

## **MSc. Applied Statistics and Data Mining**

The MSc. in Applied Statistics and Data mining is a new program This program differs from our MSc. in Statistics by being strongly applied with an emphasis on application in the commercial sector. The program focuses on the specific tools and methods that are useful and sought-after in current real-world data analysis and are largely taught within widespread commercial, rather than research, software packages e.g. SAS, SPSS. Tools are taught to cope with data that is messy and does not conform nicely to traditional methods and their rigid assumptions – the types of data most frequently encountered. The overarching goals are pragmatic solutions that provide defensible recommendations for clients, employers or researchers.

There is great demand within the job market for competent data analysts – the emphasis being on practical data manipulation, analysis and interpretation, rather than theoretical or methodological developments. Additionally there is requirement within applied analysis for familiarity with very recent methods that are not provided within typical statistics syllabuses. Graduates from this program would be seeking employment as an analyst within a company, research body, government, or as a statistical consultant.

### **Aims**

The program aims to develop the students' interest, knowledge and understanding of statistics in its practical application to real-world data analysis. The program aims to build a broad statistical knowledge through intensive, challenging modules.

The program is intended to instill in the students:

- Critical thinking skills;
- The ability to think statistically and define solutions in the presence of uncertainty;
- The ability to translate questions of interest into statistical hypotheses;
- Methods of displaying and analyzing data for specific research questions;

- IT skills in the major research and commercial statistics packages;
- Knowledge of the most recent and sought-after statistical methods, particularly those applicable to contemporary data-rich scenarios.

### **Learning Outcomes:**

**At the end of the program the students will have knowledge of:**

- Data analysis methods, covering traditional methods but with greater emphasis on modern methods that locate and address common data foibles;
- Survey design & data collection issues;
- Multivariate methods: supervised/unsupervised classification, data reduction;
- Univariate methods: both basic (e.g. t-tests, ANOVA, linear models) and advanced (e.g. Generalized Linear Models, Generalized Additive Models);
- Data mining methods: tree methods with boosting and bagging; Multivariate Adaptive Regression; Random Forests; Neural Nets;
- Model diagnostics;
- Tools for difficult data: ridge regression; basic data imputation;
- What software tools exist, along with their strengths and weaknesses, for these types of analyses?

**Similarly during the program the students will acquire understanding of:**

- How to formulate the problem and choose an appropriate analysis solution;
- The limitations of the methods studied;
- When the analysis requirements are beyond their current skill set;
- How to take a novel problem and profitably search for an existing, yet unknown, method of analysis.

By completion of the programmed the students will have the ability to:

- Conceive pragmatic solutions in the presence of uncertainty;
- Think critically - particularly with respect to conceiving multiple explanations for observed phenomena (this is key in the understanding of most statistical concepts e.g. statistical testing, inference, design, model construction and diagnostics etc), and choosing between these;
- Isolate the key questions that require answering within real-life scenarios;
- Determine the type of data required to answer such questions, and devise appropriate data collection;
- Formulate analyses suitable to answer their own or their clients' questions, both under ideal data or within the constraints of available data;
- Identify data limitations with respect to their ability to answer specific questions;
- Convey potentially complex results simply.