Applied Data Science

B.A. International Management - Fall Semester 2025/26

2025-09-09

1 Overall Goal of the Course

You will master practical data science skills using the open source programming language R. You will learn to collect, analyze, and visualize data to support managerial decision-making. Moreover, you will develop competencies in statistical modeling, machine learning, and reproducible reporting that are essential for evidence-based management practices.

2 Learning Goals

More concretely, this seminar aims to teach the following competencies:

Technical Expertise:

- Program effectively in R for data manipulation, analysis, and visualization
- Create professional reports and presentations using Quarto for reproducible research
- Apply statistical methods and machine learning techniques to business datasets
- Import, clean, and prepare real-world business data for analysis
- Develop predictive models for business forecasting and decision support

Methodological Competence:

- Design and conduct data-driven analyses to answer business questions
- Select appropriate statistical methods and visualization techniques for different types of business problems
- Interpret and communicate analytical results to non-technical stakeholders
- Evaluate the quality and limitations of data and analytical approaches

Social and Personal Skills:

- Work collaboratively on data science projects
- Present complex analytical findings clearly and transparently
- Critically evaluate data-driven claims and recommendations in business contexts

3 Course Information

3.1 Prerequisites

This seminar is designed for management students in the 3rd or 5th semester. Students must have completed basic statistics courses (Statistik I & II). Moreover, while not being required, it is helpful if you feel comfortable with the fundamental mathematics for business studies and come with basic computer literacy. Previous knowledge in programming is not required, except the content covered in *Statistik I* and *II*.

Finally, for this course it is important that you bring your own laptop on which you can install and run R.

3.2 Course Details

- **Credits:** 5 CP (150 total hours)
- **Contact Time:** 24.75 hours (11 × 2.25 hours)
- **Self-Study:** 125.25 hours
- **Location:** MAD, room 024
- **Format:** Interactive seminar combining input lectures, interactive elements, and regular hands-on work in class

3.3 Course Resources

- **Moodle Room:** 16255
- **Moodle Password:** `DataSci25`
- **Required Software:** R (free), RStudio (free), Quarto (free)
- **Recommended Textbooks:**
  - Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import, tidy, transform, visualize, and model data* (2nd edition). O'Reilly. https://r4ds.hadley.nz/
  - Ismay, C., & Kim, A. Y.-S. (2020). *Statistical inference via data science: A ModernDive, into R and the tidyverse*. CRC Press, Taylor and Francis Group. https://moderndive.com/index.html
  - Wickham, H. (2019). *Advanced R* (Second edition). CRC Press/Taylor & Francis Group. https://adv-r.hadley.nz/
  - James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (Second edition). Springer. https://www.statlearning.com/

3.4 Recommended Study Schedule

- **Time Horizon:** September 15 - December 15 (13 weeks)
- **Total Commitment:** 150 hours (11.5 hours per week average)
- **Self-Study:** 125.25 hours (9.6 hours per week average)

- **Exemplary Daily Schedule:**
  - **7 days/week:** 1.4 hours of self-study per day on average
  - **5 days/week:** 1.9 hours of self-study per day on average

4 Schedule and Logistics

> Schedule is tentative
>
> This schedule is subject to change as the course progresses. Announcements will be made via Moodle. The latest version of the syllabus will be made available here.

#	Week	Date	Topic
1	38	15.09.	Introduction & Installation
2	39	22.09.	R Programming & Data Management
3	40	29.09.	Quarto
4	41	06.10.	Statistical Recap & Inference
5	43	27.10.	Data Visualization
6	45	03.11.	AI & Machine Learning Basics
7	46	10.11.	Data Import & Preparation
8	47	17.11.	Sales Estimation & Modeling
9	48	24.11.	**Open Topic 1** - *Students Choose*
10	49	01.12.	**Open Topic 2** - *Students Choose*
11	50	08.12.	Review & Project Presentations

Available topics for Sessions 9 and 10 are described on a separate page.

4.1 Session Format

Each session includes:

- **Input Lecture**: Core concepts and theory
- **Interactive Elements**: Discussions and Q&A
- **Hands-on Practice**: Guided exercises in R/RStudio
- **Group Work**: Collaborative problem-solving, followed by a joint reflection in class

4.2 Assessment Schedule

- **Midterm Term Paper**: Pass/fail requirement
  - To be admitted to the final exam, students must pass the midterm term paper.
- **Final Exam**: 120-minute open book exam
  - Exam will be completed on students' own laptops

- Collaboration and paid tools not allowed
- All course materials and free resources permitted

4.3 Technical Requirements

- **Software**: R (4.3.0), RStudio (2023.06), Quarto (1.3)
- **Hardware**: Laptop capable of running statistical software (required for exam)
- **Data**: Real business datasets provided throughout course
- **Support**: Technical help sessions available by appointment

---

Contact Information

For questions regarding course content or logistics, please use the Moodle discussion forums or make an appointment during my office hours.