

Lausanne, April 2005

## Methodological Training in Statistical Data Mining Related to Drug Development

**Two- plus one-day training course - Monday, June 20 till Wednesday, June 22, 2005**

Given by Dr. Diego Kuonen, Statoo Consulting

Dear Madam, dear Sir

In various presentations of FDA's "Critical Path Initiative" the use of data mining technology and methodology is mentioned as one major approach to optimize various phases of discovery, pre-clinical and clinical research. But how can data mining contribute to achieving operational excellence? Is data mining in drug development worth the trouble or is it "statistical déjà vu"?

This **two- plus one-day training course** will provide you with an overview of the potential and limitations of data mining in drug development and a thorough methodological and practical coverage of state-of-art data mining techniques. It highlights its applicability to accumulated data repositories throughout the drug development life cycle – from discovery through marketing. Moreover, on the third day you will be able to apply what you have learned within a state-of-art data mining workbench using your own or benchmark data.

This **two- plus one-day training course proceeds** from tangible examples to theory and from a conceptual understanding to ability to perform all the specific statistical data mining tasks. You will see a wide range of individual tools in detail and note how they fit into the big picture.

This training is a "must" for those interested in taking drug development to its next level. You can find a detailed description of the training and a registration form on the attached pages or on our homepage – [www.statoo.com/en/PDF/DM\\_Class\\_June\\_2005\\_All.pdf](http://www.statoo.com/en/PDF/DM_Class_June_2005_All.pdf).

We look forward seeing you in Basle.

Please do not hesitate to contact us if you have any questions.

Yours sincerely,



Dr. Diego Kuonen  
CEO  
Statoo Consulting



Cosimo Caforio  
Business Development Manager  
StatSoft (Europe)

# Methodological Training in Statistical Data Mining Related to Drug Development

## Two- plus one-day training course

given by

**Dr. Diego Kuonen**

CEO, Statoo Consulting, Lausanne

### Description

This training course for professionals will provide you with a thorough methodological and practical coverage of state-of-art data mining techniques that identify expected and unexpected trends in data. This course will provide hands-on experience with data mining and throughout the course illustrations of the concepts and methods will be given. Finally, on the third day you will be able to apply what you have learned within a state-of-art data-mining workbench using your own or benchmark data.

### Training

Instruction proceeds from tangible examples to theory - from the big picture, or “whole”, to details, or “parts” - and from a conceptual understanding to ability to perform specific tasks. Consequently, the course begins with a brief discussion of the role and applicability of data mining in drug development to empower companies to extract previously unrealised information about their molecular compounds, product portfolio, clinical studies and customers from their data repositories. Next, a general overview of data mining, the art and science of learning from data, will be given. Only then we do see individual tools in detail and note how they fit into the big picture. Hence, in the main part of this training a **software-vendor independent overview** of the statistical data mining terminology and methods, resources and business issues will be given. For all techniques considered the basic methodology will be explained and illustrated with examples. Finally, the course will end by enabling you to apply the presented methodology and its underlying philosophy to your own or benchmark data.

In summary, the first two days of this three-day course will be organized in lectures covering the methodological aspects and practical applications of statistical data mining. The third day will be dedicated to hands-on practice, where you will have a chance to try on your own the methods learned in the course within a state-of-art data mining workbench provided by StatSoft - the use of which will be explained at the beginning of the third day.

## Outline Methodology – Days one and two

- Introduction
  - Role and applicability of data mining in drug development
- What is data mining?
  - Is data mining 'statistical déjà vu'?
  - What distinguishes data mining from statistical analysis?
  - Two analytical approaches to data mining
  - The knowledge discovery in databases (KDD) process
  - Data mining and business intelligence
- Data and data pre-processing
  - Data sources
  - Why data pre-processing?
  - Major tasks in data pre-processing
- Data mining techniques (unsupervised and supervised)
  - Description and visualisation
  - Characterising multivariate data
  - Dissimilarity and distance measures
  - Unsupervised methods (class discovery)
    - Principal component analysis
    - Multidimensional scaling
    - Cluster analysis (hierarchical algorithms, partitioning algorithms, using clustering in practise)
    - Kohonen's self-organizing maps
    - Affinity grouping or association rules
    - A look forward
  - Supervised methods (class prediction)
    - Classification modelling (e.g. discriminant analysis, nearest neighbour and Bayes classifiers, support vector machines)
    - Regression modelling (e.g. multiple linear models, generalized linear models, nonparametric regression models)
    - Neural networks
    - Tree-based methods (e.g. CART, C4.5 and C5.0, CHAID)
    - Ensemble learning (e.g. bagging, arcing, random forests, boosting)
    - The curse of dimensionality (feature extraction, feature selection, wrappers, embedded methods)
    - Evaluating and comparing methods
    - A look forward
- Data mining myths versus realities
- Conclusion
- References and resources

## Outline Data Mining Hands-on Practice – Day three

- Introduction to *STATISTICA* Data Miner
- Overview of its functionality and integrated data mining process
- Hands-on practice by applying the presented methodology and its underlying philosophy to your own or benchmark data
- Questions and answers

## About the speaker

Diego Kuonen, PhD in Statistics, is founder and CEO of Statoo Consulting, Lausanne, Switzerland ([www.statoo.ch](http://www.statoo.ch)). Statoo Consulting is a software-vendor independent Swiss consulting firm specializing in statistical consulting and training, data analysis, data mining, analytical CRM and bioinformatics services. Statoo Consulting's clients include companies like **Nestlé Research Center**, **Novartis Pharma**, **Novo Nordisk** and **Serono**. Dr. Diego Kuonen has several years of experience in teaching and training, and he is, also lecturer at the "Swiss Federal Institute of Technology Lausanne" and the "Swiss Federal Institute of Technology Zurich" as well at other Swiss universities. Currently, he is president of the organising committee of the "Swiss Statistics Meeting", vice president of the "Swiss Statistical Society" and president of its section "Statistics in Business and Industry".

## Prerequisites

Participants should be familiar with basic statistics, including multiple linear regression. Participants without a statistical background will be able to listen and follow the speaker in the technical discussions.

For the third day a laptop (MS-Windows XP/2000) with pre-installed *STATISTICA* Data Miner course-license. StatSoft will provide this license before the course begins.

## Presentation

The lecture will be given in English.

## Date and hour

Monday, June 20 till Wednesday, June 22, 2005 starting 08.30 AM till 5.00 PM

## Place

UBS AG, Training- and Conference Center, Viaduktstrasse 33, CH-4052 Basle, Switzerland - see [www.ubs.com/1/e/abz\\_basel.html](http://www.ubs.com/1/e/abz_basel.html)

## Course fee

Commercial:

- 3 days **SFR 2'400.00** (approx. € 1'520.00)
- 2 days - including days one and two - **SFR 1'800.00** (approx. € 1'140.00)

Academic:

- 3 days **SFR 1'600.00** (approx. € 1'010.00)
- 2 days – including days one and two - **SFR 1'200.00** (approx. € 760.00)

Prices include documentation and *STATISTICA* Data Miner course-license, coffee breaks and lunch but not VAT. Group discounts are available. All participants will receive an attendance certificate.

## Registration

See separate registration form or [www.statsoft.de/akt\\_training2.html](http://www.statsoft.de/akt_training2.html).

## Contact Information

For further information about the training please contact Cosimo Caforio, Phone +41 56 631 31 10 and Fax +41 56 631 31 12 or Email [dmdd@statoo.com](mailto:dmdd@statoo.com).

**Registration form for two- plus one-day training course**  
*Methodological Training in Statistical Data Mining Related to Drug Development*  
 given by Dr. Diego Kuonen, Statoo Consulting

To register please fill out this form completely and fax it to **+41 56 631 31 12** or register online at [www.statsoft.de/akt\\_training2.html](http://www.statsoft.de/akt_training2.html).

\* Required Information

First Name*	
Last Name*	
Company*	
Department/Function*	
Address*	
ZIP/Postal Code*	
City*	
Country*	
Phone*	
Fax	
Email*	
Date and Signature*	
Comments	

- SFR 2'400.00 Commercial – Tree days**  
 **SFR 1'800.00 Commercial – Days one and two**  
 **SFR 1'600.00 Academic – Three days** - Please attach a copy of your certification.  
 **SFR 1'200.00 Academic – Days one and two** - Please attach a copy of your certification.

**Terms and Conditions**

Prices include documentation, *STATISTICA* Data Miner course license, coffee breaks and lunch but not VAT. Payment of the course registration fee is required prior to the start of the course. Cancellations received in writing more than 30 days before the start of the course will be refunded 100% of the course fee. Cancellations received between 30 and 14 days prior to the course will be refunded 50% of the course fee. We regret that no refunds are allowed for cancellations received within 14 days of the course start date.

**Contact Information**

For further information please contact Cosimo Caforio, Phone +41 56 631 31 10 and Fax +41 56 631 31 12 or Email [dmdd@statoo.com](mailto:dmdd@statoo.com).