

Course title: SC04 - Survey Analytics from Questionnaires and Textual Social Media Analytics

Duration: 1 day

Date and time: 15 July

Venue: Faculté de Médecine et de Pharmacie de Marrakech
Address: Sidi Abbad, Marrakech 40000

Registration fee:

- Developed Country: € 240
- Host , Least Developed & Developing Country, or Student⁽¹⁾: € 150

(1) For students, proof of enrolment will be required.



Instructor 1

Fionn Murtagh, Professor, Big Data Lab, University of Derby; and Goldsmiths University of London, UK

Fionn Murtagh is a Professor of Data Science, PhD: Mathematical Statistics, Université P&M Curie, Paris 6. Previous positions included big data in education, astrophysics, and director of national research funding. Advisor to many companies. Editorial Board roles with many journals. Published about 300 refereed articles, authored or edited about 30 books. Member ISI, IASC, Member of Executive IASC. Fellow of: Royal Statistical Society, British Computer Society (FBCS), Institute of Mathematics and Its Applications (FIMA), International Association for Pattern Recognition (FIAPR), Royal Statistical Society (FRSS), Royal Society of Arts (FRSA). Elected member Royal Irish Academy (MRIA), Academia Europaea (MAE), Senior Member IEEE.



Instructor 2

Mohsen Farid, Dr, University of Derby
Department of Computing and Mathematics, UK

Mohsen Farid is a Senior lecturer (Associate Professor). He has spent a number of years working in the area of eye tracking modelling and time series analysis, and big data and text mining in the areas of cognition and criminology and Natural Language Processing. He is a Senior Member of IEEE. Mohsen has also received an "Honourable Mention" by the US Army for work on Data Mining of Breast Cancer Research. He is organizing a workshop on the Human Side of Cybersecurity in IEEE/ACM - BDSEA 2016 (the 3rd IEEE/ACM International Conference on Big data Computing, Applications and Technologies, in December 2016).



COURSE DESCRIPTION

The work of the celebrated social scientist Pierre Bourdieu (1930-2002) includes the thoughtful and creative use of the Correspondence Analysis, published in English in 1984, with title *Distinction*. It is on such a geometric data analysis approach that this course is based.

The focus is: (1) interpretation of results, graphical displays and other outputs, (2) practical implementation using the R statistical and visualization environment, and (3) providing intuition, and full understanding, relating to the geometry and statistical processing. We use data collected in various questionnaires, starting from work by Bourdieu on cultural taste. Other questionnaire analysis case studies will be related to transport, cooking and lifestyle, student experience, consumer behaviour, and music appreciation.

Next the questionnaire outcomes express both closed, fixed format questions, and, conjointly analysed, free text responses.

Finally studied will be data sourced from social media micro-blogging, i.e. Twitter.

SYLLABUS

Used is the open source and universally supported R statistical and visualization environment.

Topics.

In accompanying online course materials, there will be a practical introduction to the R language and environment. This is for participants who have not used R before.

Part 1: Questionnaire analysis case study: taking the Bourdieu taste data, detailed discussion of output, detailing the R code used.

Part 2: Geometric intuition: the methodology used for graphical display, hierarchical clustering, and putting it all together.

Part 3: Carrying out geometric data analysis, including clustering, using R. Including publication/presentation outputs, storing data for later work, and maintaining the R scripts that are used.

Part 4: Further case studies of questionnaire analysis.

Part 5: Questionnaire analysis, using conjoint, or integrally related, analysis of closed questions, and open or free text questions.

Part 6: Coverage of social media data sources, will be especially centred on Twitter.

All sessions will be associated with practical exercises, using case studies.

Final Part: Concluding short debate and discussion on potential and scope for analytics, and statistical treatment of data, and text mining.

TARGET AUDIENCE

Practitioners and researchers related to any domains that are encompassed in the case studies, and practical exercises. Students who are undertaking, or who are planning to undertake, any and all such work. Domains of general relevance include:

Health and medical surveys, marketing,
Security and forensics,



Information and data sourcing through web-based questionnaires,
Lifestyle and wellbeing analytics,
Legal studies,
Political studies,
Language and literature, digital humanities.

While presented in English, in order to have the full assembly of participants, the case studies that will be studied in lab exercises will have separate case studies, i.e. parallel versions of the data to be analysed, for: English, French, Arabic. The course instructors will liaise with student questions and answers, in, respectively English and French; and English and Arabic.