and Multi-variate Statistics - Summary tables: Grouped data, Frequency distributions, Contingency tables - Statistical graphics: Bar chart, Biplot, Box plot, Histogram

B. Descriptive Statistics: Univariate Analysis

(1 Day) Location: Mean (Arithmetic, Geometric, Harmonic), Median, Mode - Dispersion: Range, Standard deviation, Coefficient of variation, Percentiles, Interquartile range - Shape: Variance, Skewness, Kurtosis, Moments

C. Descriptive Statistics: Bivariate Analysis

(1 Day) Dependence: Pearson product-moment correlation, Rank correlation (Spearman's rho, Kendall's tau), Partial correlation, Scatter plot - Linear regression: Simple linear regression, Ordinary least squares - Regression analysis: Errors and residuals, Regression model validation, Mixed effects models

D. Inductive Statistics: Probability Theory

(0.75 Days) Probability axioms - Probability space Sample space - Elementary event - Random variable - Probability measure - Complementary event - Joint probability - Marginal probability - Conditional probability - Independence - Conditional independence - Law of total probability - Law of large numbers - Bayes' theorem - Venn diagram - Tree diagram

E. Inductive Statistics: Probability Distributions

(0.5 Days) Introduction: Probability mass function, Probability density function, Probability distribution function - Discrete univariate distributions: Binomial, Poisson, Geometric, Hypergeometric - Continuous univariate distributions: Uniform, Exponential, Normal (Gaussian)

F. Inductive Statistics: Frequentist Inference

(0.5 Days) Unbiased estimators (Mean unbiased minimum variance, Median unbiased) - Confidence interval - Testing hypotheses - Alpha-/Beta-Error and Power

G. Inductive Statistics: Specific Tests

(0.75 Days) Z (normal) - Student's t-test - F - Goodness of fit (Chi-squared) - Signed-rank (1-sample, 2-sample, 1-way anova)

(ii) Design and Analysis of Experiments using Minitab

Overview

Course ID	2024705
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Engineers, Quality Assurance
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,600.00 USD	1,550.00 USD	1,600.00 USD
30-31 Jul 24-25 Sep 19-20 Nov	10-11 Sep 05-06 Nov 31 Dec - 01 Jan	17-18 Sep 12-13 Nov

Prices plus local taxes.

Course Description

This training shows engineers and other members of the quality-assurance department to design and analyze experiments for improving the quality, efficiency and performance of working systems. It covers basic statistical methods which are useful for the analysis of experimental data, presents the Analysis of Variance (ANOVA), and teaches how to use factorial experiments, two-level factorial designs, blocking and confounding systems for two-level factorials, two-level fractional factorial designs, regression modeling, and and overview of the Response Surface Methodology.

Course Outline

A. Basic Statistical Methods

(0.25 Days) Basic Statistical Concepts - Sampling and Sampling Distributions - Inferences About the Differences in Means, Randomized Designs: Hypothesis Testing, Confidence Intervals, Choice of Sample Size, Comparing a Single Mean to a Specified Value - Inferences About the Differences in Means, Paired Comparison Designs - Inferences About the Variances of Normal Distributions

B. Analysis of Variance (ANOVA)

(0.25 Days) The Analysis of Variance - Analysis of the Fixed Effects Model: Decomposition of the Total Sum of Squares, Statistical Analysis, Estimation of the Model Parameters - Model Adequacy Checking - Determining Sample Size - The Random Effects Model - The Regression Approach to the Analysis of Variance

C. Experiments with Blocking Factors

(0.25 Days) The Randomized Complete Block Design: Statistical Analysis of the RCBD, Model Adequacy Checking, Estimating Model Parameters and the General Regression Significance Test - The Latin Square Design - The Graeco-Latin Square Design - Balanced Incomplete Block Designs

D. Factorial Experiments

(0.5 Days) The Two-Factor Factorial Design: Statistical Analysis of the Fixed Effects Model, Model Adequacy Checking, Estimating the Model Parameters, Choice of Sample Size - The General Factorial Design - Fitting Response Curves and Surfaces - Blocking in a Factorial Design

E. Two-Level Factorial Designs

(0.25 Days) The 2^2 Design - The 2^3 Design - The General 2k Design - A Single Replicate of the 2k Design - 2k Designs are Optimal Designs - The Addition of Center Points to the 2k Design - Blocking and Confounding Systems for Two-Level Factorials

F. Two-Level Fractional Factorial Designs

(0.125 Days) Process Capability Analysis Using a Histogram or a Probability Plot - Process Capability Ratios - Process Capability Analysis Using a Control Chart - Process Capability Analysis with Attribute Data - Gauge and Measurement System Capability Studies

G. The 3k Factorial Design

(0.125 Days) Notation and Motivation for the 3k Design - Confounding in the 3k Factorial Design - Fractional Replication of the 3k Factorial Design

H. Response Surface Methodology

(0.25 Days) Introduction to Response Surface Methodology - The Method of Steepest Ascent - Analysis of a Second-Order Response Surface - Experimental Designs for Fitting Response Surfaces

(iii) Engineering Statistics using Minitab

Overview

Course ID	2024701
Language	en
Duration	5 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
3,400.00 USD	3,200.00 USD	3,400.00 USD
24-28 Aug 19-23 Oct 14-18 Dec	07-11 Sep 02-06 Nov 28 Dec - 01 Jan	31 Aug - 04 Sep 26-30 Oct 21-25 Dec

Prices plus local taxes.

Course Description

This training presents a modern coverage of engineering statistics, focusing on how statistical tools are integrated into the engineering problem-solving process. All major aspects of engineering statistics are covered, including descriptive statistics, probability and probability distributions, statistical test and confidence intervals for one and two samples, building regression models, designing and analyzing engineering experiments, and statistical process control.

Course Outline

A. The Role of Statistics in Engineering

(0.25 Days) The Engineering Method and Statistical Thinking - Collecting Engineering Data - Retrospective Study - Observational Study - Designed Experiments - Random Samples - Mechanistic and Empirical Models - Observing Processes Over Time

B. Data Summary and Presentation

(0.25 Days) Data Summary and Display - Stem-and-Leaf Diagram - Histograms - Box Plot - Time Series Plots - Multivariate Data

C. Random Variables and Probability Distributions

(1 Day) Introduction - Random Variables - Probability - Continuous Random Variables: Probability Density Function, Cumulative Distribution Function, Mean and Variance - Important Continuous Distributions: Normal Distribution, Lognormal Distribution, Gamma Distribution, Weibull Distribution, Beta Distribution - Probability Plots: Normal Probability Plots, Other Probability Plots - Discrete Random Variables: Probability Mass Function, Cumulative Distribution Function, Mean and Variance - Binomial Distribution - Poisson Process: Poisson Distribution, Exponential Distribution - Normal Approximation to the Binomial and Poisson Distributions - More than One Random Variable and Independence: Joint Distributions, Independence - Functions of Random Variables: Linear Functions of Independent Random Variables, Linear Functions of Random Variables That Are Not Independent, Nonlinear Functions of Independent Random Variables - Random Samples, Statistics, and the Central Limit Theorem

D. Decision Making for a Single Sample

(0.5 Days) Statistical Inference - Point Estimation - Hypothesis Testing: Statistical Hypotheses, Testing Statistical Hypotheses, P-Values in Hypothesis Testing, One-Sided and Two-Sided Hypotheses, General Procedure for Hypothesis Testing - Inference on the Mean of a Population, Variance Known - Inference on the Mean of a Population, Variance Unknown - Inference on the Variance of a Normal Population - Inference on a Population Proportion - Other Interval Estimates for a Single Sample - Testing for Goodness of Fit

E. Decision Making for Two Samples

(0.5 Days) Introduction - Inference on the Means of Two Populations, Variances Known - Inference on the Means of Two Populations, Variances Unknown - The Paired t-Test - Inference on the Ratio of Variances of Two Normal Populations - Inference on Two Population Proportions - Completely Randomized Experiment and Analysis of Variance (ANOVA) - Randomized Complete Block Experiment

F. Building Empirical Models

(0.5 Days) Introduction to Empirical Models - Simple Linear Regression: Least Squares Estimation, Testing Hypotheses in Simple Linear Regression, Confidence Intervals in Simple Linear Regression, Prediction of a Future Observation, Checking Model Adequacy, Correlation and Regression - Multiple Regression: Estimation of Parameters in Multiple Regression, Inferences in Multiple Regression, Checking Model Adequacy - Polynomial Models - Categorical Regressors - Variable Selection Techniques

G. Design of Engineering Experiments

(1 Day) The Strategy of Experimentation - Factorial Experiments - 2k Factorial Design: 22 Design, Statistical Analysis, Residual Analysis and Model Checking, 2k Design for k 3 Factors, Single Replicate of a 2k Design - Center Points and Blocking in 2k Designs: Addition of Center Points, Blocking and Confounding - Fractional Replication of a 2k Design: One-Half Fraction of a 2k Design, Smaller Fractions (2kp Fractional Factorial Designs) - Response Surface Methods and Designs: Method of Steepest Ascent, Analysis of a Second-Order Response Surface - Factorial Experiments With More Than Two Levels

H. Statistical Process Control

(1 Day) Quality Improvement and Statistical Process Control - Introduction to Control Charts: Basic Principles, Design of a Control Chart, Rational Subgroups, Analysis of Patterns on Control Charts - R Control Charts - Control Charts For Individual Measurements - Process Capability - Attribute Control Charts: P Chart (Control Chart for Proportions) and nP Chart, U Chart (Control Chart for Average Number of Defects per Unit) and C Chart - Control Chart Performance - Measurement Systems Capability

(iv) Multivariate Analysis using Minitab

Overview

Course ID	2024694
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
17-19 Aug 12-14 Oct 07-09 Dec	10-12 Aug 05-07 Oct 14-16 Dec	03-05 Aug 28-30 Sep 23-25 Nov

Prices plus local taxes.

Course Description

Multivariate statistics is a form of statistics encompassing the simultaneous observation and analysis of more than one variable. The application of multivariate statistics is multivariate analysis. Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical implementation of multivariate statistics to a particular problem may involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied. This training is one part of a pair of courses on multivariate statistics. It helps you understand the techniques of complex and more advanced data analysis for marketing, controlling and engineering.

Course Outline

A. Multivariate Regression Analysis

(0.5 Days) Determination of a formula that can describe how elements in a vector of variables respond simultaneously to changes in others.

B. Multivariate Analysis of Variance (ANOVA and MANOVA)

(0.5 Days) Comparing multivariate means of several groups using the variance-covariance between variables in testing the statistical significance of the mean differences.

C. Discriminant Analysis

(0.5 Days) Examination whether a set of variables can be used to distinguish between two or more groups of cases.

D. Logistic Regression

(0.5 Days) Prediction of the outcome of a categorical dependent variable based on one or more predictor variables.

E. Factor Analysis

(0.5 Days) Extraction of a specified number of synthetic variables (latent variables or factors), fewer than the original set, leaving the remaining unexplained variation as error.

F. Clustering

(0.5 Days) Assignment of objects into groups (clusters) so that objects (cases) from the same cluster are more similar to each other than objects from different clusters.

(v) Statistical Quality Control using Minitab

Overview

Course ID	2024703
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Engineers, Quality Assurance
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,600.00 USD	1,550.00 USD	1,600.00 USD
17-18 Sep 12-13 Nov	27-28 Aug 22-23 Oct 17-18 Dec	03-04 Sep 29-30 Oct 24-25 Dec

Prices plus local taxes.

Course Description

This training provides a comprehensive treatment of the major aspects of using statistical methodology for quality control and improvement. Both traditional and modern methods are presented, including state-of-the-art techniques for statistical process monitoring and control and statistically designed experiments for process characterization and optimization. The training focuses on DMAIC (define, measure, analyze, improve, and control--the problem-solving strategy of six sigma).

Course Outline

A. Modern Quality Management And Improvement

(0.125 Days) The Meaning of Quality and Quality Improvement - Statistical Methods for Quality Control and Improvement - Management Aspects of Quality Improvement - The DMAIC Problem Solving Process

B. Data Summary and Presentation

(0.125 Days) Describing Variation: The Stem-and-Leaf Plot, The Histogram, Numerical Summary of Data, The Box Plot, Probability Distributions - Important Discrete Distributions - Important Continuous Distributions - Probability Plots

C. Statistical Inference In Quality Control and Improvement

(0.25 Days) Statistics and Sampling Distributions - Point Estimation of Process Parameters - Statistical Inference for a Single Sample - Statistical Inference for Two Samples - The Analysis of Variance (ANOVA)

D. Variables Control Charts

(0.5 Days) Control Charts for –x and R: Statistical Basis of the Charts, Development and Use of –x and R Charts, Charts Based on Standard Values, Interpretation of –x and R Charts, The Operating-Characteristic Function, The Average Run Length for the –x Chart - Control Charts for –x and s: Construction and Operation of –x and s Charts, The –x and s Control Charts with Variable Sample Size, The s^2 Control Chart - The Shewhart Control Chart for Individual Measurements

E. Attribute Control Charts

(0.5 Days) The Control Chart for Fraction Nonconforming: Development and Operation of the Control Chart, Variable Sample Size, Applications in Transactional and Service Businesses, The Operating-Characteristic Function and Average Run Length Calculations - Control Charts for Nonconformities (Defects): Procedures with Constant Sample Size, Procedures with Variable Sample Size, Demerit Systems, The Operating-Characteristic Function, Dealing with Low Defect Levels - Choice Between Attributes and Variables Control Charts

F. Determining Process And Measurement Systems Capability

(0.125 Days) Process Capability Analysis Using a Histogram or a Probability Plot - Process Capability Ratios - Process Capability Analysis Using a Control Chart - Process Capability Analysis with Attribute Data - Gauge and Measurement System Capability Studies

G. Designed Experiments In Process and Product Improvement

(0.25 Days) Factorial Experiments: Statistical Analysis, Residual Analysis - The 2k Factorial Design: The 2^2 Design, The 2k Design for 3 and more Factors, Blocking and Confounding in the 2k Design - Fractional Replication of the 2k Design - Fractional Replication of the 2k: The One-Half Fraction of the 2k Design, The 2k–p Fractional Factorial Design

H. Sampling Procedures

(0.125 Days) The Acceptance-Sampling Problem - Single-Sampling Plans for Attributes - Double, Multiple, and Sequential Sampling - Acceptance Sampling by Variables - Chain Sampling - Continuous Sampling

A. R

(i) Bayesian Statistics using R

Overview

Course ID	1000031
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in R and Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
07-09 Sep 02-04 Nov 28-30 Dec	24-26 Aug 19-21 Oct 14-16 Dec	31 Aug - 02 Sep 26-28 Oct 21-23 Dec

Prices plus local taxes.

Course Description

Bayesian statistics is a subset of the field of statistics in which the evidence about the true state of the world is expressed in terms of degrees of belief or, more specifically, Bayesian probabilities. The general set of statistical techniques can be divided into a number of activities, many of which have special Bayesian versions. This training shows how to use Bayesian and probabilistic thinking to analyze data, to make predictions, and to fit models. In a first part, you will see the differences between the frequentist and probabilistic approach and see how you can use R for Bayesian statistics. In a second part, you will see how you can apply Bayesian inference as a method of statistical inference in which Bayes' rule is used to update the probability for a hypothesis as evidence is acquired. A third part focuses on the formulation of statistical models where the unique feature of Bayesian statistics consists in requiring the specification of prior distributions for any unknown parameters. The training closes with a part on machine learning / Data Mining for classification. The examples and hands-on labs are carried out using both R and OpenBUGS. OpenBUGS is a software for the Bayesian analysis of complex statistical models using Markov Chain Monte Carlo (MCMC) methods.

Course Outline

A. Bayesian Statistics

(0.5 Days) Introduction: Quantifying Uncertainty Using Probabilities, Models and Prior Probabilities, Likelihoods and Posterior Probabilities, Bayesian Sequential Analysis - Review of Probability: Events and Sample Spaces, Unions - Intersections, Complements - Marginal and Conditional Probabilities - Bayes' Rule - Addition and Multiplication Rules

B. One-Parameter Models

(0.5 Days) Bayesian Models - Prior Probability and Prior Distributions - The Posterior Distribution - Conjugate Priors - Inference for a Population Proportion: Frequentist Approach, Bayesian Inference, Bayesian Point Estimates - R for Bayesian Analysis - Inference Using Nonconjugate Priors on Mean and Variance - Noninformative Priors

C. Multiparameter Models

(0.25 Days) Informative Priors for Mean and Variance - Conjugate Joint Prior Density for Mean and Variance

D. Model Fit using Markov Chain Monte Carlo (MCMC)

(0.5 Days) Sampling-Based Methods - Markov Chain Monte Carlo (MCMC) Methods - Bayesian Models - Hierarchical Models: Fitting Bayesian Hierarchical Models, Estimation Based on Hierarchical Models - Software OpenBUGS

E. Regression and Hierarchical Regression Models

(0.5 Days) Review of Linear Regression - Introduction to Bayesian Simple Linear Regression - Generalized Linear Models - Hierarchical Normal Linear Models - Model Comparison, Model Checking, and Hypothesis Testing - Bayes Factors for Model Comparison and Hypothesis Testing - Bayes Factors and Bayesian Hypothesis Testing

F. Data Mining and Classification in Bayesian Statistics

(0.75 Days) Statistics for Machine Learning - Learning as Inference - Principal Components Analysis - Naive Bayes - Nearest Neighbour Classification - Gaussian Processes

(ii) Data Mining using R

Overview

Course ID	1000020
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Information workers, IT professionals
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Manager

Course Dates

Chicago	Miami	New York
1,600.00 USD	1,550.00 USD	1,600.00 USD
03-04 Sep 29-30 Oct 24-25 Dec	20-21 Aug 15-16 Oct 10-11 Dec	27-28 Aug 22-23 Oct 17-18 Dec

Prices plus local taxes.

Course Description

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD) is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Course Outline

A. Data Mining-Grundlagen

(0.5 Days) Statistik, multivariate Statistik und Data Mining – Data Mining-Kreislauf - Daten-Vorverarbeitung: Beschreibende Datenaggregation, Datenbereinigung, Datenintegration und –transformation – Datenreduktion – Diskretisierung und Konzept-Hierarchien – Data Mining und Business Intelligence: Datenbanken, Data Warehouses und OLAP als Basis für Data Mining

B. Data Mining mit der Assoziationsanalyse

(0.25 Days) Suchen von häufigen Kombinationen (Frequent Itemset Mining) – Apriori-Algorithmus - Assoziationsregeln und Assoziationsanalyse - Warenkorbanalyse

C. Data Mining mit Entscheidungsbäumen

(0.25 Days) Ableitung von Entscheidungsbäumen – Auswahl von Attributen – Beschneidung von Bäumen – Ableitung von Regeln - Gütemaße und Vergleich von Modellen

D. Data Mining mit Wahrscheinlichkeitstheorie

(0.25 Days) Wahrscheinlichkeitstheorie und Bayes Theorem –Naïve Bayes-Algorithmus – Bayes Netze

E. Fortgeschrittene Data Mining-Verfahren für Klassifikation

(0.25 Days) Künstliche neuronale Netze und der Backpropagation-Algorithmus - Support Vector Machines für linear und nicht-linear trennbare Daten – Klassifikation mit Assoziationsanalyse – Lazy und Eager Learners

F. Cluster-Analyse

(0.5 Days) Einführung in die Cluster Analyse – Ähnlichkeits- und Distanzmessung - Varianten und grundlegende Techniken – Partitionierende Methoden: k-Means-Verfahren - Hierarchische Methoden: agglomerative und divisive Verfahren – Weitere Verfahren: Dichte- und Grid-basierte Methoden

(iii) Descriptive and Inductive Statistics using R

Overview

Course ID	2024697
Language	en
Duration	5 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	no
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,850.00 USD	2,650.00 USD	2,850.00 USD
24-28 Aug 19-23 Oct 14-18 Dec	31 Aug - 04 Sep 26-30 Oct 21-25 Dec	27-31 Jul 21-25 Sep 16-20 Nov

Prices plus local taxes.

Course Description

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data, or the quantitative description itself. Statistical inference (or inductive statistics) is the process of drawing conclusions from data that is subject to random variation, for example, observational errors or sampling variation. This training provides you with a substantial overview of both descriptive and inductive statistics. All topics are firstly explained in presentations with the fundamental mathematical theory and examples followed secondly by hands-on exercices.

Course Outline

A. Introduction to Statistics

(0.5 Days) Descriptive and Inductive Statistics - Uni-/Bi- and Multi-variate Statistics - Summary tables: Grouped data, Frequency distributions, Contingency tables - Statistical graphics: Bar chart, Biplot, Box plot, Histogram

B. Descriptive Statistics: Univariate Analysis

(1 Day) Location: Mean (Arithmetic, Geometric, Harmonic), Median, Mode - Dispersion: Range, Standard deviation, Coefficient of variation, Percentiles, Interquartile range - Shape: Variance, Skewness, Kurtosis, Moments

C. Descriptive Statistics: Bivariate Analysis

(1 Day) Dependence: Pearson product-moment correlation, Rank correlation (Spearman's rho, Kendall's tau), Partial correlation, Scatter plot - Linear regression: Simple linear regression, Ordinary least squares - Regression analysis: Errors and residuals, Regression model validation, Mixed effects models

D. Inductive Statistics: Probability Theory

(0.75 Days) Probability axioms - Probability space Sample space - Elementary event - Random variable - Probability measure - Complementary event - Joint probability - Marginal probability - Conditional probability - Independence - Conditional independence - Law of total probability - Law of large numbers - Bayes' theorem - Venn diagram - Tree diagram

E. Inductive Statistics: Probability Distributions

(0.5 Days) Introduction: Probability mass function, Probability density function, Probability distribution function - Discrete univariate distributions: Binomial, Poisson, Geometric, Hypergeometric - Continuous univariate distributions: Uniform, Exponential, Normal (Gaussian)

F. Inductive Statistics: Frequentist Inference

(0.5 Days) Unbiased estimators (Mean unbiased minimum variance, Median unbiased) - Confidence interval - Testing hypotheses - Alpha-/Beta-Error and Power

G. Inductive Statistics: Specific Tests

(0.75 Days) Z (normal) - Student's t-test - F - Goodness of fit (Chi-squared) - Signed-rank (1-sample, 2-sample, 1-way anova)

(iv) Design and Analysis of Experiments using R

Overview

Course ID	1000022
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Engineers, Quality Assurance
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,600.00 USD	1,550.00 USD	1,600.00 USD
10-11 Sep 05-06 Nov 24-25 Dec	27-28 Aug 22-23 Oct 10-11 Dec	03-04 Sep 29-30 Oct 17-18 Dec

Prices plus local taxes.

Course Description

This training shows engineers and other members of the quality-assurance department to design and analyze experiments for improving the quality, efficiency and performance of working systems. It covers basic statistical methods which are useful for the analysis of experimental data, presents the Analysis of Variance (ANOVA), and teaches how to use factorial experiments, two-level factorial designs, blocking and confounding systems for two-level factorials, two-level fractional factorial designs, regression modeling, and and overview of the Response Surface Methodology.

Course Outline

A. Basic Statistical Methods

(0.25 Days) Basic Statistical Concepts - Sampling and Sampling Distributions - Inferences About the Differences in Means, Randomized Designs: Hypothesis Testing, Confidence Intervals, Choice of Sample Size, Comparing a Single Mean to a Specified Value - Inferences About the Differences in Means, Paired Comparison Designs - Inferences About the Variances of Normal Distributions

B. Analysis of Variance (ANOVA)

(0.25 Days) The Analysis of Variance - Analysis of the Fixed Effects Model: Decomposition of the Total Sum of Squares, Statistical Analysis, Estimation of the Model Parameters - Model Adequacy Checking - Determining Sample Size - The Random Effects Model - The Regression Approach to the Analysis of Variance

C. Experiments with Blocking Factors

(0.25 Days) The Randomized Complete Block Design: Statistical Analysis of the RCBD, Model Adequacy Checking, Estimating Model Parameters and the General Regression Significance Test - The Latin Square Design - The Graeco-Latin Square Design - Balanced Incomplete Block Designs

D. Factorial Experiments

(0.5 Days) The Two-Factor Factorial Design: Statistical Analysis of the Fixed Effects Model, Model Adequacy Checking, Estimating the Model Parameters, Choice of Sample Size - The General Factorial Design - Fitting Response Curves and Surfaces - Blocking in a Factorial Design

E. Two-Level Factorial Designs

(0.25 Days) The 2^2 Design - The 2^3 Design - The General 2k Design - A Single Replicate of the 2k Design - 2k Designs are Optimal Designs - The Addition of Center Points to the 2k Design - Blocking and Confounding Systems for Two-Level Factorials

F. Two-Level Fractional Factorial Designs

(0.125 Days) Process Capability Analysis Using a Histogram or a Probability Plot - Process Capability Ratios - Process Capability Analysis Using a Control Chart - Process Capability Analysis with Attribute Data - Gauge and Measurement System Capability Studies

G. The 3k Factorial Design

(0.125 Days) Notation and Motivation for the 3k Design - Confounding in the 3k Factorial Design - Fractional Replication of the 3k Factorial Design

H. Response Surface Methodology

(0.25 Days) Introduction to Response Surface Methodology - The Method of Steepest Ascent - Analysis of a Second-Order Response Surface - Experimental Designs for Fitting Response Surfaces

(v) Exploratory Data Analysis using R

Overview

Course ID	1000032
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in R and Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
07-09 Sep 02-04 Nov 28-30 Dec	24-26 Aug 19-21 Oct 14-16 Dec	31 Aug - 02 Sep 26-28 Oct 21-23 Dec

Prices plus local taxes.

Course Description

Exploratory Data Analysis (EDA) is a statistical approach to analyzing data sets to summarize their main characteristics. This training primarily focuses on four main techniques of EDA: Principal Component Analysis (PCA) for quantitative variables, Correspondence Analysis (CA) and Multiple Correspondence Analysis (MCA) for categorical variables and finally (hierarchical and partitioning) clustering methods. As an umbrealla technique, this training also shows Factor Analysis (FA) and Multiple Factor Analysis (MFA). For the hands-on labs and practical examples the participants will use R and esp. FactoMineR - a special R package for the exploratory data analysis.

Course Outline

A. Principal Component Analysis (PCA)

(0.75 Days) Objectives of PCA and Introduction to PCA - Studying Individuals: The Cloud of Individuals, Fitting the Cloud of Individuals - Variables: The Cloud of Variables, Fitting the Cloud of Variables - Relationships - Interpreting the Data - Testing the Significance of the Components - Implementation with R and FactoMineR

B. Correspondence Analysis (CA)

(0.25 Days) Objectives and the Independence Model - Fitting the Clouds: Row and Column Profiles - Interpreting the Data - Implementation with R and FactoMineR

C. Multiple Correspondence Analysis (MCA)

(0.25 Days) Objectives: Studying Individuals, Variables, and Categories - Defining Distances between Individuals and Distances between Categories - CA on the Indicator Matrix: Relationship between MCA and CA, The Cloud of Individuals, Variables, and Categories - Implementation with R and FactoMineR

D. Clustering

(0.75 Days) Concepts of Similarity and Distance: Similarity between Individuals and Groups - Ward's Method - Partitioning and Hierarchical Clustering - Direct Search for Partitions: K-means Algorithm - Clustering and Principal Component Methods - Implementation with R and FactoMineR

E. Multiple Factor Analysis (MFA)

(0.75 Days) Factorial Analysis of Mixed Data - Weighting Groups of Variables - Comparing Groups of Variables and Indscal Model - Qualitative and Mixed Data - Multiple Factor Analysis and Procrustes Analysis - Hierarchical Multiple Factor Analysis - Implementation with R and FactoMineR

(vi) Geostatistics and the Analyis of Spatial Data

Overview

Course ID	1000029
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in R and Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,750.00 USD	1,700.00 USD	1,750.00 USD
03-04 Sep 29-30 Oct 24-25 Dec	20-21 Aug 15-16 Oct 10-11 Dec	27-28 Aug 22-23 Oct 17-18 Dec

Prices plus local taxes.

Course Description

Geostatistics is a branch of statistics focusing on spatial or spatiotemporal datasets. Such spatial and spatio-temporal data are everywhere. Beyond creating and viewing maps, spatial data analysis is concerned with questions not directly answered by looking at the data themselves. These questions refer to hypothetical processes that generate the observed data. Statistical inference for such spatial processes can be done using the statistical programming language and environment R. This training show beginners in geostatistics and participants working in the various domains of geoscience how to use R for their geostatistical analyses, visualization and plotting, model fitting, and inferential statistics. The first part of the training covers diverse techniques for handling spatial data in R, functions for import and exports of spatial data and creating diagrams and maps. The second part introduces time as a second dimension for spatio-temporal data. The third part shows you how to analyze spatial data and presents methods and functions for the anylsis of spatial point patterns and spatial point processes, interpolation, spatial prediction, the analysis of correlation, the variogram analysis as well as kriging, filtering or smoothing. This part also deals with modeling areal data and the analysis of spatial autocorrelation or fitting models.

Course Outline

A. Handling Spatial Data in R

(0.5 Days) Classes for Spatial Data in R - Visualising Spatial Data: The Traditional Plot System, Trellis/Lattice Plots, Interactive Plots, Colour Palettes and Class Intervals - Spatial Data Import and Export: Coordinate Reference Systems, Vector File Formats, Raster File Formats, Google Earth, Google Maps, Geographical Resources Analysis Support System (GRASS) - Map Overlay or Spatial Join - R-Packages: rdgal, spplot and ggplot, latticeExtra, raster, rgeos

B. Spatio-Temporal Data

(0.25 Days) Types of Spatio-Temporal Data - Handling Time Series Data - Selection, Addition, and Replacement of Attributes - Overlay and Aggregation - Visualization: Multi-panel Plots, Space-Time Plots, Animated Plots, Time Series Plots - R-Packages: xts, spacetime

C. Analyzing Spatial Data

(0.5 Days) Preliminary Analysis of a Point Pattern: G Function (Distance to the Nearest Event), F Function (Distance from a Point to the Nearest Event) - Statistical Analysis of Spatial Point Processes: Homogeneous and Inhomogeneous Poisson Processes, Estimation of the Intensity, Likelihood of an Inhomogeneous Poisson Process - Applications in Spatial Epidemiology: Case–Control Studies, Binary Regression, Accounting for Confounding and Covariates - R-Packages for the Statistical Analysis of Spatial Data: spatial, maptools, splancs, spatstat,

D. Interpolation and Geostatistics

(0.5 Days) Exploratory Data Analysis - Non-geostatistical Interpolation Methods - Estimating Spatial Correlation using the Variogram: Exploratory Variogram Analysis, Cutoff, Lag Width, Direction Dependence, Variogram Modelling, Multivariable Variogram Modelling - Spatial Prediction: Universal, Ordinary, and Simple Kriging, Kriging in a Local Neighbourhood, Multivariable Prediction: Cokriging, Trend Functions and Their Coefficients, - Kriging, Filtering, Smoothing - Model Diagnostics: Cross Validation Residuals, Cross Validation z-Scores, Multivariable Cross Validation - Geostatistical Simulation

E. Modelling Areal Data

(0.25 Days) Spatial Neighbours and Spatial Weights - Testing for Spatial Autocorrelation - Fitting Models of Areal Data

(vii) Graphical analysis of spatiotemporal data

Overview

Course ID	1000027
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in R and Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,850.00 USD	1,800.00 USD	1,850.00 USD
20-21 Aug 08-09 Oct 26-27 Nov	06-07 Aug 24-25 Sep 12-13 Nov	13-14 Aug 01-02 Oct 19-20 Nov

Prices plus local taxes.

Course Description

Space-time datasets are indexed both in space and in time. Their one- or two-dimensional analysis will typically start displaying the data in diagrams revealing the inner nature and relationships of the underlying variables. This training is organized into three parts, each devoted to different types of data. Each part comprises several topics and hands-on labs according to the various visualization methods or data characteristics. In the first part of the training, you will see how you can visualize time series data by using packages like zoo and xts for the analysis of time series data and packages like ggplot2, latticeExtra, and googleVis for their presentation. The next part of the training focuses on visualisation techniques for spatial data and presents packages like raster, rasterVis, maps, and googleVis. The third and last part finally combines variables which measure time and spatial data and teaches you how to create diagrams for such complex datasets.

Course Outline

A. Visualization of Time Series

(0.75 Days) Introduction to Displaying Time Series - Time on the Horizontal Axis - Time as a Conditioning or Grouping Variable - Time as a Complementary Variable - R Packages for time series data: zoo and xts - R Packages for visualization: ggplot2, latticeExtra, and googleVis

B. Visualization of Spatial Data

(0.75 Days) Introduction to Displaying Spatial Data - Thematic Maps: Proportional Symbol Mapping, Choropleth Maps, Raster Maps, Vector Fields - Reference and Physical Maps - Packages for working with OpenStreetMap - R Packages for spatial data: sp, maptools, gstat, and rgdal - R Packages for visualization: raster, rasterVis, maps, and googleVis

C. Visualization of Space-Time Data

(0.5 Days) Introduction to Displaying Spatiotemporal Data - Spatiotemporal Raster Data - Spatiotemporal Point Observations - R Package spacetime for spatiotemporal data

(viii) Multivariate Analysis using R

Overview

Course ID	2024695
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
17-19 Aug 12-14 Oct 07-09 Dec	10-12 Aug 19-21 Oct 14-16 Dec	03-05 Aug 28-30 Sep 23-25 Nov

Prices plus local taxes.

Course Description

Multivariate statistics is a form of statistics encompassing the simultaneous observation and analysis of more than one variable. The application of multivariate statistics is multivariate analysis. Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical implementation of multivariate statistics to a particular problem may involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied. This training is one part of a pair of courses on multivariate statistics. It helps you understand the techniques of complex and more advanced data analysis for marketing, controlling and engineering.

Course Outline

A. Multivariate Regression Analysis

(0.5 Days) Determination of a formula that can describe how elements in a vector of variables respond simultaneously to changes in others.

B. Multivariate Analysis of Variance (ANOVA and MANOVA)

(0.5 Days) Comparing multivariate means of several groups using the variance-covariance between variables in testing the statistical significance of the mean differences.

C. Discriminant Analysis

(0.5 Days) Examination whether a set of variables can be used to distinguish between two or more groups of cases.

D. Logistic Regression

(0.5 Days) Prediction of the outcome of a categorical dependent variable based on one or more predictor variables.

E. Factor Analysis

(0.5 Days) Extraction of a specified number of synthetic variables (latent variables or factors), fewer than the original set, leaving the remaining unexplained variation as error.

F. Clustering

(0.5 Days) Assignment of objects into groups (clusters) so that objects (cases) from the same cluster are more similar to each other than objects from different clusters.

(ix) Statistical Analysis with Graphics using R

Overview

Course ID	1000024
Language	en
Duration	4 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basic knowledge of statistics
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,400.00 USD	2,250.00 USD	2,400.00 USD
27-30 Jul 07-10 Sep 19-22 Oct 30 Nov - 03 Dec	24-27 Aug 05-08 Oct 16-19 Nov 28-31 Dec	31 Aug - 03 Sep 12-15 Oct 23-26 Nov

Prices plus local taxes.

Course Description

Graphics can effectively complement statistical data analysis in various ways. Successful graphics arise from a combination of good design and good implementation. This training explores mainly two R packages for statistical graphics: lattice and ggplot2. The lattice package extends the R language by providing a coherent set of tools to produce statistical graphics with an emphasis on multivariate data. ggplot2 is an R package for producing statistical, or data, graphics, but it is unlike most other graphics packages because it has a deep underlying grammar. This makes ggplot2 very powerful, because you are not limited to a set of pre-specified graphics, but you can create new graphics that are precisely tailored for your problem. The training is divided into two parts, with the first being an introduction to the development of graphics using the lattice package and the second one using the ggplot2 package for similar visualizations but also far more complex and more sophisticated visual analyses.

Course Outline

A. Introduction to graphics in R

(0.25 Days) Introduction to the lattice package - Multipanel conditioning - The "trellis"" object and its properties: the formula, data, conditioning and various plots/tiles in one diagram - Dimension and physical layout - Grouped displays - Annotation: Captions, labels, and legends

B. Graphics for Univariate Distributions

(0.75 Days) Density Plot - Histograms - Normal Q–Q plots - The empirical CDF (Cumulative Distribution Function) - Box-and-whisker plots - Strip plots - Working with small and large datasets

C. Graphics for Multivariate Distributions

(0.5 Days) Displaying Multiway Tables: Dot plots, Bar charts, Visualizing categorical data - Scatter Plots and Extensions - Trivariate Displays: Three-dimensional scatter plots, Surfaces and two-way tables

D. Advanced Graphical Parameters of the lattice Package

(0.5 Days) The parameter system: Themes and devices - Plot Coordinates and Axis Annotation: Axis annotations (ticks and labels), Limits and aspect ratio, Scale components and the axis function, Labels and Legends - Data Manipulation: Combining data sources, Subsetting, Ordering levels of categorical variables, Manipulating the "trellis" Object

E. Introduction to graphics using ggplot2

(0.5 Days) Datasets - Basic use - Colour, size, shape and other aesthetic attributes - Plot geometries - Components of the layered ggplot2 grammar: Layers, Scales, Coordinate system, Faceting

F. Visualization in ggplot2

(1 Day) Layers - Overall layering strategy - Aesthetic mappings - Creating a plot - Basic plot types - Displaying distributions - Dealing with overplotting - Surface plots - Drawing maps - Revealing uncertainty - Statistical summaries - Annotating a plot

G. Optimizing plots for publication and presentation

(0.5 Days) Themes - Customising scales and geoms - Multiple plots on the same page

(x) Statistical Quality Control using R

Overview

Course ID	1000021
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Engineers, Quality Assurance
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,600.00 USD	1,550.00 USD	1,600.00 USD
13-14 Aug 01-02 Oct 19-20 Nov	30-31 Jul 17-18 Sep 05-06 Nov 24-25 Dec	06-07 Aug 24-25 Sep 12-13 Nov 31 Dec - 01 Jan

Prices plus local taxes.

Course Description

This training provides a comprehensive treatment of the major aspects of using statistical methodology for quality control and improvement. Both traditional and modern methods are presented, including state-of-the-art techniques for statistical process monitoring and control and statistically designed experiments for process characterization and optimization. The training focuses on DMAIC (define, measure, analyze, improve, and control--the problem-solving strategy of six sigma).

Course Outline

A. Modern Quality Management And Improvement

(0.125 Days) The Meaning of Quality and Quality Improvement - Statistical Methods for Quality Control and Improvement - Management Aspects of Quality Improvement - The DMAIC Problem Solving Process

B. Data Summary and Presentation

(0.125 Days) Describing Variation: The Stem-and-Leaf Plot, The Histogram, Numerical Summary of Data, The Box Plot, Probability Distributions - Important Discrete Distributions - Important Continuous Distributions - Probability Plots

C. Statistical Inference In Quality Control and Improvement

(0.25 Days) Statistics and Sampling Distributions - Point Estimation of Process Parameters - Statistical Inference for a Single Sample - Statistical Inference for Two Samples - The Analysis of Variance (ANOVA)

D. Variables Control Charts

(0.5 Days) Control Charts for –x and R: Statistical Basis of the Charts, Development and Use of –x and R Charts, Charts Based on Standard Values, Interpretation of –x and R Charts, The Operating-Characteristic Function, The Average Run Length for the –x Chart - Control Charts for –x and s: Construction and Operation of –x and s Charts, The –x and s Control Charts with Variable Sample Size, The s^2 Control Chart - The Shewhart Control Chart for Individual Measurements

E. Attribute Control Charts

(0.5 Days) The Control Chart for Fraction Nonconforming: Development and Operation of the Control Chart, Variable Sample Size, Applications in Transactional and Service Businesses, The Operating-Characteristic Function and Average Run Length Calculations - Control Charts for Nonconformities (Defects): Procedures with Constant Sample Size, Procedures with Variable Sample Size, Demerit Systems, The Operating-Characteristic Function, Dealing with Low Defect Levels - Choice Between Attributes and Variables Control Charts

F. Determining Process And Measurement Systems Capability

(0.125 Days) Process Capability Analysis Using a Histogram or a Probability Plot - Process Capability Ratios - Process Capability Analysis Using a Control Chart - Process Capability Analysis with Attribute Data - Gauge and Measurement System Capability Studies

G. Designed Experiments In Process and Product Improvement

(0.25 Days) Factorial Experiments: Statistical Analysis, Residual Analysis - The 2k Factorial Design: The 2^2 Design, The 2k Design for 3 and more Factors, Blocking and Confounding in the 2k Design - Fractional Replication of the 2k Design - Fractional Replication of the 2k: The One-Half Fraction of the 2k Design, The 2k–p Fractional Factorial Design

H. Sampling Procedures

(0.125 Days) The Acceptance-Sampling Problem - Single-Sampling Plans for Attributes - Double, Multiple, and Sequential Sampling - Acceptance Sampling by Variables - Chain Sampling - Continuous Sampling

(xi) Statistical analysis using Bayesian Networks

Overview

Course ID	1000019
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,750.00 USD	1,700.00 USD	1,750.00 USD
03-04 Sep 29-30 Oct 24-25 Dec	20-21 Aug 15-16 Oct 10-11 Dec	27-28 Aug 22-23 Oct 17-18 Dec

Prices plus local taxes.

Course Description

A Bayesian network, Bayes network, belief network, Bayes(ian) model or probabilistic directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph. There are three main inference tasks for Bayesian networks: Structure learning, inferring unobserved variables, and parameter learning. This training presents the diverse techniques of statistical data analysis using Bayesian networks and shows in hands-on labs using R how to implement the techniques and algorithms. You will become familiar with R packages like bnlearn, deal, pcalg, and catnet for structure learning, and you will get to know packages like gRbase and gRain for inferential analysis. Time series data will be analyzed using packages like vars, lars, simone, and GeneNet.

Course Outline

A. Introduction

(0.25 Days) Introduction to Graph Theory: Graphs, Nodes, and Arcs - Bayesian Networks

B. Bayesian Networks and Static Data

(0.75 Days) Bayesian Networks: Essential Definitions and Properties: Graph Structure and Probability Factorization, Fundamental Connections, Equivalent Structures, Markov Blankets - Static Bayesian Networks Modeling: Constraint-Based Structure Learning Algorithms, Score-Based Structure Learning Algorithms, Hybrid Structure Learning Algorithms, Parameter Learning

C. Bayesian Networks and Time Series Data

(0.5 Days) Time Series and Vector Auto-Regressive Processes (VAR) - Dynamic Bayesian Networks: Essential Definitions and Properties, Dynamic Bayesian Network Representation of a VAR Process - Algorithms: Least Absolute Shrinkage and Selection Operator (LASSO), James–Stein Shrinkage, First-Order Conditional Dependencies Approximation

D. Bayesian Network Inference Algorithms

(0.25 Days) Reasoning Under Uncertainty: Probabilistic Reasoning and Evidence, Algorithms for Belief Updating: Exact and Approximate Inference, Causal Inference - Inference in Static Bayesian Networks: Exact Inference, Approximate Inference - Inference in Dynamic Bayesian Networks

A. SPSS

(i) SPSS Statistics - Statistical Data Analysis 1

Overview

Course ID	2023679
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in Statistics
Method	Presentation with examples and hands-on labs.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,000.00 USD	1,950.00 USD	2,000.00 USD
13-14 Aug 08-09 Oct 03-04 Dec	27-28 Aug 22-23 Oct 17-18 Dec	30-31 Jul 24-25 Sep 19-20 Nov

Prices plus local taxes.

Course Description

IBM SPSS Statistics is a comprehensive system for analyzing data. SPSS Statistics can take data from almost any type offile and use them to generate tabulated reports, charts and plots of distributions and trends, descriptive statistics, and complex statistical analyses. This training shows you how to use the graphical user interface of SPSS Statistics. You will learn how to make the most out of the wide range of statistical procedures for basic analyses and reports, including counts, crosstabs and descriptive statistics. For legal reasons (license management), this training can only be delivered on your hardware using your (test) license. Please contact us concerning dates and handling before the actual booking.

Course Outline

A. Data Files

Opening data files - File information - Saving data files - Comparing datasets - Data View - Variable View - Entering data - Editing data - Finding cases, variables, or imputations - Finding and replacing data and attribute values - Working with Multiple Data Sources

B. Data preparation and transformation

Variable properties - Defining Variable Properties - Setting measurement level for variables with unknown measurement level - Multiple Response Sets - Copying Data Properties - Visual Binning - Count Occurrences of Values within Cases - Identifying Duplicate Cases - Computing Variables - Recoding Values - Rank Cases - Time Series Data Transformations

C. Outputs

Viewer - Export output - Viewer printing - Saving output

D. Pivot Tables

Manipulating a pivot table - Workingwithlayers - Showing and hiding items - Table properties - Cell properties - Footnotes and captions

E. Working with Command Syntax

Syntax Rules - Using the Syntax Editor

(ii) SPSS Statistics - Statistical Data Analysis 2 (Multivariate Analysis)

Overview

Course ID	2023680
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basic knowledge of statistics
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
1,900.00 USD	1,800.00 USD	1,900.00 USD
17-19 Aug 12-14 Oct 07-09 Dec	10-12 Aug 05-07 Oct 30 Nov - 02 Dec	03-05 Aug 28-30 Sep 23-25 Nov

Prices plus local taxes.

Course Description

The IBM SPSS Statistics Premium Edition helps data analysts, planners, forecasters, survey researchers, program evaluators and database marketers – among others – to easily accomplish tasks at every phase of the analytical process. It includes a broad array of fully integrated Statistics capabilities and related products for specialized analytical tasks across the enterprise. The software will improve productivity significantly and help achieve superior results for specific projects and business goals. This training covers topics as the analysis of categorical and numeric data , linear and nonlinear models, decision trees, articial neural networks, forecasting and time series and more. For legal reasons (license management), this training can only be delivered on your hardware using your (test) license. Please contact us concerning dates and handling before the actual booking.

Course Outline

A. Regression

(0.5 Days) GLM Univariate - Linear Regression - Ordinal Regression - Curve Fitting - Partial Least Squares Regression

B. Discriminant Analysis

(0.25 Days) Defining Ranges - Selecting Cases - Stepwise Method - Classifying

C. Exploratory Factor Analysis

(0.25 Days) Selecting Cases - Descriptive Statistics - Factor Extraction - Rotation - Factor Values

D. Decision Trees

(0.5 Days) Creating Decision Trees: Selecting Categories, Validation Criteria for the Construction of the Tree - Tree Editor: Working with Large Trees, Controlling the Data Displayed in the Tree - Data Assumptions and Requirements - Construct a Valuation Model - Missing Values ??in Tree Models

E. Cluster Analysis

(0.5 Days) Distance Calculations - Nearest Neighbor Analysis - Two-Step Cluster Analysis - Hierarchical Cluster Analysis - Cluster Analysis

F. Artificial Neural Networks

(0.5 Days) Structure of Neural Networks - Multilayer Perceptron: Partitions, Architecture, Training, Editing - Radial Basis Function: Architecture, Training

G. Time Series

(0.5 Days) Transforming Data - Time Series Modeling: Exponential Smoothing, Custom ARIMA Models, Output - Seasonal Decomposition - Spectra - Forecasts - Determining Significant Predictors

(iii) SPSS Statistics - Statistical Data Analysis 3 (Questionnaires, Surveys and Market Research)

Overview

Course ID	2023681
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	Basics in Statistics
Method	Presentation with examples and hands-on labs.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,200.00 USD	2,100.00 USD	2,200.00 USD
17-19 Aug 12-14 Oct 07-09 Dec	10-12 Aug 05-07 Oct 30 Nov - 02 Dec	03-05 Aug 28-30 Sep 23-25 Nov

Prices plus local taxes.

Course Description

IBM SPSS Direct Marketing helps you understand your customers in greater depth, improve your marketing campaigns and maximize the ROI of your marketing budget. This training shows you how to conduct sophisticated analyses of your customers or contacts – and with a high level of confidence in your results. You will get to know concepts like cluster analysis, prospect profiling, Conjoint Analysis, Multidimensional Scaling, Correspondence Analysis or the analysis of complex samples. For legal reasons (license management), this training can only be delivered on your hardware using your (test) license. Please contact us concerning dates and handling before the actual booking.

Course Outline

A. Exploratory Data Analysis

(0.5 Days) Crosstabs - Summarizing Values ??- OLAP Cubes - Sample t-Tests - Components of Variance - One-Way ANOVA - Custom Tables: Simple Tables for Categorical Variables, Stacking, Nesting, and Layers with Categorical Variables, Totals and Subtotals, Computed Categories

B. Generalized Linear Models (GLM)

(0.5 Days) Generalized Linear Models - Generalized Linear Mixed Models - Model Loglinear Analysis - General Loglinear Analysis - Logit Loglinear Analysis

C. Analysis of Categorical Data

(0.5 Days) Categorical Regression: Define the Scale in the Categorical Regression, Discretization, Missing Values??, Output - Categorical Principal Components Analysis: Defining Scale and Weight, Discretization, Missing Values??, Output

D. Conjoint Analysis

(0.25 Days) Profile Method: Orthogonal Field, Experimental Stimuli, Data Collection and Analysis - Generating an Orthogonal Design - Display

E. Multidimensional Scaling

(0.25 Days) Multidimensional Scaling Analysis of Similarities, Creating Distances from Data - Defining a Model for the Multidimensional Scaling - Multidimensional Unfolding: Defining a Model Output

F. Correspondence Analysis

(0.25 Days) Defining the Row and Column Range in the Correspondence Analysis - Model - Statistics - Charts - Multiple Correspondence Analysis: Definition of the Variables´ Weight in the Multiple Correspondence Analysis, Discretization, Missing Values??, Output, Diagrams

G. Analyse von komplexen Stichproben

(0.5 Days) Stichprobenziehung mithilfe eines komplexen Plans - Vorbereiten einer komplexen Stichprobe für die Analyse - Häufigkeiten, Kreuztabellen und Deskriptive Statistiken - Regression für komplexe Stichproben: Logistische Regression, Ordinale Regression, Cox-Regression

H. Direktmarketing

(0.25 Days) RFM-Analyse - Clusteranalyse - Profile über potenzielle Kunden - Responseraten nach Postleitzahlen - Kaufneigung - Kontrollpakettest

A. Statistics

(i) Descriptive Statistics

Overview

Course ID	2020581
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,900.00 USD	1,800.00 USD	1,900.00 USD
07-09 Sep 02-04 Nov 28-30 Dec	03-05 Aug 28-30 Sep 23-25 Nov	10-12 Aug 05-07 Oct 30 Nov - 02 Dec

Prices plus local taxes.

Course Description

Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data, or the quantitative description itself. Descriptive statistics are distinguished from inferential statistics (or inductive statistics), in that descriptive statistics aim to summarize a sample, rather than use the data to learn about the population that the sample of data is thought to represent. Univariate analysis involves describing the distribution of a single variable, including its central tendency (including the mean, median, and mode) and dispersion (including the range and quantiles of the data-set, and measures of spread such as the variance and standard deviation). The shape of the distribution may also be described via indices such as skewness and kurtosis. Characteristics of a variable's distribution may also be depicted in graphical or tabular format, including histograms and stem-and-leaf display. When a sample consists of more than one variable, descriptive statistics may be used to describe the relationship between pairs of variables. In this case, descriptive statistics include quantitative measures of dependence. This training covers all the fundamentals of descriptive statistics which can be used in marketing, controlling and engineering. You will learn theory and the mathematical foundations in lectures with examples and you will train your new knowledge in practical hands-on labs and exercices.

Course Outline

A. Introduction to Statistics

(0.25 Days) Descriptive and Inductive Statistics - Uni-/Bi- and Multi-variate Statistics - Summary tables: Grouped data, Frequency distributions, Contingency tables - Statistical graphics: Bar chart, Biplot, Box plot, Histogram

B. Univariate Analysis: Measures of Central Tendency

(0.5 Days) Mean (Arithmetic, Geometric, Harmonic) - Median - Mode

C. Univariate Analysis: Measures of Dispersion

(0.5 Days) Range - Variance and Standard deviation - Coefficient of variation - Percentiles - Interquartile range - Shape: Variance, Skewness, Kurtosis, Moments

D. Univariate Analysis: Measures of Shape

(0.25 Days) Skewness - Kurtosis - Moments

E. Bivariate Analysis: Dependence

(0.75 Days) Continuous data: Pearson product-moment correlation, Partial correlation, Scatter plot - Ordinal data: Rank correlation (Spearman's rho, Kendall's tau) - Categorical data: Contingency tables, Cramer´s V, Phi coefficient, Chi coefficient

F. Bivariate Analysis: Regression

(0.75 Days) Linear regression: Simple linear regression, Ordinary least squares - Regression analysis: Errors and residuals, Regression model validation, Estimations - Overview of non-linear regression models

(ii) Descriptive and Inductive Statistics

Overview

Course ID	2020177
Language	en
Duration	5 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	no
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
2,850.00 USD	2,650.00 USD	2,850.00 USD
07-11 Sep 02-06 Nov 28 Dec - 01 Jan	03-07 Aug 28 Sep - 02 Oct 23-27 Nov	10-14 Aug 05-09 Oct 30 Nov - 04 Dec

Prices plus local taxes.

Course Description

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data, or the quantitative description itself. Statistical inference (or inductive statistics) is the process of drawing conclusions from data that is subject to random variation, for example, observational errors or sampling variation. This training provides you with a substantial overview of both descriptive and inductive statistics. All topics are firstly explained in presentations with the fundamental mathematical theory and examples followed secondly by hands-on exercices.

Course Outline

A. Introduction to Statistics

(0.5 Days) Descriptive and Inductive Statistics - Uni-/Bi- and Multi-variate Statistics - Summary tables: Grouped data, Frequency distributions, Contingency tables - Statistical graphics: Bar chart, Biplot, Box plot, Histogram

B. Descriptive Statistics: Univariate Analysis

(1 Day) Location: Mean (Arithmetic, Geometric, Harmonic), Median, Mode - Dispersion: Range, Standard deviation, Coefficient of variation, Percentiles, Interquartile range - Shape: Variance, Skewness, Kurtosis, Moments

C. Descriptive Statistics: Bivariate Analysis

(1 Day) Dependence: Pearson product-moment correlation, Rank correlation (Spearman's rho, Kendall's tau), Partial correlation, Scatter plot - Linear regression: Simple linear regression, Ordinary least squares - Regression analysis: Errors and residuals, Regression model validation, Mixed effects models

D. Inductive Statistics: Probability Theory

(0.75 Days) Probability axioms - Probability space Sample space - Elementary event - Random variable - Probability measure - Complementary event - Joint probability - Marginal probability - Conditional probability - Independence - Conditional independence - Law of total probability - Law of large numbers - Bayes' theorem - Venn diagram - Tree diagram

E. Inductive Statistics: Probability Distributions

(0.5 Days) Introduction: Probability mass function, Probability density function, Probability distribution function - Discrete univariate distributions: Binomial, Poisson, Geometric, Hypergeometric - Continuous univariate distributions: Uniform, Exponential, Normal (Gaussian)

F. Inductive Statistics: Frequentist Inference

(0.5 Days) Unbiased estimators (Mean unbiased minimum variance, Median unbiased) - Confidence interval - Testing hypotheses - Alpha-/Beta-Error and Power

G. Inductive Statistics: Specific Tests

(0.75 Days) Z (normal) - Student's t-test - F - Goodness of fit (Chi-squared) - Signed-rank (1-sample, 2-sample, 1-way anova)

(iii) Engineering Statistics

Overview

Course ID	2024700
Language	en
Duration	5 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
3,400.00 USD	3,200.00 USD	3,400.00 USD
24-28 Aug 19-23 Oct 14-18 Dec	31 Aug - 04 Sep 26-30 Oct 21-25 Dec	07-11 Sep 02-06 Nov 28 Dec - 01 Jan

Prices plus local taxes.

Course Description

This training presents a modern coverage of engineering statistics, focusing on how statistical tools are integrated into the engineering problem-solving process. All major aspects of engineering statistics are covered, including descriptive statistics, probability and probability distributions, statistical test and confidence intervals for one and two samples, building regression models, designing and analyzing engineering experiments, and statistical process control.

Course Outline

A. The Role of Statistics in Engineering

(0.25 Days) The Engineering Method and Statistical Thinking - Collecting Engineering Data - Retrospective Study - Observational Study - Designed Experiments - Random Samples - Mechanistic and Empirical Models - Observing Processes Over Time

B. Data Summary and Presentation

(0.25 Days) Data Summary and Display - Stem-and-Leaf Diagram - Histograms - Box Plot - Time Series Plots - Multivariate Data

C. Random Variables and Probability Distributions

(1 Day) Introduction - Random Variables - Probability - Continuous Random Variables: Probability Density Function, Cumulative Distribution Function, Mean and Variance - Important Continuous Distributions: Normal Distribution, Lognormal Distribution, Gamma Distribution, Weibull Distribution, Beta Distribution - Probability Plots: Normal Probability Plots, Other Probability Plots - Discrete Random Variables: Probability Mass Function, Cumulative Distribution Function, Mean and Variance - Binomial Distribution - Poisson Process: Poisson Distribution, Exponential Distribution - Normal Approximation to the Binomial and Poisson Distributions - More than One Random Variable and Independence: Joint Distributions, Independence - Functions of Random Variables: Linear Functions of Independent Random Variables, Linear Functions of Random Variables That Are Not Independent, Nonlinear Functions of Independent Random Variables - Random Samples, Statistics, and the Central Limit Theorem

D. Decision Making for a Single Sample

(0.5 Days) Statistical Inference - Point Estimation - Hypothesis Testing: Statistical Hypotheses, Testing Statistical Hypotheses, P-Values in Hypothesis Testing, One-Sided and Two-Sided Hypotheses, General Procedure for Hypothesis Testing - Inference on the Mean of a Population, Variance Known - Inference on the Mean of a Population, Variance Unknown - Inference on the Variance of a Normal Population - Inference on a Population Proportion - Other Interval Estimates for a Single Sample - Testing for Goodness of Fit

E. Decision Making for Two Samples

(0.5 Days) Introduction - Inference on the Means of Two Populations, Variances Known - Inference on the Means of Two Populations, Variances Unknown - The Paired t-Test - Inference on the Ratio of Variances of Two Normal Populations - Inference on Two Population Proportions - Completely Randomized Experiment and Analysis of Variance (ANOVA) - Randomized Complete Block Experiment

F. Building Empirical Models

(0.5 Days) Introduction to Empirical Models - Simple Linear Regression: Least Squares Estimation, Testing Hypotheses in Simple Linear Regression, Confidence Intervals in Simple Linear Regression, Prediction of a Future Observation, Checking Model Adequacy, Correlation and Regression - Multiple Regression: Estimation of Parameters in Multiple Regression, Inferences in Multiple Regression, Checking Model Adequacy - Polynomial Models - Categorical Regressors - Variable Selection Techniques

G. Design of Engineering Experiments

(1 Day) The Strategy of Experimentation - Factorial Experiments - 2k Factorial Design: 22 Design, Statistical Analysis, Residual Analysis and Model Checking, 2k Design for k 3 Factors, Single Replicate of a 2k Design - Center Points and Blocking in 2k Designs: Addition of Center Points, Blocking and Confounding - Fractional Replication of a 2k Design: One-Half Fraction of a 2k Design, Smaller Fractions (2kp Fractional Factorial Designs) - Response Surface Methods and Designs: Method of Steepest Ascent, Analysis of a Second-Order Response Surface - Factorial Experiments With More Than Two Levels

H. Statistical Process Control

(1 Day) Quality Improvement and Statistical Process Control - Introduction to Control Charts: Basic Principles, Design of a Control Chart, Rational Subgroups, Analysis of Patterns on Control Charts - R Control Charts - Control Charts For Individual Measurements - Process Capability - Attribute Control Charts: P Chart (Control Chart for Proportions) and nP Chart, U Chart (Control Chart for Average Number of Defects per Unit) and C Chart - Control Chart Performance - Measurement Systems Capability

(iv) Inferential Statistics for Probability Analysis and Testing

Overview

Course ID	2020583
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,900.00 USD	1,800.00 USD	1,900.00 USD
03-05 Aug 28-30 Sep 23-25 Nov	07-09 Sep 02-04 Nov 28-30 Dec	17-19 Aug 12-14 Oct 07-09 Dec

Prices plus local taxes.

Course Description

In statistics, statistical inference is the process of drawing conclusions from data that is subject to random variation, for example, observational errors or sampling variation. Statistical induction helps describing systems of procedures that can be used to draw conclusions from datasets arising from systems affected by random variation, such as observational errors, random sampling, or random experimentation. It is then used to test hypotheses and make estimations using sample data. This training covers all the fundamentals of inductive statistics (probability theory, probability distributions and hypotheses testing) which can be used in marketing, controlling and engineering. You will learn theory and the mathematical foundations in lectures with examples and you will train your new knowledge in practical hands-on labs and exercices.

Course Outline

A. Probability Theory

(0.75 Days) Probability axioms - Probability space Sample space - Elementary event - Random variable - Probability measure - Complementary event - Joint probability - Marginal probability - Conditional probability - Independence - Conditional independence - Law of total probability - Law of large numbers - Bayes' theorem - Venn diagram - Tree diagram

B. Probability Distributions

(0.75 Days) Introduction: Probability mass function, Probability density function, Probability distribution function - Discrete univariate distributions: Binomial, Poisson, Geometric, Hypergeometric - Continuous univariate distributions: Uniform, Exponential, Normal (Gaussian)

C. Frequentist Inference

(0.75 Days) Unbiased estimators (Mean unbiased minimum variance, Median unbiased) - Confidence interval - Testing hypotheses - Alpha-/Beta-Error and Power

D. Specific Tests

(0.75 Days) Z (normal) - Student's t-test - F - Goodness of fit (Chi-squared) - Signed-rank (1-sample, 2-sample, 1-way anova)

(v) Multivariate Analysis I

Overview

Course ID	2020576
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
07-09 Sep 02-04 Nov 28-30 Dec	03-05 Aug 28-30 Sep 23-25 Nov	10-12 Aug 05-07 Oct 30 Nov - 02 Dec

Prices plus local taxes.

Course Description

Multivariate statistics is a form of statistics encompassing the simultaneous observation and analysis of more than one variable. The application of multivariate statistics is multivariate analysis. Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical implementation of multivariate statistics to a particular problem may involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied. This training is one part of a pair of courses on multivariate statistics. It helps you understand the techniques of complex and more advanced data analysis for marketing, controlling and engineering.

Course Outline

A. Multivariate Regression Analysis

(0.5 Days) Determination of a formula that can describe how elements in a vector of variables respond simultaneously to changes in others.

B. Multivariate Analysis of Variance (ANOVA and MANOVA)

(0.5 Days) Comparing multivariate means of several groups using the variance-covariance between variables in testing the statistical significance of the mean differences.

C. Discriminant Analysis

(0.5 Days) Examination whether a set of variables can be used to distinguish between two or more groups of cases.

D. Logistic Regression

(0.5 Days) Prediction of the outcome of a categorical dependent variable based on one or more predictor variables.

E. Factor Analysis

(0.5 Days) Extraction of a specified number of synthetic variables (latent variables or factors), fewer than the original set, leaving the remaining unexplained variation as error.

F. Clustering

(0.5 Days) Assignment of objects into groups (clusters) so that objects (cases) from the same cluster are more similar to each other than objects from different clusters.

(vi) Multivariate Analysis II

Overview

Course ID	2020577
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
1,450.00 USD	1,350.00 USD	1,450.00 USD
14-16 Sep 09-11 Nov	31 Aug - 02 Sep 26-28 Oct 28-30 Dec	07-09 Sep 02-04 Nov

Prices plus local taxes.

Course Description

Multivariate statistics is a form of statistics encompassing the simultaneous observation and analysis of more than one variable. The application of multivariate statistics is multivariate analysis. Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical implementation of multivariate statistics to a particular problem may involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied. This training is one part of a pair of courses on multivariate statistics. It helps you understand the techniques of complex and more advanced data analysis for marketing, controlling and engineering.

Course Outline

A. Introduction to Data Mining

(0.125 Days) Data Mining Functionalities - Classification of Data Mining Systems - Data Mining Task Primitives - Integration of a Data Mining System with a Database or DataWarehouse System - Major Issues in Data Mining

B. Data Preprocessing

(0.125 Days) Descriptive Data Summarization - Data Cleaning - Data Integration and Transformation - Data Reduction - Data Discretization and Concept Hierarchy Generation

C. Mining Frequent Patterns, Associations, and Correlations

(0.5 Days) Basic Concepts - Efficient and Scalable Frequent Itemset Mining Methods - Mining Various Kinds of Association Rules - From Association Mining to Correlation Analysis - Constraint-Based Association Mining

D. Classification and Prediction

(0.75 Days) Issues Regarding Classification and Prediction - Classification by Decision Tree Induction - Bayesian Classification - Rule-Based Classification - Classification by Backpropagation - Support Vector Machines - Accuracy and Error Measures - Evaluating the Accuracy of a Classifier or Predictor: Holdout Method and Random Subsampling, Cross-validation - Model Selection

E. Cluster Analysis

(0.25 Days) Types of Data in Cluster Analysis - Partitioning Methods: k-Means and k-Medoids - Hierarchical Methods: Agglomerative and Divisive Hierarchical Clustering

F. Mining Time-Series and Sequence Data

(0.25 Days) Mining Time-Series Data: Trend Analysis, Similarity Search in Time-Series Analysis - Mining Sequence Patterns in Transactional Databases: Sequential Pattern Mining: Concepts and Primitives, Scalable Methods for Mining Sequential Patterns, Periodicity Analysis for Time-Related Sequence Data

(vii) Oracle and SQL

Overview

Course ID	2022765
Language	en
Duration	3 Days
Delivery mode	Classroom
Course Type	
Target Group	Business Intelligence Developer
Prerequisites	Oracle SQL, PL / SQL
Method	Lecture with examples and exercises.
Course level	Advanced

Course Dates

Chicago	Miami	New York
2,050.00 USD	1,950.00 USD	2,050.00 USD
10-12 Aug 05-07 Oct 30 Nov - 02 Dec	17-19 Aug 12-14 Oct 07-09 Dec	07-09 Sep 02-04 Nov 28-30 Dec

Prices plus local taxes.

Course Description

Oracle developers, and marketing/controlling professionals who have direct access to the Oracle database using SQL or PL/SQL can perform statistical analysis for descriptive statistics and inferential statistics using SQL queries and PL/SQL procedures and PL/SQL functions. This course presents you the numerous functions that are available directly in the Oracle database by making heavy use of scripting examples. The statistical concepts of central tendency, dispersion, correlation and regression, and statistical testing for distribution tests, contingency analysis and the analysis of variance (ANOVA) are also a part of this training.

Course Outline

A. Data analysis using Descriptive Statistics

(0.5 Days) Central tendency: Frequencies using COUNT, mode using STATS_MODE, mean values ??using AVG, MEDIAN - quantiles using PERCENTILE_CONT and PERCENTILE_DISC - Measures of dispersion: range using MIN and MAX, standard deviation using STDDEV, STDDEV_POP and STDDEV_SAMP, variance using VAR_POP, VAR_SAMP and VARIANCE - Rank and distribution using CUME_DIST, DENSE_RANK, RANK, and PERCENT_RANK

B. Correlation analysis

(0.25 Days) Covariance using COVAR_POP and COVAR_SAMP - correlation using CORR (Bravais-Pearson) - rank correlation using CORR_S (Spearman's rho) and CORR_K (Kendall's tau)

C. Regression analysis

(0.25 Days) Linear regression and the least squares method - linear equation derived using REGR_SLOPE and REGR_INTERCEPT - coefficient of determination using REGR_R2 - averages using REGR_AVGX and REGR_AVGY - model check using REGR_COUNT, REGR_SXX, REGR_SYY and REGR_SXY - prediction and residual analysis

D. Contingency

(0.25 Days) contingency and categorical variables - Chi-Square test using CHISQ_OBS and CHISQ_DF - significance using CHISQ_SIG - Contingency: Phi Coefficient using PHI_COEFFICIENT, Cramer's V using CRAMERS_V, Contingency Coefficient using CONT_COEFFICIENT and Cohen's Kappa using COHENS_K

E. Statistical Tests

(0.75 Days) Overview of probability theory - introduction to test theory - t-test using STATS_T_TEST_ONE (one sample), STATS_T_TEST_PAIRED (two samples), STATS_T_TEST_INDEP (two independent samples) and STATS_T_TEST_INDEPU (two independent samples with unequal variance) - variance comparison using STATS_F_TEST - distribution tests using STATS_BINOMIAL_TEST - Mann-Whitney test using STATS_MW_TEST - Kolmogorov-Smirnov function using STATS_KS_TEST - Wilcoxon signed ranks using STATS_WSR_TEST

F. Analysis of Variance (ANOVA)

(0.5 Days) Analysis of Variance - ANOVA performed using STATS_ONE_WAY_ANOVA: Sum of Squares using SUM_SQUARES_BETEEN and SUM_SQUARES_WITHIN, mean squares using MEAN_SQUARES_BETWEEN and MEAN_SQUARES_WITHIN, F-value using F_RATIO and significance using SIG

G. Time series analysis and trend

(0.5 Days) Fundamentals of time series analysis: Components, stationarity, autocorrelation, autocovariance, periodicity - Smoothing: moving average, exponential smoothing - Trend calculations using linear regression - seasonal decomposition and residual analysis

(viii) Structural Equation Modelling

Overview

Course ID	2020582
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,400.00 USD	1,350.00 USD	1,400.00 USD
06-07 Aug 01-02 Oct 26-27 Nov	30-31 Jul 24-25 Sep 19-20 Nov	20-21 Aug 15-16 Oct 10-11 Dec

Prices plus local taxes.

Course Description

Structural equation modelling (SEM) is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions. Structural equation models (SEM) allow both confirmatory and exploratory modeling, meaning they are suited to both theory testing and theory development. Confirmatory modeling usually starts out with a hypothesis that gets represented in a causal model. The concepts used in the model must then be operationalized to allow testing of the relationships between the concepts in the model. The model is tested against the obtained measurement data to determine how well the model fits the data. Among the strengths of SEM is the ability to construct latent variables: variables which are not measured directly, but are estimated in the model from several measured variables each of which is predicted to 'tap into' the latent variables. This allows the modeler to explicitly capture the unreliability of measurement in the model, which in theory allows the structural relations between latent variables to be accurately estimated. Factor analysis, path analysis and regression all represent special cases of SEM.

Course Outline

A. Introduction to Structural Equation Modeling

(0.25 Days) Equivalent models - Steps in performing SEM analysis: Model specification, Estimation of free parameters, Assessment of model and model fit, Model modification, Sample size and power, Interpretation and communication - Advanced uses - SEM-specific software

B. Path Analysis

(0.5 Days) Causality - Latent variable model - Path modeling - Path coefficient - Path tracing rules

C. Causal Analysis using AMOS

(0.75 Days) Analysis of SEM with latent variables (causal analysis) - General modeling and verification process - Construct operationalization - Confirmatory factor analysis for testing reflective measurement models of latent variables (hypothetical constructs) - Testing of hypothesis using the analysis of covariance

D. Variants and Extensions

(0.75 Days) Characteristics of formative measurement models - MIMIC models - Second-order factor analysis (SFA) - multi-group causal analysis and the comparative analysis of causal models in several groups (samples) - Differences between the LISREL approach and the PLS approach - Universal structure modeling

(ix) Time Series Analysis

Overview

Course ID	2020631
Language	en
Duration	2 Days
Delivery mode	Classroom
Course Type	
Target Group	Data Analysts
Prerequisites	General knowledge of math
Method	Lecture with examples and exercises.
Course level	Beginning

Course Dates

Chicago	Miami	New York
1,400.00 USD	1,350.00 USD	1,400.00 USD
17-18 Sep 12-13 Nov	27-28 Aug 15-16 Oct 03-04 Dec	20-21 Aug 22-23 Oct 10-11 Dec

Prices plus local taxes.

Course Description

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. The course provides tools for empirical work with time series data and is an introduction into the foundation of time series models. It focuses on both univariate and multivariate time series analysis. After completing this course, a student will be able to analyze univariate and multivariate time series data using available software like MS Excel, SPSS and jMulti.

Course Outline

A. Univariate analysis of time series data

(0.25 Days) Estimation of the moment-generating functions (expected value, auto-covariance) - auto-correlation: the lag operator, creating and interpretating the correlogram - smoothing of time series data: moving averages, exponential smoothing - transformation and filtering of time series data - first-order and second-order differences

B. Decomposition of time series using deterministic models

(0.5 Days) Component models: additive and multiplicative models - seasonal structures in time series: trend, seasons and identification of the seasonal pattern, prognosis and residual analysis - level shifts - linear, parabolic, logistic, exponential fit and regression of time series - polynomials - quality measures

C. Periodicities in time series

(0.25 Days) Trigonometric functions and their importance for periodic trends - period detection and frequencies - periodogram: identification and interpretation - regression models with periodic oscillations - spectra and spectral density estimation of time series - introduction to Fourier transformation for time series

D. Univariate linear time series models using AR(I)MA

(0.25 Days) Stationarity in time series - White Noise process - AR (Auto Regressive)-models - MA (Moving Average)-models - ARMA and ARIMA models - forecasting - residual analysis - statistical tests for linear time series models - quality measures and model selection

E. Analysis of multidimensional time series

(0.25 Days) Cross-correlation and cross-covariance - stationary cross-covariance - co-integration - introduction to cross-spectral analysis and coherence analysis

F. Multidimensional time series using VAR

(0.25 Days) VAR (Vector AutoRegressive) processes: modeling, prediction, residual analysis, quality measures, tests

G. Time series with exogenous influences

(0.25 Days) Regression with auto-correlated shocks - intervention analysis - transfer function models

b. Disclaimer

Comelio GmbH
Goethestr. 34
13086 Berlin
Germany

- Tel: +49.30.8145622.00
- Fax: +49.30.8145622.10

- www.comelio.com | .de | .at | .ch
- www.comelio-seminare.com
- info@comelio.com
- https://www.facebook.com/comeliogroup
- https://twitter.com/Comelio