IPSDS

SPONSORED BY THE
Federal Ministry of Education and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

COURSE CATALOG 2020-2022

MANNHEIM Master of Applied Data Science & Measurement

Document last updated: June 15th, 2020

The project on which this report is based was funded by the Federal Ministry of Education and Research under the number [16OH22064]. Responsibility for the content of this publication lies with the author.

Table of Contents

IPSDS

SPONSORED BY THE
Federal Ministry of Education and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

COURSE CATALOG 2020-2022

MANNHEIM Master of Applied Data Science & Measurement

Track: Data Insights

STUDY YEAR 1-3 (JUNE 2020 – MAY 2022)

Overview

Track "Data Insights"	Instructor(s)	ECTS	Focus Area
Year 1 – Summer Term (June 2020 – August 2020)			
Connect@MDM or Elective[1]		2	Any Area
Introduction to Real World Data Management	Alexandru Cernat, PhD	4	Data Curation/Storage
ECTS Points:		**6**	
Year 1 – Fall Term (September 2020 – December 2020)			
Fundamentals of Survey and Data Science	Alexander Wenz, PhD	6	Research Design
Modern Workflows in Data Science	Alexandru Cernat, PhD	4	Data Curation/Storage
Privacy Law	Prof. Thomas Fetzer	2	Data Output/Access
ECTS Points:		**12**	
Year 2 – Spring Term (January 2021 – May 2021)			
Generalized Linear Models	Prof. Thomas Gautschi	4	Data Analysis
Data Confidentiality and Statistical Disclosure Control	Prof. Jörg Drechsler	4	Data Output/Access
Introduction to Machine Learning and Big Data	Trent D. Buskirk, PhD, Prof. Frauke Kreuter	2	Data Analysis
ECTS Points:		**10**	
Year 2 – Summer Term (June 2021 – August 2021)			
Introduction to Record Linkage with Big Data Applications	Manfred Antoni, PhD, Prof. Stefan Bender, Christian Borgs, PhD, Prof. Joseph W. Sakshaug	4	Data Generating Process
Computer-Based Content Analysis I (Theory)	Christoph Kilian Theil	2	Data Generating Process
ECTS Points:		**6**	
Year 2 – Fall Term (September 2021 – December 2021)			
Introduction to Python and SQL	Brian Kim, PhD	2	Data Curation/Storage
Experimental Design for Surveys	Ashley Amaya, PhD	4	Data Generating Process
Machine Learning II	Trent D. Buskirk, PhD, Christoph Kern, PhD	4	Data Analysis
Computer-Based Content Analysis II (Practical Project)	Christoph Kilian Theil	2	Data Generating Process
ECTS Points:		**12**	
Year 3 – Spring Term (January 2022 – May 2022)			
Introduction to Data Visualization	Prof. Richard Traunmüller	2	Data Output/Access
Project Consulting Course	Prof. Helmut Küchenhoff, Prof. Stefan Bender	12	all areas except Research Design
ECTS Points:		**14**	
Overall ECTS Points Year 1-3:		**60**	
Year 3 – Summer/Fall Term (June 2022 – December 2022)			
Master-Project[2]		15	
Total ECTS points from all tracks and the Master-Project:		**75**	

[1] Please find more information under "Elective Courses & Connect".

[2] Detailed information regarding the Master-Project can be found in the Program Guidelines.

Course Short Descriptions

Year 1: Summer Term (June 2020 – August 2020)

Introduction to Real World Data Management *Core Course* *Focus Area: Data Curation / Storage* *4 ECTS* Instructor: Alexandru Cernat, PhD (contact@alexcernat.com)	
Video lecture: Alexandru Cernat, PhD	

Short Course Description

Data is omnipresent in the contemporary world coming in different shapes and sizes: from survey data to found data. In order to make use of such data through analysis it is necessary first to import and clean it. This is often one of the most time consuming and difficult parts of data analysis. In this course you will learn both the conceptual steps needed in preparing data for analysis as well as the practical skills to do this. The course will cover all the essential skills needed to prepare data be it survey data, administrative data or found data.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- understand the stages involved in preparing data for analysis
- understand the concept of tidy data
- understand the basics of using R
- know how to write their own functions and loop over them
- know how to import and export data
- know how to clean data in R
- know how to merge data
- know how to manipulate textual data
- know how to manipulate date/time data
- know how to use tables and graphs to explore data

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. The basics of R
2. Practical 1
3. Data importing and cleaning

4. Practical 2
5. Exploring Data
6. Practical 3
7. Learning how to deal with text and date information
8. Practical 4

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- 4 fortnightly homework assignments (worth 60% total)
- Participation in discussion during the weekly online meetings and submission of questions to the forum (demonstrating understanding of the required readings and video lectures (worth 10%)
- A final project (worth 30%)

Year 1: Fall Term (September 2020 – December 2020)

Fundamentals of Survey and Data Science

Core Course
Focus Area: Research Design

6 ECTS

Instructor:
Alexander Wenz, PhD (a.wenz@uni-mannheim.de)

Video lecture: Florian Keusch, Jennifer Sinibaldi

Short Course Description

This 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 students 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.

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.

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.

Course Composition

This is a 6 ECTS course that runs for 12 weeks. The content of the course is broken down into twelve units:

1. Introduction – How to do survey research and data science

2. Quality of Data
3. Coverage
4. Modes of Survey Data Collection
5. Data Generation from Other Sources
6. Sampling I
7. Sampling II
8. Questionnaires and Interviewing
9. Nonresponse
10. Data Preparation, Data Processing, and Data Base Management
11. Data Analysis and Data Visualization
12. Survey and Research Ethics

Learning and Teaching Methods

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. 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. Just like in an on-site course, assignments will be assigned and graded, and there will be a final exam at the end of the course.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and contributions to the forum demonstrating understanding of the required readings and video lectures (10% of the final grade).
- Ten online quizzes reviewing specific aspects of the material covered (60% of the final grade).
- A final, open-book, online exam (30% of the final grade).

Modern Workflows in Data Science

Core Course
Focus Area: Data Curation/Storage

4 ECTS

Instructor:
Alexandru Cernat, PhD (contact@alexcernat.com)

Video lecture: Alexandru Cernat

Short Course Description

Large data, fast pace of production, and collaboration are hallmarks of the new data environment. In this context, researchers must have a good understanding of data workflows and they must ensure consistent and reproducible practices in order to collaborate and consistently produce insights. This course deals with some of these essential topics. We will discuss the main types of workflows in data and survey sciences and how tools such as GitHub can enhance collaboration and insure reproducibility. We will also discuss the use of reproducible documents such as Rmarkdown or Jupyter Notebooks before covering the best practices for working with online data sources. We will finish the course by discussing the use of dashboards and how to develop such a tool using R Shiny.

Prerequisites

Intro to Real World Data Management with R course or a good knowledge of R base and tidyverse.

Course Objectives

By the end of the course, students will…

- understand the main types of workflows in data and survey sciences
- understand the principles of reproducible workflows
- know how to use Github to support reproducible flows
- Understand the basics of reproducible documents
- learn how to use Rmarkdown and Jupyter Notebooks
- learn about the main types of storage for online data (e.g., SQL, JSON)
- learn how to access distributed clusters using Spark
- learn how to manage computing clusters
- learn the principles of building a dashboard
- learn how to build a dashboard using R Shiny

Course Composition

This is a 4 ECTS course that runs for 8 weeks. The content of the course is broken down into 4 topics, each topic will be covered in two weeks:

1. Data workflows with Github

2. Reproducible documents with Rmarkdown and Jupyter Notebooks
3. Accessing data online
4. Interactive dashboards with shiny

Learning and Teaching Methods

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. 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. Just like in an on-site course, assignments will be assigned and graded.

Grading

Grading will be based on:

- Four homework assignments (worth 60% total)
- Participation in discussion during the weekly online meetings and submission of questions via e-mail (deadline: Monday, 8AM before class) demonstrating understanding of the required readings and video lectures (worth 10%)
- A final project (worth 30%)

Privacy Law

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Prof. Thomas Fetzer (fetzer@jura.uni-mannheim.de)

Video lecture: Prof. Thomas Fetzer

Short Course Description

The course will acquaint the students with the origins and basic principles of privacy law mainly in Europe. Furthermore, it will contrast the European privacy foundations with the U.S. approach. At the core of this course stands the new European General Data Protection Regulation (GDPR) and its applicability to specific cases and basic principles. Moreover, the course will cover current challenges to the existing privacy paradigms by big data and big data analytics.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- have a basic knowledge on the foundations of privacy law in Europe and the U.S.;
- have an understanding why privacy issues are treated differently in Europe and the U.S.;
- have a basic knowledge on the applicability of the General Data Protection Regulation (GDPR) and its basic principles;
- be aware of privacy issues and potential legal limitations when processing data;
- be aware of current challenges to the existing privacy paradigms by big data and big data analytics;
- be aware of currently discussed new approaches to privacy (e.g. privacy by design).

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction to privacy law
2. The GDPR I
3. The GDPR II
4. The GDPR III and current challenges to the existing privacy paradigms

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and submission of questions via e-mail demonstrating understanding of the required readings and video lectures (worth 10%)
- Three online quizzes reviewing specific aspects of the material covered (worth 60%)
- A final open-book online exam (worth 30%)

Year 2: Spring Term (January 2021 – May 2021)

Generalized Linear Models

Core Course
Focus Area: Data Analysis

4 ECTS

Instructor:
Prof. Thomas Gautschi (gautschi@sowi.uni-mannheim.de)

Video lecture: Prof. Thomas Gautschi

Short Course Description

The main focus of this course lies on the introduction to statistical models and estimators beyond linear regression useful to social and economic scientists. It provides an overview of generalized linear models (GLM) that encompass non-normal response distributions to model functions of the mean. GLMs thus relate the expected mean E(*Y*) of the dependent variable to the predictor variables via a specific link function. This link function permits the expected mean to be non-linearly related to the predictor variables. Examples for GLMs are the logistic regression, regressions for ordinal data, or regression models for count data. GLMs are generally estimated by use of maximum likelihood estimation. The course thus not only introduces GLMs but starts with an introduction to the principle of maximum likelihood estimation. A good understanding of the classical linear regression model is a prerequisite and required for the course.

Prerequisites

A sound understanding of linear regression models (OLS) is required. Knowledge in linear algebra and calculus is useful.

Course Objectives

By the end of the course, students will…

- understand how to appropriately translate research questions into statistical models
- be able to apply statistical models appropriate for non-linear problems
- estimate regression parameters using the maximum likelihood principle
- perform hypothesis tests for regression models using the maximum likelihood principle
- be able to identify limitations of non-linear regression models
- be able to identify violations of the respective regression assumptions of the discussed GLMs

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 4 units:

1. Maximum Likelihood Estimation

2. Binary Choice Models
3. Models for Ordinal Data
4. Models for Count Data

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 7 homework assignments (worth 49% total)
- Participation in online meetings and submission of questions demonstrating understanding of readings (worth 10%)
- Final Exam (worth 41%)

Data Confidentiality and Statistical Disclosure Control

Core Course
Focus Area: Data Output / Access

4 ECTS

Instructor:
Prof. Jörg Drechsler (joerg.drechsler@iab.de)

Video lecture: Prof. Jörg Drechsler

Short Course Description

This course will provide a gentle introduction to statistical disclosure control with a focus on generating synthetic data for maintaining the confidentiality of the survey respondents. The first part of the course will introduce several traditional approaches for data protection that are widely used at statistical agencies. Some limitations of these approaches will also be discussed. The second part of the course will introduce synthetic data as a possible alternative. This part of the course will discuss different approaches to generating synthetic datasets in detail. Possible modeling strategies and analytical validity evaluations will be assessed and potential measures to quantify the remaining risk of disclosure will be presented. To provide the participants with hands on experience, all steps will be illustrated using simulated and real data examples in R.

Prerequisites

The statistical software R will be used for illustrations and for (some of) the homework assignments. Thus, knowledge of R is required to be able to complete the assignments. Some background regarding general linear modelling is expected. Familiarity with the concept of Bayesian statistics is helpful but not required.

Course Objectives

By the end of the course, students will…

- know which measures are typically taken by statistical agencies to guarantee confidentiality for the survey respondents if data are disseminated to the public.
- be aware of potential limitations of these measures.
- have a practical understanding of the concept of synthetic data.
- be able to judge in which situations the approach could be useful.
- know how to generate synthetic data from their own data.
- have a number of tools available to evaluate the analytical validity of the synthetic datasets.
- know how to assess the disclosure risk of the generated data.

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. A Brief History of Data Confidentiality & Traditional Approaches for Data Protection
2. The Computer Science Perspective on Data Privacy & Introduction to Multiply Imputed Synthetic Datasets
3. Analyzing Synthetic Datasets & Relationship to Multiple Imputation for Nonresponse
4. Synthesis Models Part I (Univariate and Linear Regression Models)
5. Synthesis Models Part II (Models for Categorical Variables and Nonparametric Models) & Modeling Strategies
6. Analytical Validity & Disclosure Risk Part I (Theory)
7. Disclosure Risk Part II (Examples in R) & Discussion of the Chances and Obstacles of the Synthetic Data Approach
8. Discussion of the Third Homework Assignment

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 2 quizzes (worth 15% total)
- Participation in the weekly online meetings, engagement in discussions during the meetings and/or submission of questions via e-mail (worth 10%)
- Three homework assignments (worth 45% total)
- A final online exam (worth 30%)

Introduction to Machine Learning and Big Data

Core Course
Focus Area: Data Analysis

2 ECTS

Instructors:
Trent D. Buskirk, PhD (buskirk@bgsu.edu)
Prof. Frauke Kreuter (fkreuter@umd.edu)


Video lecture: Trent D. Buskirk, PhD, Prof. Frauke Kreuter

Short Course Description

The amount of data generated as a by-product in society is growing fast, including data from satellites, sensors, transactions, social media and smartphones, just to name a few. Such data are often referred to as “big data”, and can be used to create value in different areas such as health and crime prevention, commerce and fraud detection. Big Data are often used for prediction and classification tasks. Both of which can be tackled with machine learning techniques. In this course we explore how Big Data concepts, processes and methods can be used within the context of Survey Research. Throughout this course we will illustrate key concepts using specific survey research examples including tailored survey designs and nonresponse adjustments and evaluation.

Prerequisites

No prerequisite.
We recommend good understanding of the material typically taught in undergraduate statistics courses and some familiarity with regression techniques. Knowledge about survey data collection at the level provided in the MDM course Fundamentals of Survey and Data Science.
While not a prerequisite, familiarity with the R software package (base R or R using Rstudio) is strongly encouraged.

Course Objectives

This course covers

- an overview of key Big Data terminology and concepts
- an introduction to common data generating processes
- a discussion of some primary issues with linking Big Data with Survey Data
- issues of coverage and measurement errors within the Big Data context
- inference versus prediction
- general concepts from machine learning including signal detection and information extraction
- potential pitfalls for inference from Big Data
- key analytic techniques (e.g. classification trees, random forests, conditional forests) to process Big Data using R with example code provided

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Overview of Big Data; Working with Big Data; Classical Statistical Approaches versus Statistical Machine Learning
2. Model Evaluation/Validation; K-Means Clustering
3. K Nearest Neighbors; CARTS
4. Random Forests

Learning and Teaching Methods

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. In addition, students are encouraged to post at least one questions about the materials covered in the videos and readings of the week in the forum before the meetings. Just like in an on-site course, homework will be assigned and graded.

Grading

Grading will be based on:

- 4 online quizzes (worth 20% total)
- Participation in discussion during the weekly online meetings and submission of questions via discussion form demonstrating understanding of the required readings and video lectures (worth 20%). Obviously in the first week one question will be enough, since we just started.
- 3 homework assignments (worth 60% total)

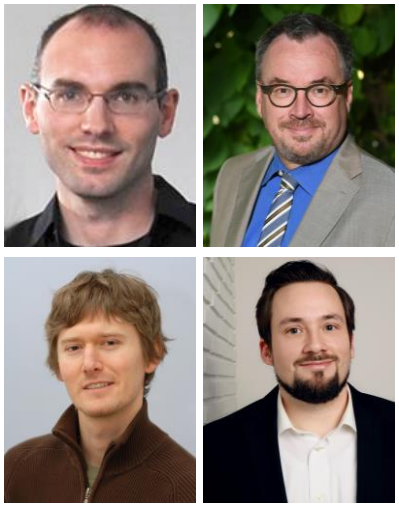
Year 2: Summer Term (June 2021 – August 2021)

Introduction to Record Linkage with Big Data Applications

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructors:
Manfred Antoni, PhD (manfred.antoni@iab.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)
Christian Borgs, PhD (christian.borgs@uni-duisburg-essen.de)
Prof. Joseph W. Sakshaug (joe.sakshaug@iab.de)



Video lecture:
Manfred Antoni, PhD, Prof. Stefan Bender, Christian Borgs, PhD, Prof. Joseph W. Sakshaug

Short Course Description

The course will address methods to combine data on given entities (people, households, firms etc.) that are stored in different data sources. By showing the strengths of these methods and by showing how each of them are performed in practice using R, the course will demonstrate the various benefits of record linkage. Participants will also learn about potential challenges that record linkage projects may face.

Prerequisites

Students should have knowledge of basic statistical concepts. They need to have an intermediate knowledge of R. Familiarity with regular expressions, the R packages ggplot2 and tidyverse is useful but not required.

Course Objectives

By the end of the course, students will…

- be familiar with a host of record linkage applications from different countries or jurisdictions that link a variety of data sources and use different types of linkage
- know how to improve the quality of linkage identifiers by applying pre-processing routines
- be familiar with different methods of increasing the efficiency of record linkage
- be able to understand, select and apply appropriate record linkage methods (e.g., deterministic and probabilistic linkage)
- be able to evaluate the success of data linkage
- be able to perform each step in the record linkage process using the R software

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introducing record linkage in the age of Big Data
2. Collecting and pre-processing linkage identifiers & blocking techniques
3. Data preprocessing and core concepts of data quality for linking
4. Comparison and classification of record pairs
5. Probabilistic record linkage and blocking (application)
6. Advanced topics, software options and literature review
7. Privacy-preserving record linkage using R
8. Evaluation and visualization of linkage quality

Learning and Teaching Methods

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. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week in the forum on the course page before the meetings.

Grading

Grading will be based on:

- 3 online quizzes (worth 30% total)
- Participation in the weekly online meetings (worth 20%): engagement in discussions during the meetings and submission of questions in the forum on the course website
- 3 homework assignments (worth 50% total)

Computer-Based Content Analysis I (Theory)

Core Course
Focus Area: Data Generating Process

2 ECTS

Instructor:
Christoph Kilian Theil (christoph@informatik.uni-mannheim.de)

Video lecture: Christoph Kilian Theil

Short Course Description

This course investigates the foundations of Natural Language Processing (NLP) as tool for analyzing natural language texts in the social sciences, thus providing an alternative to traditional ways of data generation through surveys. The course introduces general use cases for NLP, provides a guide to standard operations on text as well as their implementation in the Python-based Natural Language Toolkit (NLTK) and introduces the text mining functionalities of the WEKA Machine Learning workbench.
The theory part of the course worth two ECTS can be supplemented by an optional project part worth two more ECTS.

Prerequisites

Participants need to have attended the following MDM courses or have corresponding knowledge:

1. Introduction to Python and SQL or necessary knowledge in programming in Python: data types & structures, functions & loops, file I/O
2. Web Scraping (recommended)

Course Objectives

By the end of the course, students will be able to …:

- understand the possibilities and limitations of automatic text analysis
- judge the potential benefits of applying automatic text analysis to a given research question
- preprocess a corpus using the Natural Language Toolkit (NLTK)
- perform text classification using the WEKA Machine Learning workbench
- understand the principles of advanced text analysis methods

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction – Potential and Use Cases of Text Analysis Methods

2. The Natural Language Toolkit
3. Text Mining with WEKA
4. Advanced Text Mining

Learning and Teaching Methods

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. 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. Just like in an on-site course, homework will be assigned and graded and there will be a final project at the end of the course.

Grading

Grading will be based on:

- Participation in online meetings (worth 10%)
- Answering questions about the content of the videos – 4 quizzes (worth 15% total)
- Practical application of NLP and Machine Learning technologies – 4 assignments (worth 75% total)

Year 2: Fall Term (September 2021 – December 2021)

Introduction to Python and SQL

Core Course
Focus Area: Data Curation / Storage

2 ECTS

Instructor:
Brian Kim, PhD (kimbrian@umd.edu)

Video lecture: Brian Kim, PhD

Short Course Description

This course introduces students to the basics of Python and SQL for data analysis. Students will explore real publicly-available datasets, using the data analysis tools in Python to create summaries and generate visualizations. Students will learn the basics of database management and organization, as well as learn how to code in SQL and work with PostgreSQL databases. By the end of the class, students should understand how to read in data from CSV files or from the internet and be comfortable using either SQL or Python to aggregate, summarize, describe, and visualize these datasets.

Prerequisites

No prerequisites

Course Objectives

By the end of the course, students will

- understand the basic structure of how Python and object-oriented programming works
- be able to write basic Python code, including functions and loops
- know how to use Pandas and matplotlib packages in Python to analyze data and create visualizations
- be comfortable reading error messages and Python documentation to diagnose and debug code
- understand how relational databases work
- be able to construct a query to answer questions about the data
- understand how joins work and how to use them

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction to Python and Pandas
2. Functions, Loops and Visualizations

3. Introduction to SQL
4. Joins

Learning and Teaching Methods

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. In addition, students must post questions or comments about the materials covered in the videos and workbooks of the week to the instructor in the forum before the meetings.

Grading

Grading will be based on:

- 4 online quizzes (worth 20% total)
- Participation in discussion during the weekly online meetings and posting questions in the forum demonstrating understanding of required readings and video lectures (worth 20%)
- 4 homework assignments (worth 60% total)

Experimental Design for Surveys

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Ashley Amaya, PhD (aamaya@rti.org)

Video lecture: Roger Tourangeau, PhD

Short Course Description

This course examines how to embed experiments in surveys. It covers both the design of survey experiments and the analysis of the results.

Prerequisites

At least one prior course in data analysis. Ability to use SAS or STATA.

Course Objectives

By the end of the course, students will

- know about basic principles of experimental design
- recognize the main types of experimental designs
- improve the quality of designs used to carry out methodological research in or for surveys
- develop critical skills to spot flaws in experimental and nonexperimental designs to support causal inferences
- improve skills at analyzing results of survey experiments
- improve skills as both consumer and producer of experiments done to shed light on survey methodological issues

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introduction
2. Examples of Experiments in Surveys
3. Experimental Designs I
4. Experimental Designs II
5. Comparability and Generalizability
6. Construct Validity I
7. Construct Validity II; Statistical Validity
8. Wrap-Up

Learning and Teaching Methods

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. 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. As with an on-site course, homework will be assigned and graded and there will be three quizzes.

Grading

Grading will be based on:

- 3 Online quizzes (worth 45% total)
- 3 exercises (worth 45% total)
- Participation in online discussions (worth 10%)

Machine Learning II

Core Course
Focus Area: Data Analysis

4 ECTS

Instructors:
Trent D. Buskirk, PhD (buskirk@bgsu.edu)
Christoph Kern, PhD (c.kern@uni-mannheim.de)

Video lecture: Christoph Kern, PhD, Trent, D. Buskirk, PhD

Short Course Description

Social scientists and survey researchers are confronted with an increasing number of new data sources such as apps and sensors that often result in (para)data structures that are difficult to handle with traditional modeling methods. At the same time, advances in the field of machine learning (ML) have created an array of flexible methods and tools that can be used to tackle a variety of modeling problems. Against this background, this course discusses advanced ML concepts such as cross validation, class imbalance, Boosting and Stacking as well as key approaches for facilitating model tuning and performing feature selection. In this course we also introduce additional machine learning methods including Support Vector Machines, Extra-Trees and LASSO among others. The course aims to illustrate these concepts, methods and approaches from a social science perspective. Furthermore, the course covers techniques for extracting patterns from unstructured data as well as interpreting and presenting results from machine learning algorithms. Code examples will be provided using the statistical programming language R.

Prerequisites

Topics covered in the course Introduction to Machine Learning and Big Data (ML I), i.e.:

- Conceptual basics of machine learning (training vs. test data, model evaluation basics)
- Decision trees with CART
- Random forests

Familiarity with the statistical programming language R is strongly recommended.

Participants are encouraged to work through one or more R tutorials prior to the first-class meeting. Some resources can be found here:

- https://rstudio.cloud/learn/primers
- http://www.statmethods.net/
- https://swirlstats.com/
- https://www.rcommander.com

Course Objectives

By the end of the course, students will…

- will have a profound understanding of advanced (ensemble) prediction methods
- have built up a comprehensive ML toolkit to tackle various learning problems
- know how to (critically) evaluate and interpret results from ''black-box'' models

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Intro: Bias-variance trade-off, cross-validation (stratified splits, temporal cv) and model tuning (grid and random search)
2. Classification: Performance metrics (ROC, PR curves, precision at K) and class imbalance (over- and undersampling, SMOTE)
3. Ensemble methods I: Bagging and Extra-Trees
4. Ensemble methods II: Boosting (Adaboost, GBM, XGBoost) and Stacking
5. Variable selection: Lasso, elastic net and fuzzy/ recursive random forests
6. Support Vector Machines
7. Advanced unsupervised learning: Hierarchical clustering and LDA
8. Interpreting (Variable Importance, PDP, ...) and reporting ML results

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- 4 homework assignments (worth 40% total)
- 8 online quizzes (worth 40% total)
- Participation in discussion during the weekly online meetings (worth 20%)

Computer-Based Content Analysis II (Practical Project)

Core Course
Focus Area: Data Generating Process

2 ECTS

Instructor:
Christoph Kilian Theil (kilian@informatik.uni-mannheim.de)

Video lecture: Christoph Kilian Theil

Short Course Description

This course investigates the practical application of Natural Language Processing (NLP) for analyzing textual data with the goal to answer questions of the social sciences. More specifically, participants of the previous theoretical class will now have the opportunity to implement their own practical research project. Under the guidance of the instructor, the participants will define their own research question; develop a suitable methodology to address this question; conduct and discuss experiments based on the selected methods; and synthesize the results to answer their own-defined research question.

Prerequisites

Participants need to have attended the following MDM courses or have corresponding knowledge:

1. Introduction to Python and SQL or necessary knowledge in programming in Python: data types & structures, functions & loops, file I/O
2. Web Scraping (recommended)
3. Computer-Based Content Analysis Part 1 (Theory)

Course Objectives

By the end of the course, students will be able to:

- integrate text analysis into a research methodology and solve a research question from their field using this methodology
- define a methodology for solving a research problem that includes automatic text analysis
- select and apply appropriate methods for preprocessing textual resources relevant for their research question.
- select and apply text mining methods to the preprocessed textual resources and conduct systematic experiments.
- interpret the experimental results and draw conclusions concerning their research question.

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Early Phase Research Project: Definition of Methodology
2. Data Preparation Phase: Application of Methods for Preprocessing
3. Text Analysis Phase: Application of Text Mining Methods
4. Final phase: Correlation of Results with External Variables

Learning and Teaching Methods

In this course, you are responsible for 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. 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. Just like in an on-site course, homework will be assigned and there will be a final project report at the end of the course.

Grading

Grading will be based on:

- Participation in online meetings (worth 25%)
- Final Project Report (worth 75%)

Year 3: Spring Term (January 2022 – May 2022)

Introduction to Data Visualization

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Prof. Richard Traunmüller (rtraunmu@mail.uni-mannheim.de)

Video lecture: Prof. Richard Traunmüller

Short Course Description

Data visualization is one of the most powerful tools to explore, understand and communicate patterns in quantitative information. At the same time, good data visualization is a surprisingly difficult task and demands three quite different skills: substantive knowledge, statistical skill, and artistic sense. The course is intended to introduce participants to key principles of analytic design and useful visualization techniques for the exploration and presentation of univariate and multivariate data. This course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. Students will learn how to evaluate data visualizations based on principles of analytic design, how to construct compelling visualizations using the free statistics software R, and how to explore and present their data with visual methods.

Prerequisites

Course prerequisites are a basic understanding of statistics and bivariate linear regression. Some experience in the use of a statistical software package would help, preferable basic data management tasks in R (loading and merging data, generating and recoding variables, etc.) I will provide detailed code examples to get students up to pace.

Course Objectives

By the end of the course, students will…

- know how to evaluate and criticize data visualizations based on principles of analytic design
- be in the position to explore and present their data with visual methods
- understand which graphical formats are useful for which types of data and questions
- know how to construct compelling visualizations using the free statistics software R

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Data Visualization as a Methodology

2. Principles of Analytic Design
3. Making Visual Comparisons
4. Visualizing Relations Between Variables

Learning and Teaching Methods

In this course, you are responsible for 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. 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. Just like in an on-site course, homework will be assigned and there will be a final project report at the end of the course.

Grading

Grading will be based on:

- Participation at online meetings (worth 10%)
- Required Course Assignments (worth 40% total)
- Final Visualization Project (worth 50%)

Project Consulting Course

Core Course
Focus Area: all areas except Research Design

12 ECTS

Instructors:
Prof. Helmut Küchenhoff (kuechenhoff@stat.uni-muenchen.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)

Short Course Description

In this course, students will apply the core skills that they learned in the MDM program to address real-world problems. The course will provide experience with the steps involved in carrying out a data consulting project, such as discussing the problems to solve with a client, data handling, and communicating work in both written and oral forms. The project is completed in teams (3-4 students per team).

Prerequisites

- Successful completion of the course “Fundamentals of Survey and Data Science”
- Experience with descriptive statistics, inferential statistics and linear modeling
- Familiarity with R or Python (you should be able to clean and manipulate data using one of these programing languages)

Course Objectives

By the end of the course, students will…

- be able to apply skills and concepts learned throughout the program to solve a real-world problem for a given client,
- be able to successfully manage a project in a timely manner,
- be able to collaborate with team members as well as data clients,
- be able to effectively communicate results in both written and oral forms.

Course Composition

This is a 12 ECTS course, which runs for 14 weeks. The content of the course is broken down into 14 units:

1. General introduction
2. Cooperation with clients / Assigning projects
3. Meeting with clients
4. Teams work on projects
5. Teams work on projects
6. Teams work on projects
7. Presentations of project progress / Feedback
8. Presentations of project progress / Feedback

9. Communicating results: effective presentation and report
10. Teams work on projects
11. Teams work on projects
12. Final presentations (part 1)
13. Final presentations (part 2)
14. Teams work on projects

Learning and Teaching Methods

This is a flipped online course in which students will form groups to collaboratively work on problems involving data presented to them by outside consultants. Students will work in teams to scope a problem, analyze data provided to them, distribute work among team members, and combine their results for a joint report and presentation.

The first part of the course emphasizes individual work. Students are expected to watch video-recorded lectures, read the required literature for each unit, and attend online discussion meetings with the instructor. In some meetings, students will meet with clients. The second part of the course will emphasize project-based work involving data analysis. Student will collaborate with their group members via online tools (e.g., Zoom meetings) to complete their projects. They will provide periodic progress updates about their group projects and present their work orally and in writing. Just like in an on-site course, homework will be assigned and graded and there will be a final report at the end of the course.

Grading

Grading will be based on:

- Participation in online meetings (worth 10%)
- First meeting with clients and results protocol (worth 20%)
- Final presentation (worth 35%)
- Final report (worth 35%)

IPSDS

SPONSORED BY THE
Federal Ministry
of Education
and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

COURSE CATALOG 2020

MANNHEIM Master of Applied Data Science & Measurement

Track: Survey Statistics

STUDY YEAR 1-3 (JUNE 2020 – MAY 2022)

Overview

Track "Survey Statistics"	Instructor(s)	ECTS	Focus Area
Year 1 – Summer Term (June 2020 – August 2020)			
Connect@MDM or Elective[1]		2	Any Area
Introduction to Real World Data Management	Alexandru Cernat, PhD	4	Data Curation/Storage
ECTS Points:		**6**	
Year 1 – Fall Term (September 2020 – December 2020)			
Fundamentals of Survey and Data Science	Alexander Wenz, PhD	6	Research Design
Sampling I	Raphael Nishimura, PhD	4	Data Generation Process
Item Nonresponse and Imputation	Prof. Jörg Drechsler	2	Data Analysis
ECTS Points:		**12**	
Year 2 – Spring Term (January 2021 – May 2021)			
Generalized Linear Models	Prof. Thomas Gautschi	4	Data Analysis
Data Confidentiality and Statistical Disclosure Control	Prof. Jörg Drechsler	4	Data Output/Access
Sampling II	Raphael Nishimura, PhD	2	Data Generating Process
Multiple Imputation – Why and How	Prof. Jörg Drechsler	2	Data Curation/Storage
ECTS Points:		**12**	
Year 2 – Summer Term (June 2021 – August 2021)			
Introduction to Record Linkage with Big Data Applications	Manfred Antoni, PhD, Prof. Stefan Bender, Christian Borgs, PhD, Prof. Joseph W. Sakshaug	4	Data Generating Process
Measurement Error Models	Laura Boeschoten, PhD	2	Data Analysis
ECTS Points:		**6**	
Year 2 – Fall Term (September 2021 – December 2021)			
Practical Inference from Complex Surveys	Robert Fay, PhD	4	Data Analysis
Practical Tools for Survey Weighting (Part II)	Stefan Zins, PhD	2	Data Analysis
Elective		2	
ECTS Points:		**10**	
Year 3 – Spring Term (January 2022 – May 2022)			
Introduction to Data Visualization	Prof. Richard Traunmüller	2	Data Output/Access
Project Consulting Course	Prof. Helmut Küchenhoff	12	all areas except Research Design
ECTS Points:		**14**	
Overall ECTS Points Year 1-3:		**60**	
Year 3 – Summer/Fall Term (June 2022 - December 2022)			
Master-Project[2]		15	
Overall ECTS-Punkte of the whole Tracks:		**75**	

[1] Please find more information under “Elective Courses & Connect”.

[2] Detailed information regarding the Master-Project can be found in the Program Guidelines.

Course Short Descriptions

Year 1: Summer Term (June 2020 – August 2020)

Introduction to Real World Data Management *Core Course* *Focus Area: Data Curation / Storage* *4 ECTS* Instructor: Alexandru Cernat, PhD (contact@alexcernat.com)	
Video lecture: Alexandru Cernat, PhD	

Short Course Description

Data is omnipresent in the contemporary world coming in different shapes and sized: from survey data to found data. In order to make use of such data through analysis it is necessary first to import and clean it. This is often one of the most time consuming and difficult parts of data analysis. In this course you will learn both the conceptual steps needed in preparing data for analysis as well as the practical skills to do this. The course will cover all the essential skills needed to prepare data be it survey data, administrative data or found data.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- understand the stages involved in preparing data for analysis
- understand the concept of tidy data
- understand the basics of using R
- know how to write their own functions and loop over them
- know how to import and export data
- know how to clean data in R
- know how to merge data
- know how to manipulate textual data
- know how to manipulate date/time data
- know how to use tables and graphs to explore data

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. The basics of R
2. Practical 1
3. Data importing and cleaning

4. Practical 2
5. Exploring Data
6. Practical 3
7. Learning how to deal with text and date information
8. Practical 4

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- 4 fortnightly homework assignments (worth 60% total)
- Participation in discussion during the weekly online meetings and submission of questions to the forum (demonstrating understanding of the required readings and video lectures (worth 10%)
- A final project (worth 30%)

Year 1: Fall Term (September 2020 – December 2020)

Fundamentals of Survey and Data Science

Core Course
Focus Area: Research Design

6 ECTS

Instructor:
Alexander Wenz, PhD (a.wenz@uni-mannheim.de)

Video lecture: Florian Keusch, Jennifer Sinibaldi

Short Course Description

This 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 students 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.

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

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.

Course Composition

This is a 6 ECTS course that runs for 12 weeks. The content of the course is broken down into twelve units:

1. Introduction – How to do survey research and data science

2. Quality of Data
3. Coverage
4. Modes of Survey Data Collection
5. Data Generation from Other Sources
6. Sampling I
7. Sampling II
8. Questionnaires and Interviewing
9. Nonresponse
10. Data Preparation, Data Processing, and Data Base Management
11. Data Analysis and Data Visualization
12. Survey and Research Ethics

Learning and Teaching Methods

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. 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. Just like in an on-site course, assignments will be assigned and graded, and there will be a final exam at the end of the course.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and contributions to the forum demonstrating understanding of the required readings and video lectures (worth 10%).
- Ten online quizzes reviewing specific aspects of the material covered (worth 60% total).
- A final, open-book, online exam (worth 30%).

Sampling I

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Raphael Nishimura, PhD (raphaeln@umd.edu)

Video lecture: Raphael Nishimura, PhD

Short Course Description

Sampling is an applied statistics methods course, but differs from most statistics courses because it is concerned almost exclusively with the *design* of data collection. Little of the analysis of collected data will be discussed in the course. The course will concentrate on problems of applying sampling methods to human populations, since sampling human populations poses a number of particular problems not found in sampling of other types of units. The principles of sample selection, though, can be applied to many other types of populations.

Prerequisites

The course is presented at an intermediate statistical level. While we will not develop mathematical aspects of sampling theory, statistical notation and outlines of some algebraic proofs will be given. A sound background in applied statistics, proficiency in mathematics, including basic algebra, is necessary, since some algebraic derivations will be presented (although little emphasis will be placed on the derivations). A thorough understanding of the notation and algebraic results will be required.

Course Objectives

By the end of the course, students will…

- understand the basic ideas, concepts and principles of probability sampling from an applied perspective
- be able to identify and appropriately apply sampling techniques to survey design problems
- be able to compute the sample size for a variety of sample designs
- understand and be able to assess the impact of the sample design on survey estimates
- be able to estimate the precision of the survey statistics using different estimation techniques

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introduction; Course Perspectives
2. Simple Random Sampling, Sampling Frames, and Introduction to Clustering

3. Stratified Sampling I
4. Stratified Sampling II, Systematic selection
5. Cluster Sampling
6. Unequal-Sized Clusters I
7. Unequal-Sized Clusters II
8. Variance Estimation

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- Homework assignments (worth 50% total)
- Quizzes (worth 15% total)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail demonstrating understanding of the required readings and video lectures, and positive contributions on Piazza, see below (worth 10%)
- A final open-book online exam (worth 25%)

Item Nonresponse and Imputation

Core Course
Focus Area: Data Analysis

2 ECTS

Instructor:
Prof. Jörg Drechsler (joerg.drechsler@iab.de)

Video lecture: Prof. 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).

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.

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.

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction & Missing Data Mechanisms
2. Default Strategies of (Not) Dealing with Missing Data and Their Implications
3. Common Misconceptions Regarding Imputation & Basic Imputation Methods
4. Stratified Sampling II, Systematic selection

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- Homework assignments (worth 50% total)
- Quizzes (worth 15% total)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail demonstrating understanding of the required readings and video lectures, and positive contributions on Piazza, see below (worth 10%)
- A final open-book online exam (worth 25%)

Year 2: Spring Term (January 2021 – May 2021)

Generalized Linear Models

Core Course
Focus Area: Data Analysis

4 ECTS

Instructor:
Prof. Thomas Gautschi (gautschi@sowi.uni-mannheim.de)

Video lecture: Prof. Thomas Gautschi

Short Course Description

The main focus of this course lies on the introduction to statistical models and estimators beyond linear regression useful to social and economic scientists. It provides an overview of generalized linear models (GLM) that encompass non-normal response distributions to model functions of the mean. GLMs thus relate the expected mean E(*Y*) of the dependent variable to the predictor variables via a specific link function. This link function permits the expected mean to be non-linearly related to the predictor variables. Examples for GLMs are the logistic regression, regressions for ordinal data, or regression models for count data. GLMs are generally estimated by use of maximum likelihood estimation. The course thus not only introduces GLMs but starts with an introduction to the principle of maximum likelihood estimation. A good understanding of the classical linear regression model is a prerequisite and required for the course.

Prerequisites

A sound understanding of linear regression models (OLS) is required. Knowledge in linear algebra and calculus is useful.

Course Objectives

By the end of the course, students will…

- understand how to appropriately translate research questions into statistical models
- be able to apply statistical models appropriate for non-linear problems
- estimate regression parameters using the maximum likelihood principle
- perform hypothesis tests for regression models using the maximum likelihood principle
- be able to identify limitations of non-linear regression models
- be able to identify violations of the respective regression assumptions of the discussed GLMs

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 4 units:

1. Maximum Likelihood Estimation
2. Binary Choice Models

3. Models for Ordinal Data
4. Models for Count Data

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 7 homework assignments (worth 49% total)
- Participation in online meetings and submission of questions demonstrating understanding of readings (worth 10%)
- Final Exam (worth 41%)

Data Confidentiality and Statistical Disclosure Control

Core Course
Focus Area: Data Output / Access

4 ECTS

Instructor:
Prof. Jörg Drechsler (joerg.drechsler@iab.de)

Video lecture: Prof. Jörg Drechsler

Short Course Description

This course will provide a gentle introduction to statistical disclosure control with a focus on generating synthetic data for maintaining the confidentiality of the survey respondents. The first part of the course will introduce several traditional approaches for data protection that are widely used at statistical agencies. Some limitations of these approaches will also be discussed. The second part of the course will introduce synthetic data as a possible alternative. This part of the course will discuss different approaches to generating synthetic datasets in detail. Possible modeling strategies and analytical validity evaluations will be assessed and potential measures to quantify the remaining risk of disclosure will be presented. To provide the participants with hands on experience, all steps will be illustrated using simulated and real data examples in R.

Prerequisites

The statistical software R will be used for illustrations and for (some of) the homework assignments. Thus, knowledge of R is required to be able to complete the assignments. Some background regarding general linear modelling is expected. Familiarity with the concept of Bayesian statistics is helpful but not required.

Course Objectives

By the end of the course, students will…

- know which measures are typically taken by statistical agencies to guarantee confidentiality for the survey respondents if data are disseminated to the public.
- be aware of potential limitations of these measures.
- have a practical understanding of the concept of synthetic data.
- be able to judge in which situations the approach could be useful.
- know how to generate synthetic data from their own data.
- have a number of tools available to evaluate the analytical validity of the synthetic datasets.
- know how to assess the disclosure risk of the generated data

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. A Brief History of Data Confidentiality & Traditional Approaches for Data Protection
2. The Computer Science Perspective on Data Privacy & Introduction to Multiply Imputed Synthetic Datasets
3. Analyzing Synthetic Datasets & Relationship to Multiple Imputation for Nonresponse
4. Synthesis Models Part I (Univariate and Linear Regression Models)
5. Synthesis Models Part II (Models for Categorical Variables and Nonparametric Models) & Modeling Strategies
6. Analytical Validity & Disclosure Risk Part I (Theory)
7. Disclosure Risk Part II (Examples in R) & Discussion of the Chances and Obstacles of the Synthetic Data Approach
8. Discussion of the Third Homework Assignment

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 2 quizzes (worth 15% total)
- Participation in the weekly online meetings, engagement in discussions during the meetings and/or submission of questions via e-mail (worth 10%)
- Three homework assignments (worth 45% total)
- A final online exam (worth 30%)

Sampling II

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Raphael Nishimura, PhD (raphaeln@umd.edu)

Video lecture: Raphael Nishimura

Short Course Description

Sampling II is an applied statistics methods course, but differs from most statistics courses because it is concerned almost exclusively with the design of data collection. Little of the analysis of collected data will be discussed in the course. The course will concentrate on problems of applying sampling methods to human populations, since sampling human populations poses a number of particular problems not found in sampling of other types of units. The principles of sample selection, though, can be applied to many other types of populations.

Prerequisites

The course is presented at an intermediate statistical level. While we will not develop mathematical aspects of sampling theory, statistical notation and outlines of some algebraic proofs will be given. A sound background in applied statistics, proficiency in mathematics, including basic algebra, is necessary, since some algebraic derivations will be presented (although little emphasis will be placed on the derivations). A thorough understanding of the notation and algebraic results will be required. Sampling I or equivalent course is also required.

Course Objectives

By the end of the course, students will…

- be able to design and implement large-scale state-wide or national samples using multi-stage sampling with different sampling techniques
- understand issues related with multiple frame designs and use them in settings such as telephone sampling
- be able to identify opportunities to effectively use unequal selection probability design, like establishment surveys

Course Composition

This is a 2 ECTS course that runs for 4 weeks. The content of the course is broken down in 4 units:

1. Area Probability Sampling
2. Multiple Frame Sampling
3. Unequal Probability
4. Sampling for Panels and Longitudinal Studies

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- Homework assignments (worth 50% total)
- Quizzes (worth 15% total)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail demonstrating understanding of the required readings and video lectures, and positive contributions on Piazza, see below (worth 10%)
- A final open-book online exam (worth 25%)

Multiple Imputation – Why and How

Core Course
Focus Area: Data Curation / Storage

2 ECTS

Instructor:
Prof. Jörg Drechsler (joerg.drechsler@iab.de)

Video lecture: Prof. Jörg Drechsler

Short Course Description

This course will provide a detailed introduction to multiple imputation, a convenient strategy for dealing with (item) nonresponse in surveys. We will motivate the concept and illustrate why multiple imputation should generally be preferred over single imputation methods. The main focus of the course will be on strategies to generate (multiple) imputations and how to deal with common problems when applying the methods for large scale surveys. We will also discuss various options for assessing the quality of the imputations. All concepts will be demonstrated using software illustrations in R.

Prerequisites

Students should be familiar with generalized linear models and basic probability theory. We also expect that students know the basic concepts for dealing with nonresponse in surveys (the difference between item and unit nonresponse, formalizing the missing data mechanism, deterministic and stochastic approaches for imputation). For students unfamiliar with these concepts we highly recommend to enroll in the course “Nonresponse and Imputation” before participating in this course.

Some background knowledge in Bayesian statistics and Markov Chain Monte Carlo Methods (MCMC) is helpful but not mandatory. 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.

Course Objectives

By the end of the course, students will…

- understand why multiple imputation should be preferred over single imputation methods in most situations
- know about the two main approaches for multiple imputation
- be familiar with various imputation routines for different types of variables
- know how to implement these routines using R
- be able to deal with various problems that typically arise when imputing large scale surveys
- know about various strategies to assess the quality of the generated imputations

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. MI Intro & MI Analysis
2. MI for Continuous Variables
3. MI for Categorical Variables and Nonparametric Alternatives
4. Modeling Strategies and Quality Evaluations

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

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

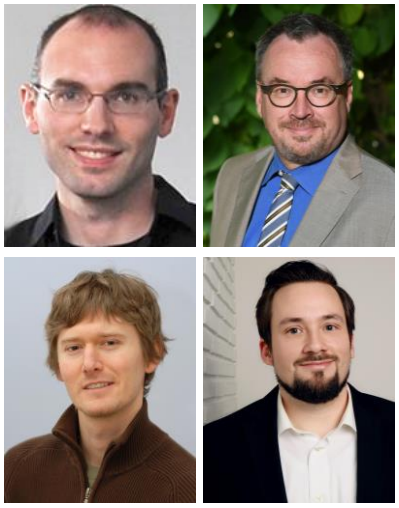
Year 2: Summer Term (June 2021 – August 2021)

Introduction to Record Linkage with Big Data Applications

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructors:
Manfred Antoni, PhD (Manfred.Antoni@iab.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)
Christian Borgs, PhD (christian.borgs@uni-due.de)
Prof. Joseph W. Sakshaug (joe.sakshaug@iab.de)



Video lecture:
Manfred Antoni, PhD, Prof. Stefan Bender, PhD, Christian Borgs, PhD, Prof. Joseph W. Sakshaug, PhD

Short Course Description

The course will address methods to combine data on given entities (people, households, firms etc.) that are stored in different data sources. By showing the strengths of these methods and by showing how each of them are performed in practice using R, the course will demonstrate the various benefits of record linkage. Participants will also learn about potential challenges that record linkage projects may face.

Prerequisites

Students should have knowledge of basic statistical concepts. They need to have an intermediate knowledge of R. Familiarity with regular expressions, the R packages ggplot2 and tidyverse is useful but not required.

Course Objectives

By the end of the course, students will…

- be familiar with a host of record linkage applications from different countries or jurisdictions that link a variety of data sources and use different types of linkage
- know how to improve the quality of linkage identifiers by applying pre-processing routines
- be familiar with different methods of increasing the efficiency of record linkage
- be able to understand, select and apply appropriate record linkage methods (e.g., deterministic and probabilistic linkage)
- be able to evaluate the success of data linkage
- be able to perform each step in the record linkage process using the R software

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introducing record linkage in the age of Big Data
2. Collecting and pre-processing linkage identifiers & blocking techniques
3. Data preprocessing and core concepts of data quality for linking
4. Comparison and classification of record pairs
5. Probabilistic record linkage and blocking (application)
6. Advanced topics, software options and literature review
7. Privacy-preserving record linkage using R
8. Evaluation and visualization of linkage quality

Learning and Teaching Methods

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. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week in the forum on the course page before the meetings.

Grading

Grading will be based on:

- 3 online quizzes (worth 30% total)
- Participation in the weekly online meetings (worth 20%): engagement in discussions during the meetings and submission of questions in the forum on the course website
- 3 homework assignments (worth 50% total)

Measurement Error Models

Core Course
Focus Area: Data Analysis

2 ECTS

Instructors:
Laura Boeschoten, PhD (lauraboeschoten@gmail.com)

Video lecture: Daniel Oberski

Short Course Description

Surveys reflect the opinions or facts researchers are after only partly – the other part will be measurement error, which can seriously bias analyses of interest. To remove such biases it is essential to estimate the extent of measurement error in survey variables, which is precisely the goal of *statistical measurement error modeling*. In this course, we will discuss how measurement error can be defined, how its presence can be detected using specialized data collection designs and models, and how to perform error-corrected statistical analyses of substantive interest.

Prerequisites

- Knowledge of basic statistics including regression analysis;
- Ability to run an R script, for example from RStudio; a cursory understanding of R;
- In-depth knowledge of R or latent variable models is NOT required.

Course Objectives

By the end of the course, students will…

- define measurement error conceptually, including the concepts of reliability and validity;
- explain the different approaches to estimating measurement error and their respective advantages and drawbacks;
- interpret the results of statistical models used to estimate measurement error in the absence of a gold standard;
- perform regression analyses from which the influence of measurement error has been removed and interpret the results.

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. What is measurement error?
2. Estimating measurement error in continuous survey variables
3. Estimating measurement error in categorical survey variables
4. Correcting regression analyses for the effects of measurement error

Learning and Teaching Methods

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. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week in the weekly forum on the course page before the meetings. 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.

Grading

Grading will be based on:

- 3 online quizzes (worth 30% total)
- Participation in discussion during the weekly online meetings and submission of questions to the weekly forum demonstrating understanding of the required readings and video lectures (worth 10%)
- A final open-book online exam (worth 60%)

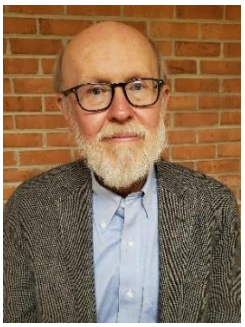
Year 2: Fall Term (September 2021 – December 2021)

Practical Inference from Complex Surveys

Core Course
Focus Area: Data Analysis

4 ECTS

Instructor:
Robert Fay, PhD (bobfay@hotmail.dom)



Video lecture: Richard Valliant

Short Course Description

We will discuss the theoretical and empirical properties of the two basic variance estimation strategies, namely Taylor series (linear) approximation and replication methods (including BRR, jackknife, and bootstrap) as they apply to several types of complex sample designs. We will study both descriptive statistics, such as means, and analytic statistics, such as linear and logistic regression. We will contrast model-based and design-based inference, the latter used as the standard in this course. Students will learn to use at least one survey software package with real survey data.

Prerequisites

A sound understanding of linear regression models (OLS), knowledge in linear algebra and calculus is important, as is previous exposure to complex sample designs and common estimation procedures. Previous exposure to maximum likelihood estimation is assumed, but students may meet this requirement by taking the course “Generalized Linear Models” in the online program previously or concurrently.

Course Objectives

By the end of the course, students will…

- understand the important theoretical properties underlying some widely used variance estimates
- calculate linear and replicate variance estimates for a variety of estimators such as a ratio mean
- determine the appropriate variance estimator(s) for a given sample design and inferential statistic
- calculate inferential statistics for descriptive estimators and analytic models using real-world survey data
- gain hands-on experience with one or more widely-used software packages that account for complex survey designs

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Variance Estimation for Complex Surveys
2. Taylor Series Linearization
3. Taylor Series Linearization (cont.)
4. Balanced Repeated Replication
5. The Jackknife and the Bootstrap and Comparisons among Methods
6. Linear Regression
7. Logistic Regression
8. Other Topics

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 8 short pre-class assignments and 7 longer post-class homework assignments (worth 60% total)
- Participation in online meetings and submission of questions demonstrating understanding of readings (worth 10%)
- Final Exam (worth 30%)

Practical Tools for Survey Weighting (Part II)

Core Course
Focus Area: Data Analysis

2 ECTS

Instructor:
Stefan Zins, PhD (st.zins@gmail.com)

Video lecture: Richard Valliant

Short Course Description

This course is a statistical methods class combining hands-on applications and general review of the theory for survey weighting.

Prerequisites

Sampling theory, Applied sampling (e.g., SURV626), or Practical Tools (Part I) for Sampling.
Some experience with variance estimation, statistical analysis using survey data, and the R statistical computing software will be helpful.

Course Objectives

By the end of the course, students will understand the…

- role of survey weights in population inference.
- steps in weighting, including computation of base weights, nonresponse adjustments, and uses of auxiliary data.
- nonresponse adjustment alternatives, including weighting cell adjustments, formation of cells using classification algorithms, and propensity score adjustments.
- weighting via poststratification, raking, general regression estimation, and other types of calibration.
- assessing if weights are not needed

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Basic Steps in Weighting
2. Basic Steps in Weighting (continued)
3. Calibration and Other Uses of Auxiliary Data in Weighting
4. Calibration (continued) and Replicate Weights

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 4 Homework assignments (worth 60% total)
- A take-home final exam (worth 30%)
- Class Participation (worth 10%) in discussion during the weekly online meetings and posting questions to the weekly forum demonstrating understanding of the required readings and video lectures

Year 3: Spring Term (January 2022 – May 2022)

Introduction to Data Visualization

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Prof. Richard Traunmüller (rtraunmu@mail.uni-mannheim.de)

Video lecture: Prof. Richard Traunmüller

Short Course Description

Data visualization is one of the most powerful tools to explore, understand and communicate patterns in quantitative information. At the same time, good data visualization is a surprisingly difficult task and demands three quite different skills: substantive knowledge, statistical skill, and artistic sense. The course is intended to introduce participants to key principles of analytic design and useful visualization techniques for the exploration and presentation of univariate and multivariate data. This course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. Students will learn how to evaluate data visualizations based on principles of analytic design, how to construct compelling visualizations using the free statistics software R, and how to explore and present their data with visual methods.

Prerequisites

Course prerequisites are a basic understanding of statistics and bivariate linear regression. Some experience in the use of a statistical software package would help, preferable basic data management tasks in R (loading and merging data, generating and recoding variables, etc.) I will provide detailed code examples to get students up to pace.

Course Objectives

By the end of the course, students will…

- know how to evaluate and criticize data visualizations based on principles of analytic design
- be in the position to explore and present their data with visual methods
- understand which graphical formats are useful for which types of data and questions
- know how to construct compelling visualizations using the free statistics software R

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Data Visualization as a Methodology
2. Principles of Analytic Design

3. Making Visual Comparisons
4. Visualizing Relations Between Variables

Learning and Teaching Methods

In this course, you are responsible for 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. 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. Just like in an on-site course, homework will be assigned and there will be a final project report at the end of the course.

Grading

Grading will be based on:

- Participation at online meetings (worth 10%)
- Required Course Assignments (worth 40% total)
- Final Visualization Project (worth 50%)

Project Consulting Course

Core Course
Focus Area: all areas except Research Design

12 ECTS

Instructors:
Prof. Helmut Küchenhoff (kuechenhoff@stat.uni-muenchen.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)

Short Course Description

In this course, students will apply the core skills that they learned in the MDM program to address real-world problems. The course will provide experience with the steps involved in carrying out a data consulting project, such as discussing the problems to solve with a client, data handling, and communicating work in both written and oral forms. The project is completed in teams (3-4 students per team).

Prerequisites

- Successful completion of the course “Fundamentals of Survey and Data Science”
- Experience with descriptive statistics, inferential statistics and linear modeling
- Familiarity with R or Python (you should be able to clean and manipulate data using one of these programing languages)

Course Objectives

By the end of the course, students will…

- be able to apply skills and concepts learned throughout the program to solve a real-world problem for a given client,
- be able to successfully manage a project in a timely manner,
- be able to collaborate with team members as well as data clients,
- be able to effectively communicate results in both written and oral forms.

Course Composition

This is a 12 ECTS course, which runs for 14 weeks. The content of the course is broken down into 14 units:

1. General introduction
2. Cooperation with clients / Assigning projects
3. Meeting with clients
4. Teams work on projects
5. Teams work on projects
6. Teams work on projects
7. Presentations of project progress / Feedback
8. Presentations of project progress / Feedback

9. Communicating results: effective presentation and report
10. Teams work on projects
11. Teams work on projects
12. Final presentations (part 1)
13. Final presentations (part 2)
14. Teams work on projects

Learning and Teaching Methods

This is a flipped online course in which students will form groups to collaboratively work on problems involving data presented to them by outside consultants. Students will work in teams to scope a problem, analyze data provided to them, distribute work among team members, and combine their results for a joint report and presentation.
The first part of the course emphasizes individual work. Students are expected to watch video-recorded lectures, read the required literature for each unit, and attend online discussion meetings with the instructor. In some meetings, students will meet with clients. The second part of the course will emphasize project-based work involving data analysis. Student will collaborate with their group members via online tools (e.g., Zoom meetings) to complete their projects. They will provide periodic progress updates about their group projects and present their work orally and in writing. Just like in an on-site course, homework will be assigned and graded and there will be a final report at the end of the course.

Grading

Grading will be based on:

- participation in online meetings (worth 10%)
- first meeting with clients and results protocol (worth 20%)
- final presentation (worth 35%)
- final report (worth 35%)

IPSDS

SPONSORED BY THE
Federal Ministry
of Education
and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

COURSE CATALOG 2020

MANNHEIM Master of Applied Data Science & Measurement

Track: Survey Methodology

STUDY YEAR 1-3 (JUNE 2020 – MAY 2022)

Overview

Track "Survey Methodology"	Instructor(s)	ECTS	Focus Area
Year 1 – Summer Term (June 2020 – August 2020)			
Connect@MDM or Elective[1]		2	Any Area
Review of Statistical Concepts	Brian Kim, PhD	6	Data Analysis
ECTS Points:		**8**	
Year 1 – Fall Term (September 2020 – December 2020)			
Fundamentals of Survey and Data Science	Alexander Wenz, PhD	6	Research Design
Privacy Law	Prof. Thomas Fetzer	2	Data Output/Access
Sampling I	Raphael Nishimura, PhD	4	Data Generation Process
ECTS Points:		**12**	
Year 2 – Spring Term (January 2021 – May 2021)			
Questionnaire Design	Gina Waleiko, PhD	4	Data Generating Process
Web Survey Methodology and Online Panels with practical survey programming	Jernej Berzelak, PhD	4	Data Generating Process
ECTS Points:		**8**	
Year 2 – Summer Term (June 2021 – August 2021)			
Introduction to Real World Data Management	Alexandru Cernat, PhD	4	Data Curation/Storage
Introduction to Record Linkage with Big Data Applications	Manfred Antoni, PhD, Prof. Stefan Bender, Christian Borgs, PhD, Prof. Joseph W. Sakshaug	4	Data Curation/Storage
ECTS Points:		**8**	
Year 2 – Fall Term (September 2021 – December 2021)			
Analysis of Complex Survey Data	Benjamin M. Reist, PhD	4	Data Analysis
Ethical Considerations for Data Science Research	Jessica Vitak, PhD	2	Research Design, Data Output/Access
Experimental Design for Surveys	Ashley Amaya, PhD	4	Data Generating Process
ECTS Points:		**10**	
Year 3 – Spring Term (January 2022 – May 2022)			
Introduction to Data Visualization	Prof. Richard Traunmüller	2	Data Output/Access
Project Consulting Course	Prof. Helmut Küchenhoff	12	all areas except Research Design
ECTS Points:		**14**	
Overall ECTS Points Year 1-3:		**60**	
Year 3 – Summer/Fall Term (June 2022 - December 2022)			
Master-Project[2]		15	
Overall ECTS-Punkte of the whole Tracks:		**75**	

[1] Please find more information under “Elective Courses & Connect”.

[2] Detailed information regarding the Master-Project can be found in the Program Guidelines.

Course Short Descriptions

Year 1: Summer Term (June 2020 – August 2020)

Review of Statistical Concepts *Core Course* *Focus Area: Data Analysis* *6 ECTS* Instructor: Brian Kim, PhD (kimbrian@umd.edu)	
Video lecture: Brian Kim, PhD	

Short Course Description

This course provides a brief overview of the basics of probability and statistics. Students will review basic probability concepts and probability distributions, the Central Limit Theorem and hypothesis testing, and linear and logistic regression. Throughout this course, students should develop and reinforce proper statistical intuition. This includes knowing how to identify a sample and a population and applying appropriate statistical methods such as hypothesis testing, as well being able to identify different types of data and using the proper methods for each type of data. By the end of the course, students should have a strong foundation in statistics with which they can start their graduate coursework.

Prerequisites

No Prerequisites.

Course Objectives

By the end of the course, students will…

- understand sample and population and know how to apply statistical methods appropriately
- be able to apply basic probability
- know basic probability distributions and how to apply them
- perform hypothesis tests and construct confidence intervals
- understand regression analysis, including multiple regression and logistic regression.

Course Composition

This is a 6 ECTS course that runs for 12 weeks. The content of the course is broken down into 11 units:

1. Introduction
2. Descriptive Statistics
3. Probability
4. The Normal Distribution and Z-Scores

5. Other Probability Distributions
6. Confidence Intervals
7. Central Limit Theorem and Hypothesis Testing
8. Inference for Numerical Data
9. Inference for Categorial Data
10. Linear Regression
11. Regression Assumptions, Multiple- and Logistic Regression

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- 11 homework assignments (worth 60% total)
- Participation in online meetings and submission of questions demonstrating understanding of readings (worth 10%)
- Online Final Exam (worth 30%)

Year 1: Fall Term (September 2020 – December 2020)

Fundamentals of Survey and Data Science

Core Course
Focus Area: Research Design

6 ECTS

Instructor:
Alexander Wenz, PhD (a.wenz@uni-mannheim.de)

Video lecture: Florian Keusch, Jennifer Sinibaldi

Short Course Description

This 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 students 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.

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.

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.

Course Composition

This is a 6 ECTS course that runs for 12 weeks. The content of the course is broken down into 12 units:

1. Introduction – How to do survey research and data science

2. Quality of Data
3. Coverage
4. Modes of Survey Data Collection
5. Data Generation from Other Sources
6. Sampling I
7. Sampling II
8. Questionnaires and Interviewing
9. Nonresponse
10. Data Preparation, Data Processing, and Data Base Management
11. Data Analysis and Data Visualization
12. Survey and Research Ethics

Learning and Teaching Methods

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. 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. Just like in an on-site course, assignments will be assigned and graded, and there will be a final exam at the end of the course.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and contributions to the forum demonstrating understanding of the required readings and video lectures (10% grade).
- Ten online quizzes reviewing specific aspects of the material covered (60% grade).
- A final, open-book, online exam (30% of grade).

Privacy Law

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Prof. Thomas Fetzer (fetzer@jura.uni-mannheim.de)

Video lecture: Prof. Thomas Fetzer

Short Course Description

The course will acquaint the students with the origins and basic principles of privacy law mainly in Europe. Furthermore, it will contrast the European privacy foundations with the U.S. approach. At the core of this course stands the new European General Data Protection Regulation (GDPR) and its applicability to specific cases and basic principles. Moreover, the course will cover current challenges to the existing privacy paradigms by big data and big data analytics.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- have a basic knowledge on the foundations of privacy law in Europe and the U.S.;
- have an understanding why privacy issues are treated differently in Europe and the U.S.;
- have a basic knowledge on the applicability of the General Data Protection Regulation (GDPR) and its basic principles;
- be aware of privacy issues and potential legal limitations when processing data;
- be aware of current challenges to the existing privacy paradigms by big data and big data analytics;
- be aware of currently discussed new approaches to privacy (e.g. privacy by design).

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction to privacy law
2. The GDPR I
3. The GDPR II
4. The GDPR III and current challenges to the existing privacy paradigms

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and submission of questions via e-mail demonstrating understanding of the required readings and video lectures (worth 10%)
- Three online quizzes reviewing specific aspects of the material covered (worth 60% total)
- A final open-book online exam (worth 30%)

Sampling I

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Raphael Nishimura, PhD (raphaeln@umd.edu)

Video lecture: Raphael Nishimura, PhD

Short Course Description

Sampling is an applied statistics methods course, but differs from most statistics courses because it is concerned almost exclusively with the *design* of data collection. Little of the analysis of collected data will be discussed in the course. The course will concentrate on problems of applying sampling methods to human populations, since sampling human populations poses a number of particular problems not found in sampling of other types of units. The principles of sample selection, though, can be applied to many other types of populations.

Prerequisites

The course is presented at an intermediate statistical level. While we will not develop mathematical aspects of sampling theory, statistical notation and outlines of some algebraic proofs will be given. A sound background in applied statistics, proficiency in mathematics, including basic algebra, is necessary, since some algebraic derivations will be presented (although little emphasis will be placed on the derivations). A thorough understanding of the notation and algebraic results will be required.

Course Objectives

By the end of the course, students will…

- understand the basic ideas, concepts and principles of probability sampling from an applied perspective
- be able to identify and appropriately apply sampling techniques to survey design problems
- be able to compute the sample size for a variety of sample designs
- understand and be able to assess the impact of the sample design on survey estimates
- be able to estimate the precision of the survey statistics using different estimation techniques

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introduction; Course Perspectives
2. Simple Random Sampling, Sampling Frames, and Introduction to Clustering

3. Stratified Sampling I
4. Stratified Sampling II, Systematic selection
5. Cluster Sampling
6. Unequal-Sized Clusters I
7. Unequal-Sized Clusters II
8. Variance Estimation

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- Homework assignments (worth 50% total)
- Quizzes (worth 15% total)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail demonstrating understanding of the required readings and video lectures, and positive contributions on Piazza, see below (worth 10%)
- A final open-book online exam (worth 25%)

Year 2: Spring Term (January 2021 – May 2021)

Questionnaire Design

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Gina Walejko, PhD, (gkwalejko@gmail.com)

Video lecture: Prof. Frauke Kreuter

Short Course Description

This course introduces students to the stages of questionnaire development. The course reviews the scientific literature on questionnaire construction, the experimental literature on question effects, and the psychological literature on information processing. It will also discuss the diverse challenges posed by self- versus proxy reporting, and special attention is paid to the relationship between mode of administration and questionnaire design. Students will also get a hands-on experience in developing their own questionnaire.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- be able to develop their own questionnaire based on a research question.
- be able to apply the knowledge about the cognitive response process to write good survey questions.
- be able to select and apply tools to pretest their questionnaire.
- know different techniques to ask respondents about sensitive topics.
- be able to develop questions that ask about facts (i.e. behaviors and events) and non-fact (i.e. attitudes and opinions).
- be able to put individual survey questions in an appropriate sequence considering the idiosyncrasies of different modes of data collection.

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introduction; Course Perspectives
2. Simple Random Sampling, Sampling Frames, and Introduction to Clustering
3. Stratified Sampling I
4. Stratified Sampling II, Systematic selection

5. Cluster Sampling
6. Unequal-Sized Clusters I
7. Unequal-Sized Clusters II
8. Variance Estimation

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- Homework assignments (worth 50% total)
- Quizzes (worth 15% total)
- Participation in discussion during the weekly online meetings, submission of questions via e-mail demonstrating understanding of the required readings and video lectures, and positive contributions on Piazza, see below (worth 10%)
- A final open-book online exam (worth 25%)

Web Survey Methodology and Online Panels with practical survey programming

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Jernej Berzelak, PhD, (Nejc.Berzelak@fdv.uni-lj.si)

Video lecture: Mario Callegaro, PhD

Short Course Description

The course introduces the students to the fundamental concepts of web surveys and online panels. The course is organized in three main sections which follow the way a proper web survey is organized: pre-fielding, fielding, and post fielding.

Prerequisites

Course 'Fundamentals of Surveys and Data Science' or knowledge in Data Collection Methods
Permission from instructor due to demonstrated knowledge of total survey error

Course Objectives

By the end of the course, students will…

- know if a web survey is appropriate for the target population of a research project
- be able to manage the entire process of collecting data via a web survey
- have the mental tools to design a questionnaire to be fielded online
- have a better understanding on online panels and how to use them
- know where to find answers to questions related to web surveys and online data collection
- learn what aspects of web surveys still deserve more research
- learn how to program a web survey

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Web surveys, mode and sampling
2. Online panels and convenience samples of internet users
3. Questionnaires for web surveys and paradata
4. Nonresponse and web surveys management
5. Postfielding & device effects
6. Data quality of online panels
7. Web survey programming exercise
8. Web survey programming exercise

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 6 homework assignments (worth 60% total)
- Participation in the weekly online meetings (worth 10%): engagement in discussions during the meetings and submission of questions via forum/e-mail
- Final Exam web survey programming (worth 30%)

Year 2: Summer Term (June 2021 – August 2021)

Introduction to Real World Data Management

Core Course
Focus Area: Data Curation / Storage

4 ECTS

Instructor:
Alexandru Cernat, PhD (contact@alexcernat.com)

Video lecture: Alexandru Cernat, PhD

Short Course Description

Data is omnipresent in the contemporary world coming in different shapes and sized: from survey data to found data. In order to make use of such data through analysis it is necessary first to import and clean it. This is often one of the most time consuming and difficult parts of data analysis. In this course you will learn both the conceptual steps needed in preparing data for analysis as well as the practical skills to do this. The course will cover all the essential skills needed to prepare data be it survey data, administrative data or found data.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- understand the stages involved in preparing data for analysis
- understand the concept of tidy data
- understand the basics of using R
- know how to write their own functions and loop over them
- know how to import and export data
- know how to clean data in R
- know how to merge data
- know how to manipulate textual data
- know how to manipulate date/time data
- know how to use tables and graphs to explore data

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. The basics of R
2. Practical 1
3. Data importing and cleaning
4. Practical 2

5. Exploring Data
6. Practical 3
7. Learning how to deal with text and date information
8. Practical 4

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

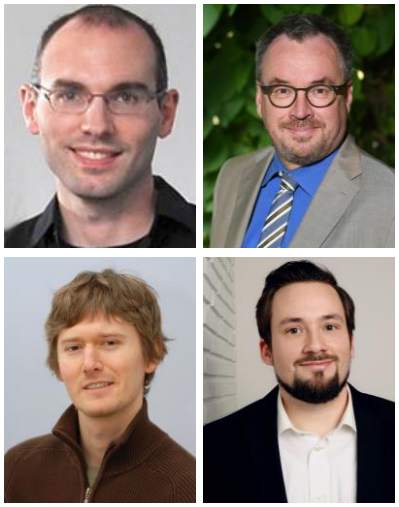
- 4 fortnightly homework assignments (worth 60% total)
- Participation in discussion during the weekly online meetings and submission of questions to the forum (demonstrating understanding of the required readings and video lectures (worth 10%)
- A final project (worth 30%)

Introduction to Record Linkage with Big Data Applications

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructors:
Manfred Antoni, PhD (manfred.antoni@iab.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)
Christian Borgs, PhD (christian.borgs@uni-duisburg-essen.de)
Prof. Joseph W. Sakshaug (joe.sakshaug@iab.de)



Video lecture:
Manfred Antoni, PhD, Prof. Stefan Bender, PhD, Christian Borgs, PhD, Prof. Joseph W. Sakshaug, PhD

Short Course Description

The course will address methods to combine data on given entities (people, households, firms etc.) that are stored in different data sources. By showing the strengths of these methods and by showing how each of them are performed in practice using R, the course will demonstrate the various benefits of record linkage. Participants will also learn about potential challenges that record linkage projects may face.

Prerequisites

Students should have knowledge of basic statistical concepts. They need to have an intermediate knowledge of R. Familiarity with regular expressions, the R packages ggplot2 and tidyverse is useful but not required.

Course Objectives

By the end of the course, students will…

- be familiar with a host of record linkage applications from different countries or jurisdictions that link a variety of data sources and use different types of linkage
- know how to improve the quality of linkage identifiers by applying pre-processing routines
- be familiar with different methods of increasing the efficiency of record linkage
- be able to understand, select and apply appropriate record linkage methods (e.g., deterministic and probabilistic linkage)
- be able to evaluate the success of data linkage
- be able to perform each step in the record linkage process using the R software

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introducing record linkage in the age of Big Data
2. Collecting and pre-processing linkage identifiers & blocking techniques
3. Data preprocessing and core concepts of data quality for linking
4. Comparison and classification of record pairs
5. Probabilistic record linkage and blocking (application)
6. Advanced topics, software options and literature review
7. Privacy-preserving record linkage using R
8. Evaluation and visualization of linkage quality

Learning and Teaching Methods

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. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week in the forum on the course page before the meetings.

Grading

Grading will be based on:

- 3 online quizzes (worth 30% total)
- Participation in the weekly online meetings (worth 20%): engagement in discussions during the meetings and submission of questions in the forum on the course website
- 3 homework assignments (worth 50% total)

Year 2: Fall Term (September 2021 – December 2021)

Analysis of Complex Survey Data

Core Course
Focus Area: Data Analysis

4 ECTS

Instructor:
Benjamin M. Reist (breist@umd.edu)

Video lecture: Steven G. Heeringa

Short Course Description

Analysis of Complex Sample Data covers the following topics: the development and handling of selection and other compensatory weights for survey data analysis; the effects of stratification and clustering on survey estimation and inference; alternative variance estimation procedures for estimated survey statistics; methods and computer software that take into account the effects of complex sample designs on survey estimation and inference; and methods for handling missing data, including weighting adjustment.

Prerequisites

The prerequisites include one or more graduate courses in statistics covering techniques through OLS and logistic regression, a course in applied sampling methods, or permission of the instructor. The course is presented at a moderately advanced statistical level. Although the course will review the fundamentals of statistical analysis methods for survey data and provide detailed examples on the use of statistical software, it will be assumed that the students are familiar with statistical methods, including multiple regression and logistic regression. The initial lectures in the course syllabus will review the various complex features of sample designs and how they influence 2 estimation and inference based on survey data. The course syllabus and level of instruction also assume that students are familiar with basic sampling procedures, including simple random sampling, stratification, cluster sampling and multi-stage sample designs. Students who do not have graduate-level training in sampling techniques should expect to devote additional time during the first weeks of the course to supplemental readings on this topic.

Course Objectives

By the end of the course, students will…

- understand the importance of accounting for the effects of complex sample designs on estimation and inference.
- be able to identify how sample design elements impact estimation and inference
- be able to estimate sampling error using:
- direct estimators
- linearization techniques

- replication methods
- be able to account for complex sample designs in:
- descriptive analysis for continuous variables
- categorical data analysis
- linear regression
- logistic regression
- be able to use standard statistical software to account for the effects of complex sample designs.

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Survey estimation and inference for complex sample designs (Part 1)
2. Survey estimation and inference for complex sample designs (Part 2)
3. Sampling error estimation for complex samples
4. Descriptive analysis for continuous variables (Part 1)
5. Descriptive analysis for continuous variables (Part 2)
6. Analysis of categorical data from complex samples
7. Linear regression for complex sample survey data
8. Logistic regression for complex survey data

Learning and Teaching Methods

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. 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. Just like in an on-site course, homework will be assigned and graded and there will be a final project at the end of the course.

Grading

Grading will be based on three criteria:

- Weekly questions submitted / class preparation / class participation (worth 20%)
- Completion of four (4) homework assignments (worth 40% total)
- Final course project (worth 40%)

Ethical Considerations for Data Science Research

Core Course
Focus Area: Research Design, Data Output / Access

2 ECTS

Instructor:
Jessica Vitak, PhD

Video lecture: Jessica Vitak, PhD

Short Course Description

The goal of research ethics is to protect human subjects from harm when they participate in a study. In the digital age, however, what constitutes "participation" has become blurry, especially with the rise of social media platforms and other online apps and services. Furthermore, new applications of big data raise important questions about how to protect consumers from harms, and what kinds of notice and consent should be obtained. This course provides and introduction and overview of research ethics in the 21st century and evaluates the many challenges to conducting ethical research.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will…

- Describe the history of research ethics and the goals of institutional review boards
- Describe the challenges data science and big data raise for protecting individuals' rights and privacy
- Identify ethical issues in the study design, data collection, and data analysis process
- Detail best practices for conducting ethical research

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Social and Ethical Implications of Data
2. Foundations of Research Ethics
3. Ethical Concerns Around Data Collection, Storage, and Analysis
4. Best Practices for Conducting Ethical Research

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on three criteria:

- Participation in discussion during the weekly online meetings and contributions to weekly discussion forums demonstrating understanding of the required readings and video lectures (10% of grade)
- Four open-book quizzes assessing comprehension of course material (20% of grade; 5% each)
- Three online homework assignments reviewing specific aspects of the material covered (45% of grade; 15% each)
- Final paper covering overarching themes of the class (25%).

Experimental Design for Surveys

Core Course
Focus Area: Data Generating Process

4 ECTS

Instructor:
Ashley Amaya, PhD (aamaya@rti.org)

Video lecture: Roger Tourangeau, PhD

Short Course Description

This course examines how to embed experiments in surveys. It covers both the design of survey experiments and the analysis of the results.

Prerequisites

At least one prior course in data analysis. Ability to use SAS or STATA.

Course Objectives

By the end of the course, students will

- know about basic principles of experimental design
- recognize the main types of experimental designs
- improve the quality of designs used to carry out methodological research in or for surveys
- develop critical skills to spot flaws in experimental and nonexperimental designs to support causal inferences
- improve skills at analyzing results of survey experiments
- improve skills as both consumer and producer of experiments done to shed light on survey methodological issues

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Introduction
2. Examples of Experiments in Surveys
3. Experimental Designs I
4. Experimental Designs II
5. Comparability and Generalizability
6. Construct Validity I
7. Construct Validity II; Statistical Validity
8. Wrap-Up

Learning and Teaching Methods

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. 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. As with an on-site course, homework will be assigned and graded and there will be three quizzes.

Grading

Grading will be based on:

- 3 Online quizzes (worth 45% total)
- 3 exercises (worth 45% total)
- Participation in online discussions (worth 10%)

Year 3: Spring Term (January 2022 – May 2022)

Introduction to Data Visualization

Core Course
Focus Area: Data Output / Access

2 ECTS

Instructor:
Prof. Richard Traunmüller (rtraunmu@mail.uni-mannheim.de)

Video lecture: Prof. Richard Traunmüller

Short Course Description

Data visualization is one of the most powerful tools to explore, understand and communicate patterns in quantitative information. At the same time, good data visualization is a surprisingly difficult task and demands three quite different skills: substantive knowledge, statistical skill, and artistic sense. The course is intended to introduce participants to key principles of analytic design and useful visualization techniques for the exploration and presentation of univariate and multivariate data. This course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. Students will learn how to evaluate data visualizations based on principles of analytic design, how to construct compelling visualizations using the free statistics software R, and how to explore and present their data with visual methods.

Prerequisites

Course prerequisites are a basic understanding of statistics and bivariate linear regression. Some experience in the use of a statistical software package would help, preferable basic data management tasks in R (loading and merging data, generating and recoding variables, etc.) I will provide detailed code examples to get students up to pace.

Course Objectives

By the end of the course, students will…

- know how to evaluate and criticize data visualizations based on principles of analytic design
- be in the position to explore and present their data with visual methods
- understand which graphical formats are useful for which types of data and questions
- know how to construct compelling visualizations using the free statistics software R

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Data Visualization as a Methodology

2. Principles of Analytic Design
3. Making Visual Comparisons
4. Visualizing Relations Between Variables

Learning and Teaching Methods

In this course, you are responsible for 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. 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. Just like in an on-site course, homework will be assigned and there will be a final project report at the end of the course.

Grading

Grading will be based on:

- Participation at online meetings (worth 10%)
- Required Course Assignments (worth 40% total)
- Final Visualization Project (worth 50%)

Project Consulting Course

Core Course
Focus Area: all areas except Research Design

12 ECTS

Instructors:
Prof. Helmut Küchenhoff (kuechenhoff@stat.uni-muenchen.de)
Prof. Stefan Bender (stefan.bender@bundesbank.de)

Short Course Description

In this course, students will apply the core skills that they learned in the MDM program to address real-world problems. The course will provide experience with the steps involved in carrying out a data consulting project, such as discussing the problems to solve with a client, data handling, and communicating work in both written and oral forms. The project is completed in teams (3-4 students per team).

Prerequisites

- Successful completion of the course "Fundamentals of Survey and Data Science"
- Experience with descriptive statistics, inferential statistics and linear modeling
- Familiarity with R or Python (you should be able to clean and manipulate data using one of these programing languages)

Course Objectives

By the end of the course, students will…

- be able to apply skills and concepts learned throughout the program to solve a real-world problem for a given client,
- be able to successfully manage a project in a timely manner,
- be able to collaborate with team members as well as data clients,
- be able to effectively communicate results in both written and oral forms.

Course Composition

This is a 12 ECTS course, which runs for 14 weeks. The content of the course is broken down into 14 units:

1. General introduction
2. Cooperation with clients / Assigning projects
3. Meeting with clients
4. Teams work on projects
5. Teams work on projects
6. Teams work on projects
7. Presentations of project progress / Feedback
8. Presentations of project progress / Feedback

9. Communicating results: effective presentation and report
10. Teams work on projects
11. Teams work on projects
12. Final presentations (part 1)
13. Final presentations (part 2)
14. Teams work on projects

Learning and Teaching Methods

This is a flipped online course in which students will form groups to collaboratively work on problems involving data presented to them by outside consultants. Students will work in teams to scope a problem, analyze data provided to them, distribute work among team members, and combine their results for a joint report and presentation.

The first part of the course emphasizes individual work. Students are expected to watch video-recorded lectures, read the required literature for each unit, and attend online discussion meetings with the instructor. In some meetings, students will meet with clients. The second part of the course will emphasize project-based work involving data analysis. Student will collaborate with their group members via online tools (e.g., Zoom meetings) to complete their projects. They will provide periodic progress updates about their group projects and present their work orally and in writing. Just like in an on-site course, homework will be assigned and graded and there will be a final report at the end of the course.

Grading

Grading will be based on:

- participation in online meetings (worth 10%)
- first meeting with clients and results protocol (worth 20%)
- final presentation (worth 35%)
- final report (worth 35%)

IPSDS

SPONSORED BY THE
Federal Ministry
of Education
and Research

AUFSTIEG DURCH BILDUNG
OFFENE HOCHSCHULEN

COURSE CATALOG 2020

MANNHEIM Master of Applied Data Science & Measurement

Elective courses & Connect

STUDY YEAR 1-3 (JUNE 2020 – MAY 2022)

Elective Courses

In addition to the compulsory courses of your track you can also take elective courses. You can either choose courses from other tracks or courses that are not part of any track. Below you find a list of our electives that are not assigned to a track.
The terms/times given in the overview below are examples. You are free to choose when you want to participate in an elective.

Elective Courses (that are not part of a track)	Instructor(s)	ECTS	Focus Area
Year 2 – Spring Term (January 2021 – May 2021)*			
Usability Testing for Survey Research	Emily Geisen Jennifer Romano-Bergstrom, PhD	2	Data Generating Process
Web Scraping and APIs	Simon Munzert, PhD	4	Data Generating Process
Year 2 – Summer Term (June 2021 – August 2021)*			
Survey Design and Implementation in International Contexts	Zeina Mneimneh, PhD	2	Data Generating Process
Year 2 – Fall Term (September 2021 – December 2021)*			
Practical Tools for Sampling (Part I)	Stefan Zins, PhD	4	Data Analysis
Year 3 – tbc			
Introduction to Official Statistics	Walter J. Radermacher, PhD	2	Data Output/Access
Advanced Topics in Official Statistics	Hanna Brenzel Piet Daas Marco Puts	2	Data Output/Access

Year 2: Spring Term (January 2021 – May 2021)

Usability Testing for Survey Research

Elective Course
Focus Area: Data Generating Process

2 ECTS

Instructors:
Emily Geisen (egeisen@qualtrics.com)
Jennifer Romano-Bergstrom, PhD
(jennifer@romanocog.com)

Video lecture: Emily Geisen, Jennifer Romano-Bergstrom, PhD

Short Course Description

This course introduces the concepts of usability and usability testing and why they are needed for survey research. The course provides a theoretical model for understanding the respondent-survey interaction and then provides practical methods for incorporating iterative user-centered design and testing into the survey development process. The course provides techniques and examples for designing, planning, conducting and analyzing usability studies on web or mobile surveys.

Prerequisites

Students are expected to be familiar with questionnaire design. Experience with cognitive testing is a plus, but not a requirement.

Course Objectives

By the end of the course, students will

- understand what usability and usability testing are and how to apply usability testing to survey research
- learn about moderating techniques, such as the think-aloud protocol and verbal probing
- learn when to test, where to test (lab vs. field vs. remote) and who to test with (type and number of participants)
- be able to plan for usability testing (develop protocol guide, determine test metrics, consider hardware/software)
- learn what to test: conceptual testing, paper prototypes, wireframes
- understand how to collect, record, and analyze usability data

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction

2. Moderating Techniques
3. Test Materials, Metrics, Equipment, Location
4. What to test/ Analyzing Results

Learning and Teaching Methods

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. 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.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings (worth 10%)
- Weekly online exercises reviewing specific aspects of the material covered (worth 60% total)
- A final open-book online exam (worth 30%)

Web Scraping and APIs

Elective Course
Focus Area: Data Generating Process

2 ECTS

Instructor:
Simon Munzert, PhD (munzert@hertie-school.org)

Video lecture: Simon Munzert, PhD

Short Course Description

The short course provides a condensed overview of web technologies and techniques to collect data from the web in an automated way. To this end, students will use the statistical software R. The course introduces fundamental parts of web architecture and data transmission on the web. Furthermore, students will learn how to scrape content from static and dynamic web pages and connect to APIs from popular web services. Finally, practical and ethical issues of web data collection are discussed.

Prerequisites

Students are expected to be familiar with the statistical software R. Besides base R, knowledge about the "tidyverse" packages, in particular, dplyr, plyr, magrittr, and stringr, are of help. If you are familiar with R but have no experience in working with these packages, the best place to learn them is the primary reading "R for Data Science".

Course Objectives

By the end of the course, students will…

- have an overview of state-of-the-art research that draws on web-based data collection,
- have a basic knowledge of web technologies,
- be able to assess the feasibility of conducting scraping projects in diverse settings,
- be able to scrape information from static and dynamic websites as well as web APIs using R, and
- be able to tackle current research questions with original data in their own work.

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction – Web Technologies
2. Scraping static webpages
3. Scraping dynamic webpages and good practice
4. Tapping APIs

Learning and Teaching Methods

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. In addition, students are encouraged to post questions about the materials covered in the videos and readings of the week to the “Place to post your questions” forum before the meetings.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and submission of questions via the forum demonstrating understanding of the required readings and video lectures (worth 10%)
- Weekly quizzes that check factual knowledge about the course topics (worth 30% total)
- Weekly assignments that require students to implement and practice scraping techniques in R (worth 60% total)

Year 2: Summer Term (June 2021 – August 2021)

Survey Design and Implementation in International Contexts

Elective Course
Focus Area: Data Generating Process

2 ECTS

Instructor:
Zeina Mneimneh, PhD (zeinam@umich.edu)

Video lecture: Zeina Mneimneh, PhD

Short Course Description

This course focuses on design and implementation considerations for different phases of the survey lifecycle when conducting surveys internationally or outside of one's home country. Overview and considerations related to ten topics are discussed: Total Survey Error framework, project stakeholders, triple constraints, bids and contracts, sampling and sample management, questionnaire and instrument design, translation and adaptation, pretesting and cognitive interviews, interviewers and data collection, and interviewer monitoring.

Prerequisites

There are no specific prerequisites; however, some background in survey operations is helpful.

Course Objectives

By the end of the course, students will understand fundamental design and implementation considerations related to key stages of the survey lifecycle for conducting surveys in a different national context including:

- Overall survey design and the Total Survey Error framework
- Project management including stakeholders and the triple constraints of time, cost, and scope
- Bids and contracts
- Sampling and sample management
- Questionnaire and instrument design
- Translation and adaptation
- Pretesting and cognitive interviews
- Interviewers, data collection, and interviewer monitoring

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. The Total Survey Error framework, project stakeholders, and the triple constraints of time, cost and scope
2. Bids and contracts and sampling and sample management
3. Questionnaire and instrument design, translation and adaption, and pretesting and cognitive interviews
4. Interviewers, data collection and interviewer monitoring

Learning and Teaching Methods

In this course, you are responsible for watching video recorded lectures and reading the required literature or watching required videos 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. 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. Just like in an on-site course, homework will be assigned and graded and there will be quizzes planned within a specific time window.

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and submission of questions (worth 10%)
- 4 online quizzes (worth 20% total)
- 2 homework assignments (worth 70% total)

Year 2: Fall Term (September 2021 – December 2021)

Practical Tools for Sampling (Part I)

Elective Course
Focus Area: Data Analysis

4 ECTS

Instructor:
Stefan Zins, PhD (st.zins@gmail.com)

Video lecture: Richard Valliant

Short Course Description

This course is a statistical methods class combining hands-on applications and general review of the theory behind different approaches to survey sampling.

Prerequisites

Sampling theory (e.g., SURV440) and applied sampling (e.g., SURV626).
Some experience with the R statistical computing software is helpful.

Course Objectives

By the end of the course, students will be able to…

- sample size calculations using estimation targets based on coefficient of variation, margin of error, and power requirements.
- mathematical programming to determine sample sizes needed to achieve estimation goals for a series of subgroups and analysis variables.
- resources for designing area probability simples.
- methods of sample allocation for multistage samples.

Course Composition

This is a 4 ECTS course, which runs for 8 weeks. The content of the course is broken down into 8 units:

1. Overview / The R Statistical Package / Sample Design and Sample Size in Single-stage Surveys
2. Sample Design and Sample Size in Single-stage Surveys (continued)
3. Power Computations and Sample Size Determination
4. Mathematical Programming/ Study Performance rates and Effect on Sample Size
5. Designing Multistage Samples
6. Designing Multistage Samples (continued)
7. Area Probability Sampling
8. Area Probability Sampling (continued) / Wrap-up

Learning and Teaching Methods

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. 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. 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.

Grading

Grading will be based on:

- 8 Homeworks (worth 60% total)
- A take-home final exam (worth 20%)
- 6 Quizzes (worth 10% total)
- Class Participation (worth 10%) in discussion during the weekly online meetings and posting questions to the weekly forum demonstrating understanding of the required readings and video lectures

Year 3: tbc

Introduction to Official Statistics *Elective Course* *Focus Area: Data Output/Access* *2 ECTS* Instructor: Walter J. Radermacher, PhD (wjr@outlook.de)	
Video lecture: Walter J. Radermacher, PhD	

Short Course Description

Students will be familiarised with the DNA of official statistics, what quality means, how to achieve it, especially under the challenges of modern societies. In times of data revolution, artificial intelligence and a ubiquitous flood of permanently produced and consumed information, orientation and differentiation of good and less good information, of fact-checks, and of verification of statistical evidence is crucial. The public infrastructure called 'official statistics' plays an essential role in this respect.

Prerequisites

No prerequisites.

Course Objectives

By the end of the course, students will …

- know the production process of official statistics as well as the institutional set-up of European statistics
- know the types of products and services provided by official statistics, including their respective quality profiles
- know the quality management of official statistics and understand how (civil) society participates in the making of statistics
- understand the opportunities offered by new data sources as well as their risks; know the main principles for modern public statistics
- be able to apply the knowledge to the design of evidence for decision making in more complex situations

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. The Art of Statistics
2. Official statistics: The Making of
3. The Era of Digitalisation and Globalisation
4. Evidence and Decision Making: The Case of Sustainable Development

Learning and Teaching Methods

In this course, you are responsible for watching video-recorded lectures, reading the required literature for each unit and doing the quizzes (graded). We offer a discussion forum for each unit where students have the chance to discuss the materials from a unit with the fellow students and the tutor of the course.

Grading

Grading will be based on:

- 4 online quizzes (worth 100% total)

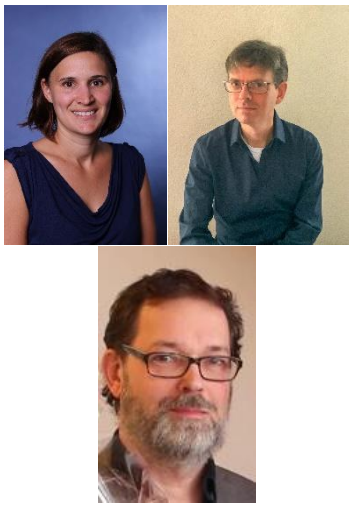
Year 3: tbc

Advanced Topics in Official Statistics

Elective Course
Focus Area: Data Output/Access

2 ECTS

Instructors:
Hanna Brenzel, (hb@hanna-brenzel.de)
Piet Daas, (pjh.daas@cbs.nl)
Marco Puts, (m.puts@cbs.nl)



Video lecture: Hanna Brenzel, Piet Daas, Marco Puts

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.

Prerequisites

No prerequisites.

Course 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

Course Composition

This is a 2 ECTS course, which runs for 4 weeks. The content of the course is broken down into 4 units:

1. Introduction to Big Data in official statistics

2. Using Big Data for official statistics
3. The importance and use of georeferenced data in official statistics
4. Introduction to Microsimulation in official statistics

Learning and Teaching Methods

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. 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.

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 demonstrating understanding of the required readings and video lectures (10% of grade)
- 2 online quizzes (worth 40% total and 20% each)
- 2 assignments (worth 50% total and 25% each)

Connect@MDM

You can also take part in Connect@MDM and get 2 ECTS under certain conditions. The annual CONNECT@MDM conference usually takes place in May/June in Mannheim. It is dedicated to facilitating networking and information exchange in the area of survey and data science across various sectors. You will meet your fellow students and instructors, learn from our panelists, participate in interesting workshops and listen to exciting lectures. If you attend both days and successfully complete the workshop (it includes some homework), you can receive 2 ECTS credits, and the course will count towards your ECTS total.