Advanced Topics in Official Statistics
SURV 699G
1 credit/2 ECTS
Fall 2020

Instructors

Hanna Brenzel, Federal Statistical Office Germany, hb@hanna-brenzel.de
Piet Daas, Statistics Netherlands, pjh.daas@cbs.nl
Marco Puts, Statistics Netherlands, m.puts@cbs.nl

Video lecture by

Hanna Brenzel, Federal Statistical Office Germany
Piet Daas, Statistics Netherlands
Marco Puts, Statistics Netherlands

Short Course Description

The course gives an overview about advanced topics in official statistics such as Big Data, georeferenced data, and microsimulations. The benefits and downsides of using Big Data as a data source for official statistics production are discussed and examples are given. In addition, the course provides insights into the use of georeferenced data in official statistics and compares the current achievements in Germany with international developments. Moreover, an overview of the past, the present, and the future state-of-the-art of microsimulation methods and applications within official statistics is given.

Course and Learning Objectives

By the end of the course, students will…

- have a basic knowledge on how to use Big Data in official statistics
- be able to list the advantages and downsides of particular Big Data sources
- be able to explain what geo information and geo data is and understand the importance and usefulness of georeferenced data
- have basic knowledge on the developments of microsimulations within and outside Germany
- be able to describe what microsimulation is and what different types of microsimulations exist

Prerequisites

No prerequisites.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor. Just like in an on-site course, homework will be assigned and graded.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching prerecorded videos, attending the live online meetings), and follow-up activities (working on assignments and exams) – as in an on-site course. As a rule of thumb, you can expect to spend approximately 3h/week on in-class-activities and 9 hours per week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes and exams). Therefore, the workload in all courses will be approximately 12h/week. This is a 1-credit/2 ECTS course that runs for 4 weeks. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Mondays, 12:00 PM – 12:50 PM ET / 6:00 PM – 6:50 PM CET, starting November 2nd

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on https://www.elms.umd.edu/. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are required to post questions or comments about the materials covered in the videos or readings of the week in the forum before the meetings (deadlines for posting are 24 hours before the online meetings).

Students have the opportunity to use the Zoom meeting room set up for this course to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are encouraged to post the times that they will be using the room to the course website forum to avoid scheduling conflicts. Students are not required to use Zoom and can use other online meeting platforms such as Google Hangout or Skype.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and submission of questions in the forum on the course page before the meetings (deadline: 24 hours before the online meetings) demonstrating understanding of the required readings and video lectures (20% of grade)
- 4 online quizzes (worth 80% total and 20% each)

Dates of when quizzes will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructor's discretion.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between students and the instructor during the weekly online meetings. Therefore, we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than $20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

Long Course Description

The course gives an overview about various current topics in official statistics such as Big Data, georeferencing and microsimulation. The topics are therefore independent of one another but must be considered together when it comes to current advanced topics in official statistics. The course provides an overview on the statistical use of Big Data and provides essential background knowledge to enable this. The course consists of 3 sections: one on Big Data, composed of two lectures, and two on using georeferenced data (i.e. data that is attached to a unique location) and applying microsimulation techniques; one lecture each. A considerable number of examples will be discussed. The Big Data section will provide a general look at the benefits and downsides of Big Data in official statistics. It also is the starting point for Big Data methodology development. In addition, the relation between Big Data analysis and the various Big Data IT environments is discussed.

The second part of the course lies on the importance and use of georeferenced data. The goal of Eurostat, the European National Statistical Offices and the Federal Statistical Office is to provide statistical information that is necessary for decision-making processes in a democratic society. In times of open data, georeferencing

creates a new, expanded basis for evidence-based decisions with regional relevance. The expansion of the regional reference consists in the spatial depth and flexibility of the analyzed regional units. Information systems based on integrated statistical and geographic information can support the political process and forecasts, while at the same time bringing the different dimensions of sustainable development (ecological, economic, social) into a coherent picture. Using geospatial information in the production of statistics has numerous advantages and creates a statistic overarching potential for analysis, which goes far beyond small-scale cartographic presentation of single statistics. In addition, the analysis potential based on official data will be extended by integrating geospatial information.

A third focus of the course lies on microsimulation. The planning and further development of political decisions increasingly requires the use of special simulation and calculations that go beyond the published statistical standard tables of official statistics, in order to be able to evaluate and estimate consequence of political measures. The course provides a basic overview about the idea of microsimulation, its origin and development over time and highlights the strong relationship between the developments of microsimulation and access to individual data. Moreover, the different types of microsimulation will be presented as well as insights into ongoing projects in Germany will be given.

Readings
There is no text book for this course.

Required and recommended readings are provided below for each specific unit.

Academic Conduct
Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III-100A.pdf (University of Maryland) and

https://www.uni-mannheim.de/en/research/good-research-practice/ (University of Mannheim).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or

failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office to register in person for services. Please call the office to set up an appointment to register with an ADS counselor. Contact the ADS office at 301.314.7682; https://www.counseling.umd.edu/ads/.

Students at the University of Mannheim should contact the Commissioner and Counsellor for Disabled Students and Students with Chronic Illnesses at http://www.uni-mannheim.de/studienbueros/english/counselling/disabled_persons_and_persons_with_chronic_illnesses/.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Class Schedule

Please note that assignments and dates are subject to change. Information (e.g., articles and assignments) posted to the course website supersedes the information noted here.

Unit 1: Introduction to Big Data in official statistics

Video lecture: available Monday, October 26, 2020

Online meeting: Monday, November 2nd, 2020, 12:00 PM – 12:50 PM ET / 6:00 PM – 6:50 PM CET

Online quiz 1: due Wednesday, November 4, 2020, 12:00 PM ET / 6:00 PM CET

Readings:
Daas, P.J.H., Puts, M.J.H. (2014) Big Data as a Source of Statistical Information. *The Survey Statistician* 69, 22-31. Available at: http://isi-iass.org/home/wp-content/uploads/N69-2014-01-ISSN.pdf

Recommended (optional):
Daas et al. (2015) Big Data and Official Statistics. Journal of Official Statistics 31(2), 249-262. Available at:

https://content.sciendo.com/downloadpdf/journals/jos/31/2/article-p249.xml

Kitchen, R. (2014) Big Data, new epistemologies and paradigm shifts. Big Data & Society, April-June, 1-12. Available at: https://journals.sagepub.com/doi/pdf/10.1177/2053951714528481

Unit 2: Using Big Data for official statistics

Video lecture: available Monday, November 2nd

Online meeting: Monday, November 9, 2020, 12:00 PM – 12:50 PM ET / 6:00 PM – 6:50 PM CET

Online quiz 2: due Wednesday, November 11, 2020, 12:00 PM ET / 6:00 PM

Readings:
Ginsberg et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457, 1012-1014. Available at: https://www.nature.com/articles/nature07634.pdf

Lazer et al. (2014) The Parable of Google Flu: Traps in Big Data Analysis. Science 343(6176), 1203-1205. Available at: https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf

Recommended (optional):
Puts et al. (2015) Finding Errors in Big Data. *Significance* 12 (3), 26-29. Available at: https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2015.00826.x

Unit 3: The importance and use of georeferenced data in official statistics

Video lecture: available Monday, November 9, 2020

Online meeting: Monday, November 16, 2020, 12:00 PM – 12:50 PM ET / 6:00 PM – 6:50 PM CET

Online quiz 3: due Wednesday, November 18, 2020, 12:00 PM ET / 6:00 PM CET

Recommended (optional):
UN-GGIM: Europe. 2019. The integration of statistical and geospatial information — a call for political action in Europe, European Union/United Nations, Luxembourg. Available at:

https://ec.europa.eu/eurostat/documents/4031688/10158240/KS-03-19-423-EN-N.pdf/c8f75ee1-2181-288c-1efa-1622c5abb980

UN-GGIM. 2019. The global statistical and geospatial framework, United Nations, New York. Available at: http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf

INSEE-Eurostat. 2018. Handbook of Spatial Analysis. Theory and Application with R, N° 131. Available at: https://ec.europa.eu/eurostat/documents/3859598/9462709/INSEE-ESTAT-SPATIAL-ANA-18-EN.pdf/c4f87d5b-b508-4aff-ad7d-264da463077e

Unit 4: Introduction to microsimulation in official statistics

Video lecture: available Monday, November 16, 2020

Online meeting: Monday, November 23, 2020, 12:00 PM – 12:50 PM ET / 6:00 PM – 6:50 PM CET

Online quiz 4: due Wednesday, November 25, 2020, 12:00 PM ET / 6:00 PM CET

Readings:
Orcutt (2007) A new type of socio-economic system. *International Journal of Microsimulation (Reprinted)*, 1(1), pp. 3-9. Available at: https://www.microsimulation.org/IJM/V1_1/IJM_1_1_2.pdf

Recommended (optional):
Li and O'Donoghue (2013) A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* 6(2), pp. 3-55. Available at: https://www.microsimulation.org/IJM/V6_2/2_IJM_6_2_2013_Li_Odonoghue.pdf

Rahman and Harding (2017) Small area estimation and microsimulation modelling. Chapman and Hall/CRC, London, UK, Chapter 4.

Merz (1991) Microsimulation - a survey of principles, developments and applications. *International Journal of Forecasting, 7(1),* pp. 77-104. Available at: https://mpra.ub.uni-muenchen.de/7232/