

# THE 2010 CATALOGUE OF TRAINING COURSES IN STATISTICS AND DATA ANALYSIS

Basic Concepts  
in Statistics

*Analysis of Variance*

multifactorial  
data:

- Description and analysis
- Decision and prediction

Small  
Samples  
Statistical  
Processing

The Measure of risk  
factors and control of  
bias

*Introduction to the R software*

The survey : construction and  
analysis of questionnaires

Statistical Analysis of  
Open-ended Questions  
in Surveys

**Statistics and Data Analysis  
Training - Studies - Consulting**

**VIES ANIMALES**

**Les Vigneaux**

**37220 RILLY SUR VIENNE**

**☎ 02 47 58 70 33**

**[caroline.vieulle@free.fr](mailto:caroline.vieulle@free.fr)**

**[www.anastats.fr](http://www.anastats.fr)**

 **AnaStats.fr**

# We offer training courses in statistics that are unusual

## Three important special features:

### 1. They are given by USERS of statistics

- Coming from experimental biological sciences, the trainers are boosted by many years of statistic's applications in their research field. In collaboration with eminent statisticians, they searched for statistical solutions for experimentalist's problems. During more than ten years they trained researchers, engineers and technicians of industry, CNRS, INRA and INSERM, whose *they share the language and the concerns*.
- Over the years they refined rigorous training courses, and nevertheless that can be understood without special ability in mathematics.

### 2. They focus on EFFICIENCY

- Choosing the most suitable method or test is essential before any analysis; however this is often a problem for many people. Participants will be soundly trained by a lot of exercises and explanations.
- The knowledge of statistical methods is not sufficient: participants must *understand the software's outputs*. They are scrutinized in a very detailed way. The computing tool comes down to everybody's level.
- Practical exercises are carried out on computer by the trainees themselves, allowing a first familiarization with the software and a confrontation with the concrete problems encountered during the analyzes.
- The courses are made in interaction with the participants, listening to their request. Examples are brought by the participants themselves, starting from their daily preoccupations.

### 3. MODERN APPROACHES are implemented

The power of computers allowed the development of new statistical methods, well adapted to the concrete situations met by the users: validations of the results by re-sampling, calculations adapted to small samples, exact tests of permutations, exploration of multifactorial data.

## **A range of nine independent courses**

### **1. Basic Concepts in Statistics**

- Three days open courses in Tours  
January 27 - 29 / 2010
- Three days intra-firm, max. 12  
participants.

### **2. Small Samples Statistical Processing**

- Three days open courses in Tours  
March 10 – 12 / 2010
- Three days intra-firm, max. 12  
participants.

### **3. Analysis of variance**

- Three days open courses in Tours  
March 31 - April 02 / 2010
- Three days intra-firm, max. 12  
participants.

### **4. Description and analysis of multifactorial data**

- Three days open courses in Tours  
May 05 - 07 / 2010
- Three days intra-firm, max. 12  
participants.

### **5. Decision and prediction with multifactorial data : discriminant analysis and regressions**

- Three days open courses in Tours  
June 16 - 18 / 2010.
- Three days intra-firm, max.12  
participants.

### **6. Introduction to the R software**

- Three days open courses in Tours  
February 03 - 05 / 2010 or  
September 15 - 17 / 2010.
- Three days intra-firm, max. 12  
participants.

### **7. The Measure of risk factors and control of bias**

- Two days open courses in Tours  
October 14 - 15 / 2010.
- Two days intra-firm, max. 12  
participants.

### **8. Survey: construction and analysis of questionnaires**

- Three days open courses in Tours  
November 24 - 26 / 2010.
- Three days intra-firm, max. 12  
participants..

### **9. Statistical Analysis of Open-ended Questions in Surveys**

- Two days open courses in Tours  
December 09 - 10 / 2010.
- Two days intra-firm, max. 12  
participants.

*Courses 8 and 9 intend more for  
sociologists having survey data to  
collect and to analyse.*

The trainings are given by approved trainers

See the detailed programme on the following pages

**In addition, if you have data which you cannot analyze but you do not  
have time to follow a training course, we also have  
a “STUDIES” department**

## 1. Basic Concepts in Statistics

- **Audience**

Any person having to treat data, to characterize or compare samples, to analyze results of experiments, but lacking in basic knowledge in statistics.

- **Objectives**

To be able to describe a series of values. To pose problems which could be solved by statistics in a correct manner To choose and implement the statistical test adapted to each situation. To interpret the corresponding software outputs.

- **Method**

In each heading, the principles of calculation are first given, without recourse to complex mathematical concepts, giving priority to an intuitive comprehension. Limits of techniques and requirements to use them are laid down. Each one put into practice by using the statistical functions of the Excel software (a computer by participant). Description and interpretation of the software outputs are seen together Illustrations of each technique are given. Concrete situations met by the trainees are examined and discussed.

### Programme

- ✓ Function of statistics: description and decision.
- ✓ Characterisation of a sample: distribution, central values and dispersion.
- ✓ Comparison of several series of quantitative measures: different scales of measurement and matching tests; one- and two-sided hypothesis; risk and sample size; power of a test; comparison of two samples; introduction to the analysis of variance.
- ✓ Measuring the relationship between quantitative variables: correlation and regression.
- ✓ Measuring the relationship between qualitative variables : the Chi-square test.

**Duration** : three days

**Open course** :

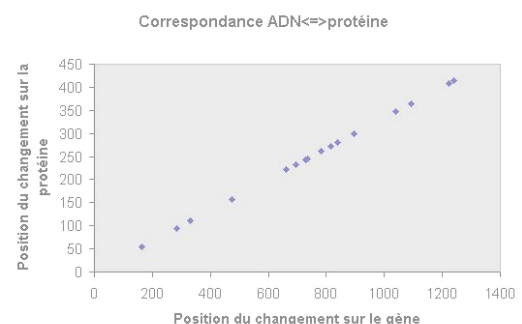
Rate : 669 € H.T. per person  
Max. 6 participants

**Dates** : january 27 – 29 / 2010

**Intra-firm** :

Rate : 3,500 € H.T.

*A review of the essential basic concepts to approach, understand and use the various statistical methods*



## 2. Small Samples Statistical Processing

- **Audience**

Technicians, engineers or researchers confronted with the statistical treatment of small samples. Basic knowledge in statistics is very desirable.

- **Objectives**

Description of a short series of data. To choose a test adapted to the questions and to the data.

To compare two or more samples, to measure the relationship between several variables.

To implement these tests and interpret the software's outputs.

- **Method**

For each question, the course is divided in four parts:

1. Principle of available tests. Choice of the solution adapted to the situation.
2. Computing method.
3. Power and validity conditions.
4. Software solutions, implementation and interpretation of the results.

### Programme

#### ✓ One day of theory

Reminder of the functions of statistics – Discussion of the sample size – Parameters of a small sample – Different measurement scales – How to choose a risk threshold and the power of a test – Independent and related samples – Choice of a non-parametric test – "Manual" computation of some tests – Principle of exact tests.

#### ✓ Two days of practice

Carrying out numerous tests with the 'StatXact' software and with utilities using Excel: choice of the test, implementation, interpretation of outputs. Discussions starting from the concrete situations presented by the participants; work possibly on their data sets. Presentation of the available tools in Excel.

***The landscape of statistics is largely modified by modern methods of statistical processing. Recent techniques make it possible to approach easily and surely the analysis of small samples.***

**Duration** : three days

**Open course** :

**Rate:** 669 € H.T. per person

*Max. 6 participants*

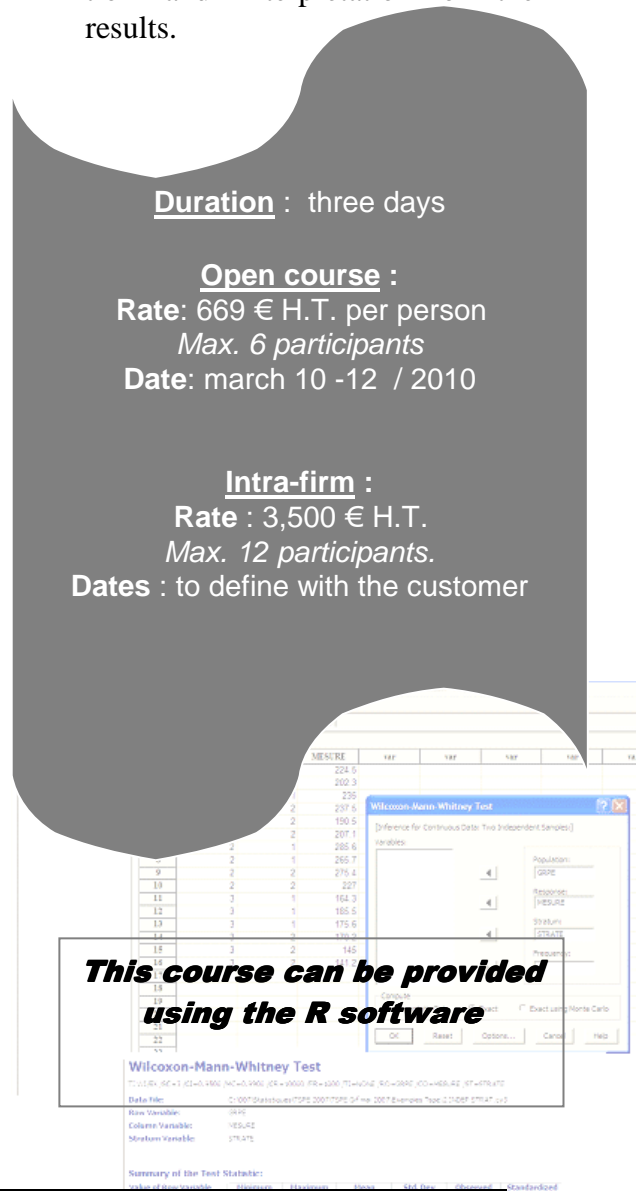
**Date:** march 10 -12 / 2010

**Intra-firm** :

**Rate :** 3,500 € H.T.

*Max. 12 participants.*

**Dates :** to define with the customer



***This course can be provided using the R software***

2-Selected Point Pro  
0.1569  
0.2008 0.038

### 3. Analysis of Variance

- **Audience**

Any person having to compare samples, to analyze experimental results, but lacking a good control of the methods of analysis of variance. Basic knowledge in statistics is very desirable.

- **Objectives**

To be able to choose the right method matching with the data and the issues. To be able to control the conditions of use. To be able to prepare the data and to use an adapted software. To interpret the results.

- **Method**

Computation methods are explained without complex mathematical knowledge. We alternate between theoretical questions and practice with the XLSTAT software.

The trainees carry out many applications starting from very diverse experimental situations (a computer by trainee). The outputs are discussed and interpreted with the help of charts.

#### Programme

- ✓ General principle of analysis of variance.
- ✓ Validity conditions of the ANOVA.
- ✓ One-way ANOVA for independent or related data.
- ✓ Factorial analysis.
- ✓ Multivariate analysis of variance (MANOVA).
- ✓ The General Linear Model and covariance analysis.

***This course can be provided using the R software***

**Duration** : three days

**Open course** :

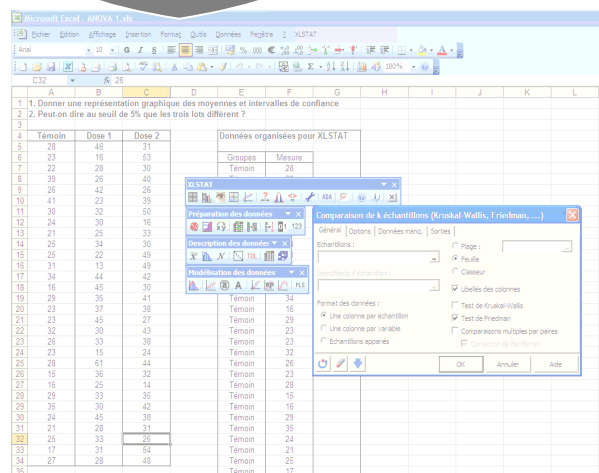
**Rate:** 669 € H.T. per person  
**Max. 6 participants**

**Dates** : March 31 - April 02 / 2010.

**Intra-firm** :

**Rate** : 3,500 € H.T.  
**Max. 12 participants.**

**Dates** : to define with the customer



## 4. Description and analysis of multifactorial data

### • Audience

Technicians, engineers or researchers confronted with large data sets and analysing complex relationships between numerous variables. Basic knowledge in statistics is not necessary.

### Programme

- ✓ Different measurement scales.
- ✓ Quantitative data: Principal Component Analysis. Principle and implementation. Numerous examples of applications. Help to interpret the results of a PCA.
- ✓ Two categorical variables: correspondence analysis. Principle and implementation. Examples.
- ✓ Numerous categorical data: Multiple Correspondence Analysis. Principle, implementation and interpretations.
- ✓ Multiple Factorial Analysis.
- ✓ Clustering techniques. Interest and limits. Optimal use.

### • Objectives

Facing a complex problem, how to organize the data and to build analysable tables. To choose the technique adapted to its data and its questions. How to use a software of multifactorial analysis. How to interpret the results.

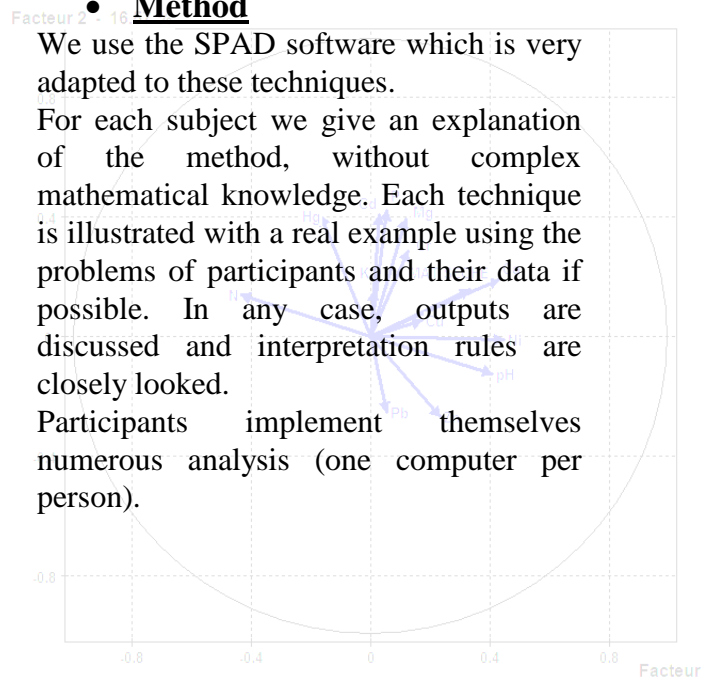
*Users of multifactorial methods to analyse large tables are increasingly numerous. The statistics make it possible today to explore these tables and to analyze processes influenced by several factors.*

### • Method

We use the SPAD software which is very adapted to these techniques.

For each subject we give an explanation of the method, without complex mathematical knowledge. Each technique is illustrated with a real example using the problems of participants and their data if possible. In any case, outputs are discussed and interpretation rules are closely looked.

Participants implement themselves numerous analysis (one computer per person).



Duration : three days

#### Open course :

Rate: 669 € H.T. per person

Max. 6 participants

Dates : May 05 – 07 / 2010

#### Intra-firm :

Rate : 3,500 € H.T.

Max. 12 participants.

Dates : to define with the customer

## 5. Decision and prediction with multifactorial data: discriminant analysis and regressions

- **Audience**

Technicians, engineers or researchers confronted with the thorough analysis of large tables of numerical data, interested by the relationships between variables and the explanation of variability, looking at the best variables to explain a process and to perform forecasts.

- **Objectives**

To choose and implement modern methods of detailed analysis of complex data. To use adapted software and to interpret the outputs.

- **Method**

XLSTAT and Tanagra software are used during this course. For each subject we give an explanation of the method, without complex mathematical knowledge. Each technique is illustrated with a real example using the problems of participants and their data if possible. In any case, outputs are discussed and interpretation rules are closely looked at. Participants implement themselves numerous analysis (one computer per person).

### Programme

- ✓ Methods of forecasting : linear discriminant analysis, non parametric discriminant analysis using binary decision trees.
- ✓ Univariate and multivariate linear regression, logistic regression, PLS and PLS-DA regression methods.

**Duration** : three days

**Open course** :

**Rate**: 669 € H.T. per person

*Max. 6 participants*

**Dates** : June 16 -18 / 2010

**Intra-firm** :

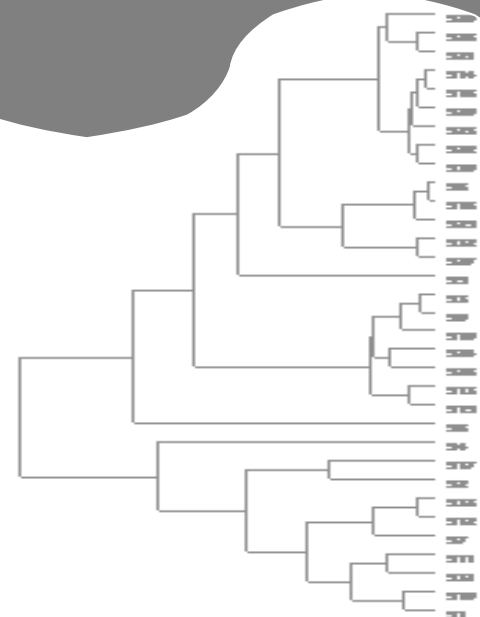
**Rate** : 3,500 € H.T.

*Max. 12 participants.*

**Dates** : to define with the customer

***Data Mining tools will enable you to extract information from your data.***

***A wide range of modern multifactorial methods of large table's analysis now enables to carry out forecasts using all known information and bring a decision-making aid.***



## 6. Introduction to the 'R' software

- **Audience**

Technicians, engineers or researchers confronted with statistical analysis. Basic knowledge in statistics is necessary

- **Objectives**

The purpose of this course is to initiate the participants into the use of the software, in support with a large number of current applications.

- **Method**

Numerous examples are implemented by trainees (one computer per person). Outputs are discussed and interpreted.

### Programme

- ✓ Presentation of the "R project". How to download and to install the software.
- ✓ Basic concepts; how to use the 'Help' and the documentation available on the web.
- ✓ Basis of the 'S' language. Simple computations in the R console.
- ✓ Managing 'objects' in the memory: data manipulation and recording.
- ✓ The 'Rcmdr' interface: loading files, data managing, graphical functions, simple statistical tests.
- ✓ Statistical analysis with Rcmdr: non-parametric tests, analysis of variance and linear regressions, general linear model.
- ✓ Multifactorial analysis with R: the 'FactoMineR' library and his graphical interface. Implementation of PCA, Correspondence Analysis, clustering, partition and description of classes. Implementation of a factorial discriminant analysis.
- ✓ Introduction to graphical functions: usual graphical commands are implemented by the participants.
- ✓ Basics of programming in R.

**Duration** : three days

**Open course:**

Rate: 669 € H.T. per person  
Max. 6 participants

**Dates** : February 03 - 05 / 2010  
or September 15 - 17 / 2010

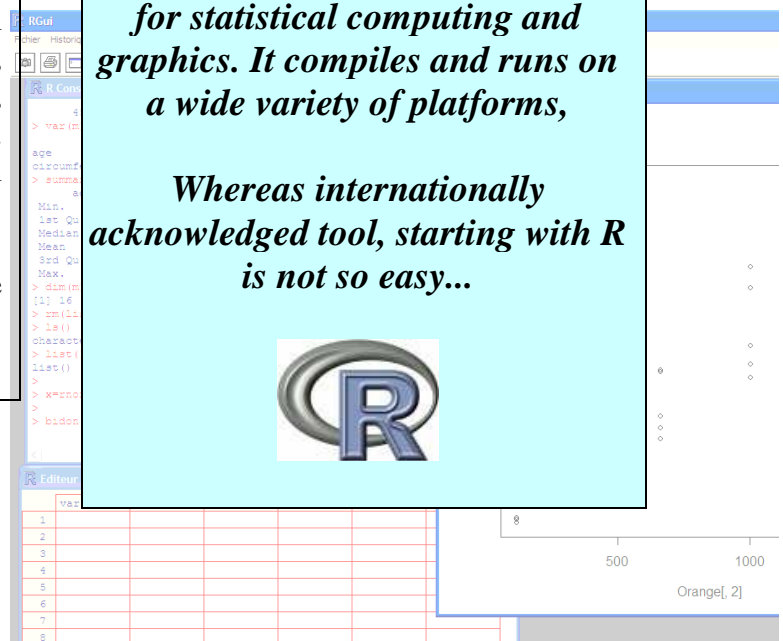
**Intra-firm:**

Rate : 3,500 € H.T.  
Max. 12 participants.

**Dates** : to define with the customer

***R is a free software environment  
for statistical computing and  
graphics. It compiles and runs on  
a wide variety of platforms,***

***Whereas internationally  
acknowledged tool, starting with R  
is not so easy...***



## 7. The Measure of risk factors and control of bias

- **Audience**

Any person having to collect and analyze data whose sampling is difficult.

- **Objectives**

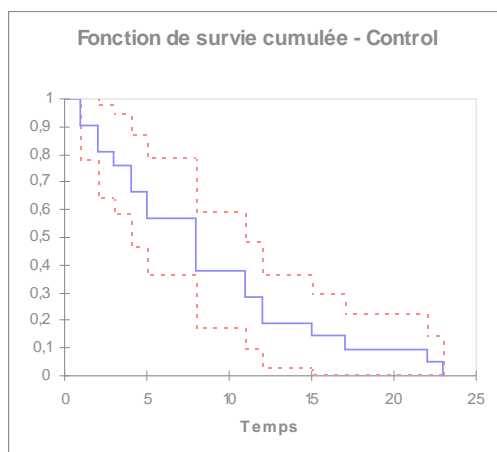
To be able to identify and measure the risk factors of an event. To avoid bias during an investigation. To use and interpret results of uni- and multi-variate analysis. To discuss the main types of surveys, particularly epidemiologic. To know the methods of analysis aiming at the improvement of the statistical power.

- **Method**

For each subject we give an explanation of the method, without complex mathematical knowledge. Each technique is illustrated with a real example. In any case, outputs are discussed and interpretation rules are closely looked. Participants implement themselves numerous analysis (one computer per person).

### Programme

- ✓ Special features of surveys data: cohort, case-control, cross-sectional.
- ✓ Measure of risk and measure of association: prevalence, incidence, relative risk, odds ratio.
- ✓ Bias: sampling bias, classification bias, confusion factors.
- ✓ Estimation and test methods, Bayes's theorem, predictive values.
- ✓ Multivariate analysis, standardisation, adjusting method, general principles
- ✓ Logistic regression
- ✓ Cox Model



**Duration:** Two days

**Open course:**

Rate: 446 € H.T. per person

Max. 6 participants

Dates : October 14 – 15 / 2010

**Intra-firm:**

Rate : 2,334 € H.T.

Max. 12 participants.

Dates : to define with the customer

***Basic principles of data analysis to avoid bias and for risk estimate.***

## 8. The survey: questionnaire design and analysis

- **Audience**

Any researcher having to carry out surveys and/or to carry out the statistical analysis of the results. Basic knowledge of statistics is desirable.

- **Objectives**

To choose a method of survey and to build a questionnaire. To organize and control the quality of the data. To choose and to use statistical tools and to interpret the results.

- **Method**

The Excel spreadsheet and the SPAD software are used. Participants study real examples. Outputs are discussed and interpreted. One computer per participant.

### Programme

- ✓ To choose a method: interviews, opened, closed or mixed questionnaires, according to objectives and sampling.
- ✓ Design of the questionnaire: methodology of survey.
- ✓ Sampling methods: random, systematic, stratified, convenience, judgment, quota or snowball sampling. Sampling mistakes.
- ✓ Recoding, adjusting and data quality control.
- ✓ Data analysis:
  - Cross-tables, chi-square test
  - Factor analysis: active and supplementary variables, methods of description, of decision and prevision and of opened questions analysis.
- ✓ Synthesis: computation of indexes (satisfaction index...), layout of results.

**Duration** : three days

**Open course**:  
 Rate: 669 € H.T. per person  
 Max. 6 participants

**Dates** : November 24 - 26 / 2010

**Intra-firm**:  
 Rate : 3,500 € H.T.  
 Max. 12 participants.  
 Dates : to define with the customer

The screenshot shows a questionnaire titled 'Connaissance du programme' with several questions and radio button options. The questions are numbered 1 through 5. The options include 'Oui', 'Non', 'Je ne sais pas', 'C'est compliqué', 'C'est facile', 'C'est difficile', 'C'est intéressant', 'C'est ennuyeux', 'C'est utile', 'C'est inutile', 'C'est amusant', 'C'est pas amusant', 'C'est utile', 'C'est pas utile', 'C'est intéressant', 'C'est pas intéressant', 'C'est utile', 'C'est pas utile', 'C'est intéressant', 'C'est pas intéressant'.

***The methods of survey are varied and the statistical analysis of the results can be very interesting if one knows the modern adapted tools.***

## 9. Statistical analysis of opened questions in inquiries

- **Audience**

Technicians, engineers or researchers confronted with statistical analysis of textual data. Basic knowledge in multivariate statistics is desirable.

- **Objectives**

To collect and describe natural language textual data. To analyse the relations between words and categorical variables. To find similarities between groups of repliers on the basis of their answers.

- **Method**

The SPAD software is used because it includes very efficient tools for textual analysis and numerous methods of multifactorial analysis are associated.

The different steps of the analysis are explained using a real-size example. Practice by the participants follows the explanations (one computer per person). Outputs are then discussed and interpreted. The full analysis of an opened question is then performed by the participants, this allowing self evaluation of knowledge.

Important times are spared for the discussion of the concrete situations met by the trainees.

### Programme

- ✓ Preparation (lemmatization) of the text before analysis.
- ✓ Description of the vocabulary: recurrent segments, concordances, most significant words and sentences.
- ✓ Cross-table 'words x repliers groups'. Description of the table by correspondence analysis and clustering.
- ✓ Cross-table 'individuals x words'. Looking for sub-groups of repliers on the basis of their answers. Co-occurrences (associations of ideas). Textual discriminant analysis.

Type d'Établissement

Taux de réponse : 100.0%

Lycée	17	20.5%
Lycée Professionnel	11	13.3%
Collège	55	66.3%
Total	83	100.0%

Indiquez le ou les matériels dont vous disposez dans votre établissement.

Taux de réponse : 100.0%

Somme des pourcentages différente de 100 du fait des réponses multiples et des suppressions.

Appareil photo numérique	58	69.9%
Camescope	58	69.9%
Ordinateur de bureau	68	81.9%
Ordinateur portable	31	37.3%
Ordinateur réservé à l'E.P.S.	38	45.8%
Téléviseur pour l'E.P.S.	30	36.1%
Assistant numérique (Palm - Pocket PC ...)	5	6.0%
Scanner	38	45.8%
Lecteur de DVD	35	42.2%
Imprimante	69	83.1%

**Duration:** Two days

**Open course:**

Rate: 446 € H.T. per person

Max. 6 participants

Dates : December 09 – 10 / 2010

**Intra-firm:**

Rate : 2,334 € H.T.

Max. 12 participants.

Dates : to define with the customer

***Specialized software allow today the automatic analysis of the natural languages. The multivariate methods allow a thorough exploitation of these data, open questions in the investigations, talks, technological survey...***

## **The « STATISTICAL STUDIES » department**

**Numerous researchers, engineers, manufacturers are confronted with complex data, from which interesting information could be extracted. If you are lacking time or competence, our Statistical Studies Department can help you.**

### **DOMAIN**

Because we use the most performing tools of data mining we are able to perform this complex studies:

- ✓ What variables are the most sensitive to a treatment;
- ✓ How are the performances correlated ?
- ✓ How can we explain the variability of this process?
- ✓ How can a given sub-group of individuals be characterised?
- ✓ What parameters of a manufacturing process can explain a malfunctioning?
- ✓ Detailed analysis of enquiries;
- ✓ Statistical analysis of open questions and relationships between answers and characteristics of sub-populations;
- ✓ Technological survey.

### **COMMUNICATION**

Listening your request which is not always formulated in precise statistical terms, our reports are easy to use, readable by nonspecialists. They are the result of our great experience in this field, with partners from various origins (industry, scientific research, trade, sociology ...)

**You have data,  
do not lose the results of your investigations,  
do not forget them in the bottom of a drawer,  
do not hesitate to contact us !**

## Training team

### Doctors of Biology

University lecturers and researchers in biology and data analysis.

#### **Some of our customers for training courses, data studies and consulting:**

- INSERM, CNRS, INRA, IFREMER, AFSSAPS, Institut Pasteur, several universities and Grandes Ecoles.
- Companies: L'Oréal, Sanofi~Aventis, Danone, Novartis, Wella, Veolia, BorsodChem (Hungary)...
- LE FOYER insurances (Luxembourg).

#### **Contribution to the development of Data Mining software.**

#### Contact :

**Caroline VIEUILLE**  
**Les Vigneaux**  
**F - 37220 Rilly Sur Vienne**  
Tél. 33 - (0)2 47 58 70 33  
E.mails : [caroline.vieuille@free.fr](mailto:caroline.vieuille@free.fr)  
[severine.montaudouin@neuf.fr](mailto:severine.montaudouin@neuf.fr)  
[sandrine\\_longis@yahoo.fr](mailto:sandrine_longis@yahoo.fr)  
[gilles.lepape@neuf.fr](mailto:gilles.lepape@neuf.fr)

[www.anastats.fr](http://www.anastats.fr)

