# STID
## Statistics and Business Intelligence

## IUT Roubaix

## Lille 2 University

## France

**Sylvia CANONNE**

**Description of teaching modules. September 2014**

# Course descriptions subject to change

<u>**Term 1**</u>

| M1101A -Mathematics | **Mathematics Elective** | term 1 – Coefficient 2 – 30h |
|---|---|---|
| With this module (aimed at those from non-scientific background) the students' level of Mathematics will be brought to the required level needed for this University's training. The content of the course enables students to study functions and to calculate integrals. | | |

| M1101A -Economics | **Economics Elective (only for those who did not study Economics in high school)** | term 1 – Coefficient 2 – 30h |
|---|---|---|
| The aim of this class is to introduce students to economics and provide them with common knowledge in the domain. The first part of the course aims at defining economics and the role of economists in society. History of economic thought will be addressed. Micro and macroeconomics will be introduced and studied by the means of case studies. | | |

| M1102 - Statistics | **Univariate Statistics** | term 1 – Coefficient 2 – 30h |
|---|---|---|
| Provide students with the required tools to describe a variable distribution of one or more populations. Use of spreadsheet programs (and their specialized macros) for the exploration, presentation and calculation of statistical indicators, the visualization of categorical and quantitative data in the form of appropriate tables and graphs. | | |

| M1103 –Project | **Personal and Professional Project** | term 1 – Coefficient 1- 20h |
|---|---|---|
| PPP sessions are designed to accompany students in the elaboration of a project which deals with both their motivations and skills and the perspectives offered by the professional world. To help students get a better grasp of the job market, students elaborate job descriptions, visit companies and attend as well as organize conferences to which actors of the professional world participate. | | |

| M1201 - Statistics | **Bivariate Statistics** | term 1 – Coefficient 2 – 30h |
|---|---|---|
| Provide students with the tools necessary for the description of the distribution of two variables observed in one same population.<br><br>Use of spreadsheet programs (and their specialized macros) for the presentation and visualization of data by crossing two variables (categorical and/or quantitative) in the form of double entry tables and graphs and the calculation of appropriate measures of association. | | |

| M1202- Statistics | **Probability and Simulation 1** | term 1 – Coefficient 3 – 40h |
|---|---|---|

The purpose is to consolidate the concept of univariate random variable. Beginning with the background theory of sets, the notions of random event, inclusion, disjunction and independence are thoroughly studied and discerned in the context of conditional probabilities. The discreet and continuous random variables are then carefully designed through simulations provided by software such as R and Excel. The density and the distribution functions are intensively used and graphically illustrated. Everything is systematically provided with numerous concrete examples.

| M1203 - Statistics | **Statistical surveys and case studies** | term 1 – Coefficient 2 – 30h |
|---|---|---|

Acquisition of the methodology needed to carry out a statistical study implementing the basic methods of descriptive statistics on a real life case (simple survey type) and to present its results.

Use of specialized software (survey study, spatial data)

| M1204- Mathematics | **Mathematics for Probability and Statistics** | term 1 – Coefficient 3 – 45h |
|---|---|---|

Continuing from M1202, couples of random variables are introduced both in the context of discreet and continuous random variables. The links with statistical bivariate analysis are underlined. Expanding the study to random vectors, the operations on multiple random variables are studied. Sum of independent random variables are used to reach the building of some usual laws of random variables and some important results as the law of large numbers and the central limit theorem are finally put in evidence.

| UE13/M1301- Computer Science | **Basics of Programming** | term 1 – Coefficient 3 – 45h |
|---|---|---|

This course covers the basics of computer science and programming. It aims at providing students with an understanding of the role computation can play in solving problems. It first covers the common data structures and algorithms used to solve these problems. Topics include variables, arrays, conditional statements, loops, functions, etc. It also focuses on planning and organizing programs, as well as the grammar of the VBA programming language.

| UE13/M1302 – Computer Science | **Data  Exploitation** | term 1  – Coefficient 3 – 40h |
|---|---|---|

In this course, students first learn to understand the way a data file (spreadsheet) is structured and how to exploit its content using spreadsheet program functions. Then, they are introduced to the foundations of relational databases. The relational data model and the relational algebra are presented as well as the SQL (Structured Query Language) syntax for accessing databases. The DBMS (Databases Management Systems used are ACCESS and/or MySQL).

| UE13/M1303 – Computer Science and Management | **Reporting Tools 1** | term1 – Coefficient 3 -45h |
|---|---|---|

This modules aims at helping students :

- Get acquainted with the professional world, with companies in general.
- Understand financial and accounting information systems by analyzing documents such as balance sheets and income statements
- Learn about accounting and financial reporting

| UE14/M1401- Economics and Management | **General Economics and company organisation** | term1 – Coefficient 2 – 35h |
|---|---|---|

This course is organized along two main themes. The first one is the origin of different existing economic information measurements (unemployment rate, Gross National Product, etc.). The second one is an introduction to company organization.

| UE 14 / M1402 - Expression | **Basis of Communication** | term1 – Coefficient 2 – 30h |
|---|---|---|

The aim of the class is to explain how communication functions on a theoretical and practical level and the functions of language in order to become more comfortable in writing and orally and to see how language is used in surveys and polls.

After having discovered the theories of communication, students learn how to better communicate in their domain, by adapting their discourse to their public and to the nature of their support (oral or written). They will see how an email is written, how a questionnaire is worded, how to transform figures into words or in graphs in a statistical report. They will also have to know how to make a statistical presentation using PPT.

| UE 14 / M1403- English | **Introduction to English for Special Purposes** | term1 – Coefficient 2 – 30h |
|---|---|---|

Students will become aware of the existence of the specific English used in their scientific field and will develop their capacity to communicate using it (orally and in writing).

They will be brought to present tables and graphs and comment on statistical studies.

**Term 2**

| UE 21 / M2101 -Statistics | **Introduction to inferential statistics** | Term 2 – Coefficient 2,5 – 45h |
|---|---|---|
| This course gives a first introduction to the sampling distributions of a mean, a variance or a proportion when calculated on Gaussian samples. The Chi-square, Student and Fisher distributions are defined and then applied to the construction of the classical parametric one-sample and two-sample statistical tests. | | |

| UE 21 / M2102 -Statistics | **Curve fitting and time series** | Term 2 – Coefficient 2 – 30h |
|---|---|---|
| In this class, students will be provided with the necessary tools for curve fitting (linear adjustment and adjustment of curves with two quantitative characteristics) and for time series description (smoothing of addition and multiplication decomposition models so as to calculate trends and components, seasonally adjusted series, exponential smoothing – (single and double). | | |

| UE 21 / M2104 - Mathematics | **Mathematics for data analysis** | Term 2 – Coefficient 3 – 60H |
|---|---|---|
| To prepare data analysis, the basic concepts of linear algebra are introduced. Matrix tools are used to study linear systems, diagonalization of matrices and orthogonality in conjunction with formal computing software and dynamical geometric software. Examples are extracted from concrete and real situations. | | |

| UE 22/M2201– Computer Science | **Software engineering and Web technologies** | Term 2 – Coefficient 2,5 -45h |
|---|---|---|
| This module is twofold. | | |
| One course introduces the basis of software engineering methods. Starting from an analysis of requirements, students are introduced to analytical techniques and processes essential for specifying, designing and implementing a software system. | | |
| One course gives students an insight into web technologies. Aspects covered include: historic development of the Web; client server architecture; protocols; content markup (XHTML and CSS) and JavaScript and PHP programming activities. | | |
| This module includes by a case study to specify, design and implement a simple web application. | | |

| UE 22/M2202 – Computer Science | **Data Structuring** | Term 2 – Coefficient 1,5 – 30h |
|---|---|---|
| This course focuses on the development of a database to meet specified requirements. Students learn the steps involved in modeling, developing and implementing a database. It examines tasks and techniques appropriate for each step in the development process. Topics include Entity Relationship model, schema normalization, SQL (data definition instructions of the language). | | |

| UE 22 / M2203 – Statistics and Computer Science | **Statistical Programming 1** | Term 2 – Coefficient 1,5 – 30h |
|---|---|---|

This course gives the methodological bases for data control and data management, using statistical software. The following aspects are addressed :

1. Importing and exporting data frames ;
2. Frequency tables ; cross tabulation ;
3. Missing data, outliers, faulty data ;
4. Coding and recoding data
5. Script programming, macros.

| UE 22 / M2204 – Statistics and Management | **Reporting Tools 2** | Term 2 – Coefficient 1,5 – 30h |
|---|---|---|

Here students will :

- Understand the various methods of cost calculations (accounting management) especially full-costing and variable costing methods (direct costing),

- Learn how to choose the key performance indicators by using reporting tools (classic spreadsheets and balance scoreboards) so as to make them good decision-making tools

- Follow and analyze these data in order to know how well a company is doing

| UE 23 / M2301 – Economics and Management | **General economics and Organisation Management** | Term 2 – Coefficient 2 – 45h |
|---|---|---|

Students will first be asked to further their knowledge of macro-economy: economic growth, income distribution, consumption and investment. Then students will be introduced to strategic decision making in an organization.

| UE 23 / M2302 - Expression | **Expression – Communication 2 : communication, information and argumentation** | Term 2 – Coefficient 2 – 30h |
|---|---|---|

The aim of this class is to enrich students' general knowledge and their knowledge of socio-economic aspects. Students learn how to choose, find, present and synthetize quality information. They will be brought to write long essays, in particular the report of their supervised student project (PPP).

In small groups and in the library, students learn how to use document databases, to carry out online research and read the national press. We also focus on the visualization of graphs and complex data (infographics). They also write and orally present their report of their PPP projects, thus seeing how office software applications can be used in a professional environment.

| UE 23 / M2303 - English | **English for Specific Purposes II** | Term 2 – Coefficient 2 – 30h |
|---|---|---|
| This course aims at deepening students' knowledge of the English of their field as well as their communication skills in a professional environment. Students will be able to develop their opinions about news items and business in general. | | |

| UE 23 / M2304– Project | **Personal and Professional Project 2** | Term 2 – Coefficient 1 – 20h |
|---|---|---|
| This class is geared toward the elaboration of students' professional and personal project. Students will be asked to reflect on their motivations and career choices. They will be invited and helped to approach firms so as to find an internship that meets their expectations. | | |

| UE 24 / M2401 – Project | **Project Management** | Term 2 – Coefficient 2 – 20h |
|---|---|---|
| This course aims at giving students basic skills to be able to conduct their own project (see "Project 1" M2402). The different steps of a project are reviewed on a theoretical viewpoint. Then, a great deal of emphasis is put on practical issues: students learn how to apply a number of tools to their project like the five Ws method, the Gantt and Pert charts, or the responsibilities matrix. | | |

| UE 24 / M2402 - Project | **Project 1** | Term 2 – Coefficient 4- 120h (of personal work) |
|---|---|---|
| The supervised project is for the student the opportunity to put in practice, on real-life cases, the multidisciplinary knowledge acquired in Computer and Statistics classes from their training. It also aims at teaching the student group-work, making them follow the guidelines given by the supervisor (academic or professional) and applying project management techniques. | | |

**Term 3**

| E 31 / M3101 - Statistics | **Data analysis** | Term 3 – Coefficient 3 – 45h |
|---|---|---|
| Understanding the techniques of usual multivariate exploratory analyses (Principal Component Analysis, Correspondence Analysis, Clustering) and knowing how to interpret their results. Students will use professional software to illustrate the implementation of these methods on data taken from real-life issues. | | |

| UE 31 / M3102 - Statistics | **Estimates and tests** | Term 3 – Coefficient 3 – 40h |
|---|---|---|
| This course deepens the introductory course "Introduction to Statistical Inference" held in the previous year. Various operational techniques of statistical inference are taught in a three-fold program : <br><br>1. Hypotheses testing : study of the classical one-sided or two-sided parametric tests ; concept of p-value ; classical nonparametric tests ; goodness-of-fit tests ; correlation tests ; independence chi-square test. <br>2. Estimate: the moment method, the maximum likelihood method, quality of an estimator. <br>3. Confidence intervals: pivot function, confidence interval for a population mean, a population variance or a proportion. | | |

| UE 31 / M3103 - Statistics | **Linear models** | Term  3 – Coefficient 3 – 45h |
|---|---|---|
| This class will enable students to understand the concepts and master the tools of linear modelisation (simple and multiple linear regression). Choice of the model, sensitivity, performance and production, statistical inference and prediction:  tests of hypothesis, ANOVA. | | |

| UE 32 / M3201C – Computer Science | **Introduction to Business Intelligence Systems** | Term 3 – Coefficient 3 – 45h |
|---|---|---|
| This course introduces students to Business intelligence (BI), a set of theories, methodologies, architectures, and technologies that transform raw data into meaningful and useful information for business purposes. The course then focuses on principles, concepts and tools involved in both the design and development of a data warehouse. The multidimensional model for instance will be presented. | | |

| UE 32 / M3202 – Computer Science | **Design and development of a decision-making software** | Term 3 – Coefficient 2 – 30h |
|---|---|---|
| In this course, student will be in charge of the design and development of small decision-making software according to tender specifications. | | |

| UE 32 / M3203 - Management | **Management for Decision Making** | Term 3 – Coefficient 2 – 30h |
|---|---|---|

Students will be presented with budgetary management (provisional management of sales, supply and inventory management, cash management, etc.)

A business game may enable students to manipulate these notions in a practical way.

| UE 33 / M3301 - Economics | **Economics** | Term 3 – Coefficient1,5 – 30h |
|---|---|---|

In this class, students will go over the different economic policies enacted by the government: fiscal policies, monetary policies, growth policies, etc. These will be studied on an international level. The approach is mainly descriptive and aims at highlighting the disputes public decision-makers are confronted with.

| UE 33 / M3302 - Expression | **Expression – communication 3 : Professional Communication** | Term 3 – Coefficient 1,5 – 30h |
|---|---|---|

The aim of this class is to prepare students to find a job internship.

Students learn how to write their CVs and cover letters, to read and reply to a job offer. Students discover how a professional meeting is organized and chaired. They are brought to defend their points of view and participate to debates that they may be asked to organize.

| UE 33 / M3303 - English | **Business English and International Cooperation** | Term 3 – Coefficient 2 – 45h |
|---|---|---|

In this class students will use Business English, deepen their use of English for specific purposes, develop a critical mind and will be introduced to the issues of multiculturalism.

They will write their CV and Cover Letter in English and learn how to communicate in English in a professional environment (telephoning, email, business letters, etc.)

| UE 33 / M3304 - Project | **Personnal and professional project 3** | Term 3 – Coefficient 1 – 20h |
|---|---|---|

Here, students apply on real data the statistical analysis methods and business intelligence software they have been working on.

They carry out a case study, using clear specifications and requirements, thereby reacting to the demand of the person/organization who commissioned the project as they would in their job.

| UE 34 / M3401C – Statistics and Computer Science | **Case studies in statistics and BI** | Term 3 – Coefficient 2,5 – 30h |
|---|---|---|

To be determined (this is a new class).

| UE 34 / M3402C - Statistics | **Fields of Application 1** | Term 3 – Coefficient 1,5 – 30h |
|---|---|---|

**Objectives:**
Initiation to applications of statistics
**Contents:**
Presentation of specific problems and statistical methods used in fields such as:
Management audit, Marketing, Surveys, etc.
**Survey** (9 hours): specific statistical methods  for multiple choice questions, for customer satisfaction surveys, for behavioral analysis, for  open ended questions (introduction)
**Keywords:**
fields of application

| UE 34 / M3403 – Computer Science and Statistics | **Statistics programming 2** | Term 3 – Coefficient 2 – 30h |
|---|---|---|

This course aims at presenting the SAS programming environment and the SAS language: data spreadsheets, statistical analysis, graphics automation of complex tasks.

| UE 34 / M3404 – Project | **Project 2** | Term 3 – Coefficient 2 – 90H (of personal work) |
|---|---|---|

To be determined (this is a new class).

**Term 4**

| UE 41 / M4101 – Statistics and Computer Science | **Data Mining** | Term 4 – Coefficient 3 – 45h |
|---|---|---|

This course will give students knowledge and skills to understand and apply data-mining methods. They will deal with the principles behind supervised classification *(decision tree, logistic regression, discriminant analysis,* the k-*Nearest Neighbors* algorithm).This course will be associated with a course entitled "Big data," where the students will deal with the notion of "structured," "unstructured" data, etc. and they will see an overview of the architecture data and the methods used in Big data.

| UE 41 / M4102C - Statistics | **Surveys** | Term 4 – Coefficient 2 – 35h |
|---|---|---|

**Objectives**
To know the elementary principles of the theory of probability samples; to be capable of criticizing published survey; to be able to propose and implement a simple sample design and to ensure the quality control of the data collected.
**Contents:**
Basic principles concerning the theory of random surveys, the various sources of bias in surveys;  simple random samples; stratified samples; cluster samples; post-stratification; practice in institutes : quota samples
**Keywords:**
Base survey, sample design, fair sample, estimation, bias, precision of a survey, stratification

| UE 41 / M4103C – Computer Science | **Advanced Databases** | Term 4 – Coefficient 2 – 30h |
|---|---|---|

This course covers many additional key database topics from the design and application-building perspective: indexes, views, transactions, authorization, integrity constraints, triggers, on-line analytical processing (OLAP).

| UE 42 / M4201C – Economics, Management, Law | **Economics, Management, Law** | Term 4 – Coefficient 3 – 40h |
|---|---|---|

In this class, students will be asked to think about the main issues of a global economy (globalization, foreign exchange, balance of payments, etc.)

Students will also be introduced to law so as to be able to position themselves in their future professional environment. There will be an emphasis on digital copyright and labor law.

Students will also deepen their knowledge of management: quality control, marketing, human resources management and budgetary and fiscal management.

| UE 42 / M4202 - Expression | **Expression – communication 4 : communication in organisations** | Term 4 – Coefficient 1,5 – 30h |
|---|---|---|

In this class, students are prepared for job interviews (by enacting individual and group mock interviews) and the writing of their job placement report. They study specialized articles from their scientific domain on which they have to prepare a critical presentation (oral and written) in which they highlight the methods used, the results and their limitations.

| UE 42 / M4203 - English | **Scientific English** | Term 4 – Coefficient 2,5 – 45H |
|---|---|---|

This course aims at introducing students to the different types of communication used in a professional environment (meetings, videoconferencing, teamwork).

C1 level from the Common European Framework of Reference for Languages is aimed.

Use and knowledge of Scientific English will be promoted through various activities (reading of scientific publications, conferences, abstracts, etc.).

Students will also be trained for the TOEIC

| UE 42 / M4204C - Statistics | **Field of application 2** | Term 4 – Coefficient 2 – 35H |
|---|---|---|

**Objectives:**
Initiation to some applications of statistics
**Contents:**
Presentation of specific problems and statistical methods used in fields such as:
Biostatistics, Environment (SIG), reliability, etc.
**Survey** (9 hours): specific statistical methods for multiple choice questions, for customer satisfaction surveys, for behavioral analysis, for open ended questions (introduction)
**Keywords:**
fields of application

| UE 43 / M4301 – Project and Job Placement | **Project 3** | Term 4 – Coefficient 2 – 90h |
|---|---|---|

To be determined (this is a new class).

| UE 43 / M4302 – Job Placement | **Internship** | Term 4 – Coefficient 6 – A minimum of 10 weeks |
|---|---|---|

**To complete the professional bachelor's degree, an internship is required in a company for a period of 10 weeks. The key objectives of the internship are for students to :**

- **Familiarize themselves with the company's organization**

- **Be able to explain their position and responsibilities**

- **Apply knowledge and skills learned during the courses**

- **Help in the choice of future career**

This internship is to be considered a professional experience and performance will be assessed upon the following criteria:

- Ability to perform in a professional manner, as indicated by ability to arrive on time, ability to meet deadlines, ability to take initiative in learning, and ability to interact with supervisors and colleagues
- Evaluation of the supervisor in company
- An oral presentation and written report describing the learning experience and performed tasks