Syllabus

Item Nonresponse and Imputation
1 credit/2 ECTS

Prof. Jörg Drechsler

Video lecture by
Prof. Jörg Drechsler

June 20 – July 11, 2022

Short Course Description

Missing data are a common problem, which can lead to biased results if the missingness is not taken into account at the analysis stage. Imputation is often suggested as a strategy to deal with item nonresponse allowing the analyst to use standard complete data methods after the imputation. However, several misconceptions about the aims and goals (isn't imputation making up data?) of imputation make some users skeptical about the approach. In this course we will illustrate why thinking about the missing data is important and clarify which goals a useful imputation method should try to achieve (and which not).

Course Objectives

By the end of the course, students will…

- understand why the default way of dealing with missing data as implemented in most statistical software is often problematic.
- realize that it is better not to account for the missingness instead of applying simplistic imputation methods such as mean imputation or last-observation carried forward.
- know what is meant by a missing data mechanism and understand the implication of the different mechanisms.
- be familiar with the principle ideas and concepts of multiple imputation.

Prerequisites

Students should be familiar with generalized linear models and basic probability theory. **The statistical software R will be used for illustrations and for (some of) the homework assignments. Thus, basic knowledge of R is required to be able to complete the assignments.**

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 and then "attending" 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 and there will be a final exam at the end of the course.

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 videos, participating in 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 course that runs for 4 weeks. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Section 1: Mondays, 12-1 pm ET/6-7 pm CET, June 20 - July 11, 2022

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on mannheim.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 Sunday before the meeting).

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

Grading

Grading will be based on:

- 2 online quizzes (worth 20% total)
- 2 homework assignments (40% total)
- Participation in the weekly online meetings, engagement in discussions during the meetings and/or submission of questions to the weekly discussion forums (10% of grade)
- A final online exam (30% of grade)

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

The grading scale is a base scale recommended by the MDM. Variations for grading on a scale are at the discretion of the instructor.

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 assignment will be due are indicated in the syllabus. There will be a grace period for late assignments (**not for quizzes**), but late assignments will be penalized according to the following rules:
1 day late: 10% off
2 days late: 25% off
3 days late: 50% off
4+ days late: no credit

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between the students and the instructors 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

Most analysis procedures assume that the data to be analyzed are fully observed. However, in practice this will hardly ever be the case. If the data are collected through surveys, missingness can occur because respondents refuse to participate (unit nonresponse) or are unwilling or unable to respond to some of the questions in the survey (item nonresponse). In experimental designs, data might be missing due to drop-out or because some measures were not taken from all the participants due to cost constraints. In administrative data, information might be missing because implausible data entries were set to missing during the data editing stage or because no information was provided for those fields that are not mandatory. Using only the fully observed cases for analysis, which is the standard procedure implemented in most statistical software, will usually be inefficient and can introduce bias in the analysis results if the probability to be missing is related to the collected data. Nevertheless, strategies for taking the missing data properly into account are never discussed in most statistics courses. As a consequence, applied researchers tend to either rely on standard procedures implemented in the statistical software of their choice or they apply simple fixes to the problem that can cause more damage than ignoring the missingness completely.
The aim of this course is to raise the awareness that missing data can have substantial negative impacts and should always be addressed. The course will focus on imputation as a convenient tool for obtaining valid inferences in the presence of missing data. The course consists of four sections. In the first section a formal framework for modeling the response mechanism which underlies most nonresponse adjustment strategies will be introduced. In the second section students will learn which restrictive assumptions are required to obtain valid inferences based only on fully observed data. The third part of the course discusses some general misconceptions about the goals of imputation and illustrates the drawbacks of several ad-hoc imputation procedures. The final part of the course will introduce multiple imputation as a strategy to account for missing data in a principled manner.

Readings

Primary Readings to Be Purchased by Participants

Carpenter, J. and Kenward, M. (2012). *Multiple imputation and its application*. New York: John Wiley & Sons

Note: We will only use some introductory sections of this book for this course. But if you are planning to also take the multiple imputation course (SURV726) later, I highly recommend to buy this book now, since we will heavily rely on this book in the multiple imputation course.

Required and Recommended Readings

List of required and recommended readings for each class 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 in the MBS Honor Code, signed at the beginning of the program.

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 Mannheim Business School 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. Participation is entirely voluntary and highly appreciated.

Sessions

Week 1: Introduction & Missing Data Mechanisms

Video lecture: available Monday, June 13, 2022

Online meeting: Monday, June 20, 2022, noon EST / 6 PM CET

Quiz 1: due Tuesday, June 21, 2022, noon EST / 6 PM CET

Required Readings:
Carpenter, J. and Kenward, M. (2012). *Multiple imputation and its application*. New York: John Wiley & Sons, Chapter 1.1 to Chapter 1.4.4

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (2004) Survey Methodology, Wiley, Chapter 6

Week 2: Default Strategies of (Not) Dealing with Missing Data and Their Implications

Video lecture: available Monday, June 20, 2022

Online meeting: Monday, June 27, 2022, noon EST / 6 PM CET

Assignment 1: due Tuesday, June 28, 2022, noon EST / 6 PM CET

Required Readings:
Carpenter, J. and Kenward, M. (2012). *Multiple imputation and its application*. New York: John Wiley & Sons, Remainder of Chapter 1

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data (2nd ed.), New York: John Wiley & Sons, Sections 3.1, 3.2, and 3.4.

Week 3: Common Misconceptions Regarding Imputation & Basic Imputation Methods

Video lecture: Monday, June 27, 2022

Online meeting: Monday, July 4, 2022, noon EST / 6 PM CET

Assignment 2: due Tuesday, July 5, 2022, noon EST / 6 PM CET

Required Readings:
Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis with Missing Data (2nd ed.), New York: John Wiley & Sons, Chapter 4.

Brick, J.M. and Kalton, G. (1996). Handling missing data in survey research. Statistical Methods in Medical Research, 5, 215-238. Sections 1 and 3.1.

Recommended Readings:
Brick, J.M. and Kalton, G. (1996). Handling missing data in survey research. Statistical Methods in Medical Research, 5, 215-238. Sections 2 and 4.

Week 4: More Advanced Imputation Methods & Multiple Imputation

Video lecture: Monday, July 4, 2022

Online meeting: Monday, July 11, 2022, noon EST / 6 PM CET

Quiz 2 **due Monday, July 11, 2022, 11:30 AM EST / 5:30 PM** (note that this Quiz is due **before** the online meeting)

Required Readings:
Carpenter, J. and Kenward, M. (2012). *Multiple imputation and its application*. New York: John Wiley & Sons, Chapter 2.1 to Chapter 2.4

Rubin, D.B. (1986). Basic ideas of multiple imputation for nonresponse. Survey Methodology, 12, 37-47.

Recommended Readings:
Rässler, S., Rubin, D.B., Zell, E.R (2007). Incomplete data in epidemiology and medical statistics. In: Rao CR, Miller J, Rao DC (eds) Handbook of Statistics, 27, Elsevier, pp 569-601.

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of statistical software, 45(3).

Final exam

Due: Monday, July 18, 2022, noon EST / 6 PM CET