**IB/CS 385 INTRODUCTION TO BUSINESS PROGRAMMING**
IES Abroad Milan

**DESCRIPTION:**
This course enables students to acquire basic skills in one of the most fundamental skills of the present period: computer programming. The course introduces students to data analysis programs with the aim of growing their skills in using those programs to predict future outcomes. Students will first learn how to write programming codes from scratch using *R*, one of the most intuitive and popular programming languages. The course will then familiarize students with the most important predictive techniques – from linear regressions to machine learning – and ask them to implement what they've learned theoretically in applications related to management and finance.

**CREDITS:** 3 credits

**CONTACT HOURS:** 45 hours

**LANGUAGE OF INSTRUCTION:** English

**PREREQUISITES:** An introductory knowledge of statistics is required. Students must bring their own laptop in class

**ADDITIONAL COST:** None

**METHOD OF PRESENTATION:**
Theoretical lectures illustrated by examples followed by programming sessions. Students should plan to bring their own laptops (at least 2 GB of RAM) to class on the dates indicated in the schedule. In the first class session all relevant concepts in calculus and statistics will be introduced to students.

Since the conceptual material is dense and requires both theoretical and practical modules in order to be adequately processed and understood by students, the course often pairs theoretical sessions with lab components during which students will get hands-on practice with the week's material by working on assigned lab activities.

**REQUIRED WORK AND FORM OF ASSESSMENT:**
- Course Participation - 10%
- Midterm Exam - 25%
- Final Exam - 30%
- Homework - 15%
- Group Project - 20%

**Course Participation**
Students are expected to actively participate in class and to demonstrate respect and responsibility in the classroom. A rubric for participation will be available on Moodle and distributed on the first day of class.

**Midterm Exam**
Students will be asked to solve exercises on the techniques covered in the first half of the class.

**Final Exam**
Students will be asked to solve exercises on the techniques covered in the second half of the class.

**Homework**
Students will be asked to perform weekly exercises that require them to practice and apply the techniques discussed in class.
- Two of the weekly assignments (in Week 4 and Week 9) will receive a letter grade and together will count for 10% of the grade. These graded assignments will require students both to complete problem sets and to give written explanation of their answers and methods. Students will receive feedback from the instructor on their work.

- During all other weeks homework exercises will be graded for completion and will be worth 5% of the grade. The solutions to the problems will be discussed extensively in class.

**Group Project**
Students will work in groups to apply techniques used in the course to a real-world case study involving a publicly available dataset. They will be asked to write a final report presenting their results of between 1500 and 2000 words (excluding diagrams, graphs, and tables) due the last week of the course. The project will require students to move from a business-related objective to the creation of a research question and ultimately a full statistical report. Students will be expected to (a) identify important variables; (b) perform statistical analyses to address their question; (c) produce tabular and graphical summaries to support their findings; and (d) write a summary report detailing their methodological approach, findings, and potential limitations of their results.

**LEARNING OUTCOMES:**
By the end of the course, students will be able to:

- Apply the logic of computer programming
- Write code that collects, manipulates, and visualizes data in R
- Evaluate a data set and identify analytical techniques that can extract relevant information from it
- Implement the most important predictive techniques in R and apply them to financial or business concerns

**ATTENDANCE POLICY:**
Regular class attendance is mandatory. IES Abroad Milano allows a maximum of **TWO (2)** excused absences per course before the final course grade is penalized. This margin is specifically intended to cover any absences arising from unavoidable and unexpected events or emergencies related to health and family. Each additional absence will automatically result in a penalty of 2 points off (2/100) the final grade. **SEVEN (7)** absences per course will result in a failing grade (the three excused absences included).

**CONTENT:**
The main reference text will be *An Introduction to Statistical Learning with Application in R* (Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, 2013).

In the first class all relevant concepts in calculus and statistics are introduced to students.

Each theoretical session is followed by a practical session in which students are asked to implement the techniques highlighted that week on their own laptops.

| Session | Content | Assignments |
|---|---|---|
| **Session 1** | **Statistical learning and model accuracy.**<br><br>**Part I: Crash course in calculus and statistics** | • Review instructor's slides |
| **Session 2** | **Statistical learning and model accuracy.**<br><br>**Part II: What is statistical learning?** | • Ch 2, pp. 15-52 |
| **Session 3** | **Lab session: introduction to R (bring your own laptop)** | • Review instructor's slides |
| **Session 4** | **Linear regression** | • Ch 3, pp. 59-120 |

| Session | Content | Assignments |
|---|---|---|
| **Session 5** | **Lab session: linear regression (bring your own laptop)** | • Ch 3, pp. 59-120 |
| **Session 6** | **Classification, logistic regressions and LDA** | • Ch 4, pp. 127-168 |
| **Session 7** | **Lab session: classification (bring your own laptop)** | • Ch 4, pp. 127-168 |
| **Session 8** | **Resampling Methods** | • Ch 5, pp. 175-197<br>• **1st graded homework assignment due.** |
| **Session 9** | **Lab session: resampling methods (bring your own laptop)** | • Ch 5, pp. 175-197 |
| **Session 10** | **Linear model selection and regularization** | • Ch 6, pp. 203-259 |
| **Session 11** | **Lab session: model selection (bring your own laptop)** | • Ch 6, pp. 203-259 |
| **Session 12** | **MIDTERM EXAM** | |
| **Session 13** | **Moving beyond linearity: splines and generalized additive models** | • Ch 7, pp. 265-297 |
| **Session 14** | **Lab session: non-linear modelling (bring your own laptop)** | • Ch 7, pp. 265-297 |
| **Session 15** | **Tree based methods: decision trees, random forest, and boosting** | • Ch 8, pp. 303-332 |
| **Session 16** | **Lab session: decision trees (bring your own laptop)**<br>• | • Ch 8, pp. 303-332 |
| **Session 17** | **Support Vector Machines** | • Ch 9, pp. 337-368<br>• **2nd graded homework assignment due.** |
| **Session 18** | **Lab session: support vector machines** | • Ch 9, pp. 337-368 |

| Session | Content | Assignments |
|---|---|---|
| **Session 19** | **Unsupervised learning: principal component and clustering (Ch 10, pp. 373-413)** | • Ch 10, pp. 373-413 |
| **Session 20** | **Lab session: unsupervised learning** | • Ch 10, pp. 373-413 |
| **Session 21** | **Introduction to Simulations, text as data, and network science** | • Instructor slides |
| **Session 22** | **R and Python: similarities and differences (instructor's slides)** | • Instructor slides<br>• **Group project due** |
| **Session 23** | **FINAL EXAM** | |

**REQUIRED READINGS:**

- James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani. An Introduction to Statistical Learning with Application in R. Springer, 2013.

**RECOMMENDED READINGS:**

- Larose, Chantal D. and Daniel T. Larose. *Data Science Using Python and R*. (Hoboken, NJ: Wiley, 2019).