Syllabus

Fundamentals of Survey and Data Science

3 credits/6 ECTS

Instructor: Dr. Ruben Bach
Teaching Assistant: João Areal

Video lecture by
Prof. Florian Keusch, Prof. Frauke Kreuter,
Jennifer Sinibaldi PhD and Prof. Richard Valliant

September 3 – December 3, 2025

Short Course Description

The course introduces the student to a set of principles of survey and data science that are the basis of standard practices in these fields. The course exposes the student to key terminology and concepts of collecting and analyzing data from surveys and other data sources to gain insights and to test hypotheses about the nature of human and social behavior and interaction. It will also present a framework that will allow the student to evaluate the influence of different error sources on the quality of data.

Course Objectives

By the end of the course, students will...

- be able to apply the key terminology used by survey methodologists and data scientists.
- be able to assess the quality of data from different sources based on a data quality framework.
- be able to select an appropriate data source to answer different types of research questions.
- understand the influence of coverage, sampling, and nonresponse on data quality and know how to deal with deficiencies of the data.
- have a clear understanding of the steps involved in data preparation, data processing, data analysis, and data visualization.
- be able to comply with ethical standards in survey research and data science.

Prerequisites

Students are expected to be familiar with basic statistical concepts, such as mean, standard deviation, variance, and distributions (at the level of an undergraduate

course) and have exposure to elements of social science perspectives on human behavior.

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.

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-classactivities and 9h/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. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Wednesdays, 11:00 AM EDT/5:00 PM CEST, starting September 3, 2025

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on umd.instructure.com 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 encouraged to post questions about the materials covered in the videos and readings of the week in the forum before the meetings (deadline for posting questions is Tuesday, 11:00 AM EDT/5:00 PM CEST).

There will also be weekly office hours offered by the teaching assistant where students have an additional opportunity to ask questions about the course materials. The office hours will be held in the hour before the weekly online meetings (Wednesdays, 10:00 AM EDT/4:00 PM CEST – 11:AM EDT/5:00 PM, starting September 10, 2025) through Zoom.

Students have the opportunity to use the Conferences feature in Canvas to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are not required to use Canvas Conferences and can of course use other online meeting platforms such as Google Hangouts, Skype or Microsoft Teams.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and contributions to the weekly discussion forum **(deadline is Tuesday, 11:00 AM EDT/5:00 PM CEST)**, demonstrating understanding of the required readings and video lectures, is worth 10% of the final grade.
- Ten online quizzes, worth 100 points each, reviewing specific aspects of the material covered. The simple average of the points across all quizzes is worth 60% of the final grade.
- A final open-book online exam, worth 100 points, is worth 30% of the final grade.

A+ 100 - 97
A 96 - 93
A- 92 - 90
B+ 89 - 87
B 86 - 83
B- 82 - 80
Etc.

The final grade will be communicated under the assignment "Final Grade" in the Canvas course. Please note that the letter grade written in parentheses in Canvas is the correct final grade. The point grade displayed alongside the letter grade is irrelevant and can be ignored.

Dates of when assignments will be due are indicated in the syllabus. Quizzes and assignments will automatically close on the course website at the indicated time. Assignments can be completed at any time leading up to the deadline. Extensions will be granted sparingly and are at the instructor's discretion.

The final exam is open book and open notes, and students are on the honor system (see Academic Conduct below). **The exam will take place on Wednesday, December 3, 2025 at 11:00 AM EST/5:00 PM CET. Students have** two days **to complete the exam and must submit their exam before 1:00 PM EST/7:00 PM CET.**

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between the 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 cameras are available for less than $20 each. We ask students to refrain from using built-in web cameras and speakers on their desktop computers 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 fields of survey methodology and data science draw on theories and practices developed in several academic disciplines – mathematics, statistics, psychology, sociology, computer science, and economics. To become an accomplished professional in these fields requires a mastery of research literatures as well as experience designing, conducting, and analyzing surveys and data from other sources, such as administrative records, social media, or transactions.

This course introduces the student to a set of principles of survey design and data science that are the basis of standard practices in these fields. The course exposes the student to research literatures that use both observational and experimental methods to test key hypotheses about the nature of human behavior and other factors that affect the quality of data. It will also present important statistical concepts and techniques in sample design, execution, and estimation, as well as models of behavior describing errors in responding to survey questions. Thus, both social science and statistical concepts will be presented.

The course uses the concept of total error as a framework to discuss coverage properties of sampling frames and organic data, alternative sample designs and their impacts on standard errors of statistics, different modes of data collection and generation, the role of interviewers and respondents in surveys, impacts of nonresponse and missing data on statistics, measurement errors in data, data processing, and data/research ethics.

The course is intended as an introduction to the fields of survey methodology and data science, taught at a graduate level. Lectures and course readings assume that students understand basic statistical concepts (at the level of an undergraduate course) and have exposure to elements of social science perspectives on human behavior. For those lacking such a background, supplementary readings are recommended.

Readings

Primary readings will be from the following volumes:

Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology, 2nd Edition.* New York: Wiley. (available at local bookstores and online retailers)

Peng, R.D. & Matsui, E. (2015). *The Art of Data Science. A Guide for Anyone Who Works with Data.* Leanpub. (available online at https://leanpub.com/artofdatascience).

Required and Recommended Readings

Lists of required and recommended readings are provided below for each unit.

None of the information in the recommended readings will be included on the online quizzes or the final exam.

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)

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/.

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. Participation is voluntary but highly appreciated.

Sessions

Please note that assignments and dates are subject to change. Information posted in Canvas supersedes the information noted here.

Unit 1: Introduction – How to do survey research and data science

Video lecture: available Wednesday, August 27, 2025

Online meeting: Wednesday, September 3, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 1: due Thursday, September 4, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
Groves et al. (2009). Chapters 1.4 and 1.5

Peng & Matsui (2015). Chapters 1-3

Leek, J.T. and Peng, R.D. (2015). What is the question? *Science, 347*, 1314-1315.

Recommended Readings:
Couper, M. (2013). Is the sky falling? New technology, changing media, and the future of surveys.
Survey Research Methods, 7(3), 145-156. https://ojs.ub.uni-konstanz.de/srm/article/view/5751/5289

Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *PS: Political Science and Politics, 48*, 80-83. https://web.stanford.edu/~jgrimmer/bd_2.pdf

Groves, R. (2011). Three eras of survey research. *Public Opinion Quarterly, 75*(5), 861-871. http://www.uvm.edu/~dguber/POLS234/articles/groves.pdf

Unit 2: Quality of Data

Video lecture: available Wednesday, September 3, 2025

Online meeting: Wednesday, September 10, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 2: due Thursday, September 11, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
Groves et al. (2009). Chapter 2

Biemer, P. (2010). Total Survey Error. Design, implementation, and evaluation. *Public Opinion Quarterly, 74*(5), 817-848.

Recommended Readings:
EUROSTAT (2007). *Handbook on data quality assessment methods and tools.* https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS %20%20I.pdf

Morganstein, D. & Marker, D.A. (2012). Continuous quality improvement in statistical agencies. In L. Lyberg et a. (Eds.), *Survey Measurement and Process Quality*, John Wiley & Sons: Hoboken, NJ, 475500.

Lavrakas, P. (2013). Presidential Address: Applying a Total Error Perspective for Improving Research Quality in the Social, Behavioral, and Marketing Sciences. *Public Opinion Quarterly, 77*(3), 831-850.

AAPOR (2015). *AAPOR Report on Big Data.* https://www.aapor.org/getattachment/Education-Resources/Reports/BigDataTaskForceReport_FINAL_2_12_15_b.pdf.aspx

Unit 3: Coverage

Video lecture: available Wednesday, September 10, 2025

Online meeting: Wednesday, September 17, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 3: due Thursday, September 18, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
Groves et al. (2009). Chapter 3

Hargittai, E. (2015). Is bigger always better? Potential biases of Big Data derived from social network sites. *The Annals of the American Academy of Political and Social Science, 659*, 63-76.

Kreuter, F. and Peng, R.D. (2014). Extracting information from Big Data: Issues of measurement, inference and linkage. In J. Lane, V. Stodden, S. Bender, and H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 257-275.

Unit 4: Modes of Survey Data Collection

Video lecture: available Wednesday, September 17, 2025

Online meeting: Wednesday, September 24, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 4: due Thursday, September 27, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
Groves et al. (2009). Chapters 1.3 and 5

Recommended Readings:
Couper, M.P., Antoun, C., & Mavletova, A. (2017). Mobile Web Surveys. In P.P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L.E. Lyberg, N.C. Tucker, & B.T. West (Eds.), *Total Survey Error in Practice*, 133-154.

Blom, A.G., Bosnjak, M., Cornilleau, A., Cousteaux, A.S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review, 34*(1), 8-25. https://journals.sagepub.com/doi/pdf/10.1177/0894439315574825

de Leeuw, E.D. (2018). Mixed-Mode: Past, present, and future. *Survey Research Methods, 12*(2), 75-89. https://ojs.ub.uni-konstanz.de/srm/article/view/7402/6582

Unit 5: Data Generation from Other Sources

Video lecture: available Wednesday, September 24, 2025

Online meeting: Wednesday, October 1, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 5: due Thursday, October 2, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
boyd, d. & Crawford, K. (2012). Critical questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Lazar, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in Big Data analysis. *Science, 343*(6176), 1203-1205.

Stier, S., Breuer, J., Siegers, P. & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data:
Key Issues in Developing an Emerging Field. *Social Science Computer Review, 38*(5), 503-516.

Recommended Readings:
Benzeval, M., Bollinger, C.R., Burton, J., Couper, M.P., Crossley, T.F., Jäckle, A. (2020). Integrated data: Research potential and data quality. *Understanding Society Working Paper, 2020-02*.
https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2020-02.pdf

AAPOR (2014). *Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys:*
Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. https://www.aapor.org/getattachment/Education-Resources/Reports/REVISED_Mobile_Technology_Report_Final_revised10June14.pdf.aspx

Schober, M.F, Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F.G. (2016). Social media analyses for social measurement. *Public Opinion Quarterly, 80*(1), 180-211.

Unit 6: Sampling I

Video lecture: available Wednesday, October 1, 2025

Online meeting: Wednesday, October 8, 2025, 11:00 AM EDT/5:00 PM

CEST Online quiz unit 6: due Thursday, October 9, 2025, 5:59 PM

EDT/11:59 PM CEST

Required Readings:
Groves et al. (2009). Chapters 4.1-4.6

Unit 7: Sampling II

Video lecture: available Wednesday, October 8, 2025

Online meeting: Wednesday, October 15, 2025, 11:00 AM EDT/5:00

PM CEST Online quiz unit 7: due Thursday, October 16, 2025, 5:59

PM EDT/11:59 PM CEST

Required Readings:
Baker, R. et al. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly, 74*(4), 711-781.

Recommended Readings:
Baker, R. et al. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology, 1*(2), 90-143.

Cornesse, C. et al. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology, 8*(1), 4-36.

Valliant, R. (2020). Comparing alternatives for estimation from nonprobability samples. *Journal of Survey Statistics and Methodology, 8*(2), 231-263.

Unit 8: Questionnaires and Interviewing

Video lecture: available Wednesday, October 15, 2025

Online meeting: Wednesday, October 22, 2025, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 8: due Thursday, October 23, 2025, 5:59 PM EDT/11:59 PM

CEST Daylight saving time ends in Europe on October 26, 2025, and clocks are forward 1 hour.

Required Readings:
Groves et al. (2009). Chapters 8.1-8.8 and 9.1-9.7

Recommended Readings:
West, B.T. & Blom, A. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175-211.

Unit 9: Nonresponse

Video lecture: available Wednesday, October 22, 2025

Online meeting: Wednesday, October 29, 2025, 11:00 AM EST /5:00 PM CET

Online quiz unit 9: due Thursday, October 30, 2025, 5:59 PM EST /11:59 PM CET

Daylight saving time ends in the US on November 2, 2025 and clocks are turned back 1 hour.

Required Readings:

Groves et al. (2009). Chapters 6 and 10.6

Recommended Readings:
AAPOR (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 9th edition.
https://www.aapor.org/AAPOR_Main/media/publications/StandardDefinitions20169theditionfinal.pdf

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics, 29*(3), 329-353.
https://content.sciendo.com/view/journals/jos/29/3/article-p329.xml

Kalton, G. & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics, 19*(2), 81-97.
http://www.websm.org/uploadi/editor/1369063357Kalton_2003_Weighting_methods.asp.pdf

Unit 10: Data Preparation, Data Processing, and Data Base Management

Video lecture: available Wednesday, October 29, 2025

Online meeting: Wednesday, November 5, 2025, 11:00 AM EST/5:00

PM CET Online quiz unit 10: due Thursday, November 6, 2025, 5:59

PM EST/11:59 PM CET

Required Readings:
Groves et al. (2009). Chapters 10.1-10.5

Foster, I. & Heus, P. (2016). Databases. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, & J. Lane (Eds.), *Big Data and Social Science: A Practical Guide to Methods and Tools*. Chapman & Hall, 93-124.
https://textbook.coleridgeinitiative.org/index.html

Unit 11: Data Analysis and Data Visualization

Video lecture: available Wednesday, November 5, 2025

Online meeting: Wednesday, November 12, 2025, 11:00 AM EST/5:00

Required Readings:
Peng & Matsui (2015). Chapters 4, 5, 7, 8, and 10

Yalcin, A. & Plaisant, C. (2016). Information Visualization. In I. Foster, R. Ghani, R.S. Jarmin, F. Kreuter, & J. Lane (Eds.), *Big Data and Social Science: A Practical Guide*

to Methods and Tools. Chapman & Hall, 243-263.
https://textbook.coleridgeinitiative.org/index.html

Unit 12: Survey and Research Ethics

Video lecture: available Wednesday, November 12, 2025

Online meeting: Wednesday, November 19, 2025, 11:00 AM EST/5:00 PM CET

*** Thanksgiving Break, November 26- November 30, 2025***

Required Readings:
Groves et al. (2009). Chapter 11

Barocas, S. & Nissenbaum, H. (2014). Big Data's end run around anonymity and consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 44-75.

Final Exam

Wednesday, December 3, 2025, 11:00 AM EST/5:00 PM CET