

SPRING-SCHOOL IN QUANTITATIVE METHODS 2017

University of Hohenheim, June 6th to June 8th

The Spring School in Quantitative Methods at University of Hohenheim is hosted by the Faculty of Business, Economics and Social Sciences of the University of Hohenheim. It provides the opportunity to explore quantitative methods and their applications in the fields of Social Sciences, Economics, Finance and Business Administration. The school addresses the needs of doctoral students and post-docs who intend to strengthen their methodical skills, discuss current research projects with other participants and discover new ways to investigate their problems at hand.

Topic: Quantitative Text Analysis

Instructor: Prof. Kenneth Benoit

(Department of Methodology, London School of Economics)

Content:

The course surveys methods for systematically extracting quantitative information from political text for social scientific purposes, starting with classical content analysis and dictionary-based methods, to classification methods, and state-of-the-art scaling methods and topic models for estimating quantities from text using statistical techniques. The course lays a theoretical foundation for text analysis but mainly takes a very practical and applied approach, so that students learn how to apply these methods in actual research. The common focus across all methods is that they can be reduced to a three-step process: first, identifying texts and units of texts for analysis; second, extracting from the texts quantitatively measured features – such as coded content categories, word counts, word types, dictionary counts, or parts of speech – and converting these into a quantitative matrix; and third, using quantitative or statistical methods to analyse this matrix in order to generate inferences about the texts or their authors. The course systematically covers these methods in a logical progression, with a practical, hands-on approach where each technique will be applied using appropriate software to real texts.

Objectives:

The course is also designed to cover many fundamental issues in quantitative text analysis such as inter-coder agreement, reliability, validation, accuracy, and precision. It focuses on methods of converting texts into quantitative matrixes of features, and then analysing those features using statistical methods. The course briefly covers the qualitative technique of human coding and annotation but only for the purposes of creating a validation set for automated approaches. These automated approaches include dictionary construction and application, classification and machine learning, scaling models, and topic models. For each topic, we will systematically cover published applications and examples of these methods, from a variety of disciplinary and applied fields but focusing on political science. Lessons will consist of a mixture of theoretical grounding in content analysis approaches and techniques, with hands on analysis of real texts using content analytic and statistical software.

Literature:

- Manning, Raghavan and Schütze. *An Introduction to Information Retrieval*, Cambridge UP (2008, Ch. 13)
- Lantz *Machine Learning with R*. Packt Publishing (2013, Ch. 3–4)
- Laver, Benoit and Garry (2003) *Extracting Policy Positions from Political Texts Using Words as Data*. *American Political Science Review* 97(2, May): 311-331.

Course outline:

Tuesday, June 6 th	
	Quantitative text analysis overview and fundamentals This session will cover fundamentals, including the continuum from traditional (non-computer assisted) content analysis to fully automated quantitative text analysis. We will cover the conceptual foundations of content analysis and quantitative content analysis, discuss the objectives, the approach to knowledge, and the particular view of texts when performing quantitative analysis. We will also discuss issues including where to obtain textual data; formatting and working with text files; indexing and meta-data; units of analysis; and definitions of features and measures commonly extracted from texts, including stemming, and stop-words. Descriptive statistical methods for textual analysis Here we focus on quantitative methods for describing texts, focusing on summary measures that highlight particular characteristics of documents and allowing these to be compared. These methods include characterizing texts through concordances, co-occurrences, and keywords in context; complexity and readability measures; and an in-depth discussion of text types, tokens, and equivalencies. We will also discuss weighting strategies for features, such as tf-idf. The emphasis will be on comparing texts, through concordances and keyword identification, dissimilarity measures, association models, and vector-space models.

	Practical Session
Wednesday, June 7th	
	Automated dictionary-based approaches Automatic dictionary-based methods involve association of pre-defined word lists with particular quantitative values assigned by the researcher for some characteristic of interest. This topic covers the design model behind dictionary construction, including guidelines for testing and refining dictionaries. Hand-on work will cover commonly used dictionaries such as LIWC, RID, and the Harvard IV-4, with applications. We will also review a variety of text pre-processing issues and textual data concepts such as word types, tokens, and equivalencies, including word stemming and trimming of words based on term and/or document frequency. Practical Session
Thursday, June 8th	
	Machine Learning and Scaling for Texts Classification methods permit the automatic classification of texts in a test set following machine learning from a training set. We will introduce machine learning methods for classifying documents, including one of the most popular classifiers, the Naive Bayes model. The topic also introduces validation and reporting methods for classifiers and discusses where these methods are applicable. Building on the Naive Bayes classifier, we introduce the “Wordscores” method of Laver, Benoit and Garry (2003) for scaling latent traits, and show the link between classification and scaling. We also cover applications of penalized regression to score and scale texts. Practical Session

R-Tutorial:

The practical examples during the Spring School will mainly be discussed using the statistical software package R, therefore basic knowledge in R is desirable. Participants with little or no knowledge in R are invited to register for a free of charge two half-day introductory course in R at **June 1st** and **2nd** prior to the Spring School.

Application, fees, dates and credits:

Eligible for admission are doctoral students and post-docs in the fields of social sciences, economics, finance, business administration and related subjects.

Knowledge of a specific programming language is not required, but a basic understanding of the R-language is helpful for understanding the illustrations of the implementation of the methods discussed.

The Spring School will take place at the University of Hohenheim.

Members of the University of Hohenheim can participate in the Spring School free of charge.

The registration fee for the all other participants is **250 Euros**. It covers the tuition, coffee breaks and drinks throughout the Spring School.

Interested candidates should apply by [mail](#) before **April 17, 2017**. After this deadline, your place may not be guaranteed. The application requires personal details, information on current or previous university study and current employment information. Moreover information on the level of knowledge in statistical methods and proficiency in R should be provided. Applications will be evaluated and candidates will be informed of the decision by **April 24, 2017**. For successful applicants a document will be attached to our response with detailed payment information.

At the conclusion of the Spring School, participants will receive a certificate for the number of hours attended. 6 ECTS credits can be earned by completing a written assignment/exam.

Contact:

For further information please contact

Dr. Johannes Klenk
Faculty of Business, Economics and Social Sciences
University of Hohenheim
Schloss Hohenheim 1 B
70599 Stuttgart

e-mail: wiso_forschung@uni-hohenheim.de