Item Nonresponse and Imputation
SURV 725
1 credit/2 ECTS
Fall 2018, Sec 1

Instructor(s)
Jörg Drechsler

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 and Learning 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.

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, for each credit offered by a course, students can expect to spend one hour per week on in-class activities and three hours per week on out-of-class activities over the span of a full 12-week term. This is a 1-credit course that runs for 4 weeks. Hence, the total average workload is about 12 hours per week.

**Mandatory Weekly Online Meetings**
Wednesday, 2:00pm-2:50pm EST/20:00-20:50 CET
Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on [https://www.elms.umd.edu/](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 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 Wednesday, 7:00am EST/13:00 CET).

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 of course use other online meeting platforms such as Google Hangout or Skype.

**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 via e-mail (10% of grade)
- A final online exam (30% of grade)
Dates of when assignment 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**
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**

Some of the required readings will be from the following volume


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.

Additional required and recommended readings will be made available on the course website: [https://www.elms.umd.edu/](https://www.elms.umd.edu/)

**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

[http://www.graduate.umd.edu/policies/misconduct.html](http://www.graduate.umd.edu/policies/misconduct.html) (University of Maryland) and


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 Disability Support Services (DSS) office to register in person for services. Please call the office to set up an appointment to register with a DSS counselor. Contact the DSS office at 301.314.7682; http://www.counseling.umd.edu/DSS/.

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.

Europe switched to winter time on October 28, while the clocks in the USA are only turned back by one hour on November 4. Please note the different time for the first online meeting: On Wednesday, October 31, the meeting will take place at 3 pm ET/8 pm CET!

Unit 1: Introduction & Missing Data Mechanisms
Online meeting (Jörg Drechsler): Wednesday, October 31, 3:00pm-3:50pm EST/20:00-20:50 CET

Homework assignment 1: Sunday, November 4, 2018, 1:00pm EST/19:00 CET

Video lecture (Jörg Drechsler): available online Wednesday, October 24

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

Unit 2: Default Strategies of (Not) Dealing with Missing Data and Their Implications
Online meeting (Jörg Drechsler): Wednesday, November 7, 2:00pm-2:50pm EST/20:00-20:50 CET

Online quiz 1: due Wednesday, November 7, 1:00pm EST/19:00 CET

Video lecture (Jörg Drechsler): available online Wednesday, October 31

Readings:


Unit 3: Common Misconceptions Regarding Imputation & Basic Imputation Methods
Online meeting (Jörg Drechsler): Wednesday, November 14, 2:00pm-2:50pm EST/20:00-20:50 CET

Homework assignment 2: Sunday, November 18, 2018, 1:00pm EST/19:00 CET

Video lecture (Jörg Drechsler): available online Wednesday, November 7

Readings:

Recommended (optional):

Unit 4: More Advanced Imputation Methods & Multiple Imputation
Online meeting (Jörg Drechsler): Wednesday, November 21, 2:00pm-2:50pm EST/20:00-20:50 CET

Online quiz 2: due Wednesday, November 21, 1:00pm EST/19:00 CET

Video lecture (Jörg Drechsler): available online Wednesday, November 14

Readings:


Recommended (optional):


**Final Exam**

Due: Wednesday, November 28, 3:00pm EST/21:00 CET

**Note:** Student access to the course website will be revoked two weeks after the final exam.