DSS Statistics Seminar
November 20, 2020, 12:00 a.m.
https://meet.google.com/sqz-deps-qdi

Computational and Other Challenges in Using Text as a New Source of Data

P. Winker
University of Giessen

The use of textual information gained momentum over the last years in economics. Text is considered as the new data in fields such as financial markets, innovation activities and economic history. For drawing meaningful conclusions from this type of data, a substantial number of steps in pre-processing and analyzing the data has to be taken. Usually, the implementation of these methods is based on previous experience, statistical methods, or human judgement. Thus, typical issues present when dealing with conventional quantitative data also apply to textual information. They might just be disguised differently, while new challenges show up. In the talk some of the relevant steps will be sketched and it will be demonstrated using an example of conference abstracts, how decisions in each step might affect the outcomes. Finally, some remarks on the use of the generated indicators in further analysis will be provided and a substantial set of open issues regarding, e.g. computational complexity and robustness of the methods might serve as starting point for further research.

SAPIENZA
UNIVERSITÀ DI ROMA