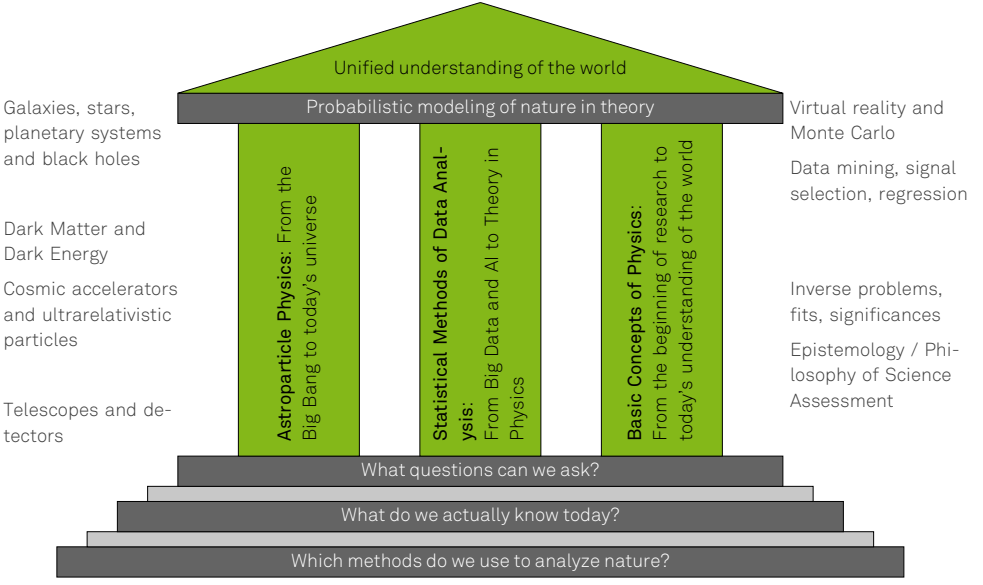
technische universität
dortmund

Statistical Methods of Data Analysis

Organization

Prof. Dr. Dr. Wolfgang Rhode Dr. Maximilian Linhoff

2023


Unified understanding of the world
Probabilistic modeling of nature in theory
Galaxies, stars, planetary systems and black holes
Dark Matter and Dark Energy
Cosmic accelerators and ultrarelativistic particles
Telescopes and detectors
Astroparticle Physics: From the Big Bang to today's universe
Statistical Methods of Data Analysis: From Big Data and AI to Theory in Physics
Basic Concepts of Physics: From the beginning of research to today's understanding of the world
Virtual reality and Monte Carlo
Data mining, signal selection, regression
Inverse problems, fits, significances
Epistemology / Philosophy of Science Assessment
What questions can we ask?
What do we actually know today?
Which methods do we use to analyze nature?

	4. Sem	5. Sem	6. Sem	1. Sem		2. Sem	Pool		
			Astroteilchen I (VL)	Radioastronomie (Sem)	Astroteilchen II (VL)	Neutrino-/Gamma- Astronomie (Sem)	Kosmische Strahlung (Sem)	Masterarbeit	Promotion
			←Grundbegriffe→	Fortgeschrittene Theorie (VL) ⇐	ART, QFT, Kosmologie ⇒	Fortgeschrittene Theorie (VL)			
			Spez. Themen d. Teilchenphysik (Sem)	ETT (VL)		Exp. Aspekte der Teilchenphysik (VL)	Flavourphysik (VL)		
		KET (VL)		Detektorphysik (VL)			Falsche Entdeckungen (Sem)		
Toolbox Workshop	SMD A	SMD B		Schlüsselexperimente (Sem)		Detektor-Systeme (Sem)	Lesekurs TP (Sem)		
						Elektronik (VL)			
						Elektronik (PR)	Ext. Schulen		
				FP (PR) / Begleitsem. (Sem)		FP II (Block PR)	NF: Informatik, Statistik, E-Technik		
				Datenanalyse (Sem)	SMD II (VL)		Maschinelles Lernen (Sem)		

Bachelorarbeit

Statistical Methods of Data Analysis

- Teaches the methods necessary to understand experimental data (for experimentalists and theorists)
- Provides skills needed to evaluate the relevance of theories
- Significantly better results are achieved in bachelor and master theses
- Acquired skills are also highly relevant outside of academia (→ industry, Digitale Agenda des Bundes, ...)

Plan: SMD before bachelor thesis, then more data analysis modules before master thesis

SMD / Big Data / Data Science

Experimental physics

- How can we extract relevant information from the very few things we can actually measure?
- How do we draw conclusions about the actual physical quantities from observed distributions?

Theoretical physics

- How well can a theory be scientifically tested?
- Am I a physicist or a meta-physicist?

Economics

- How can we draw optimal consequences from existing information?

Engineering

- Predictive maintenance
- Autonomous machines
- ...

Philosophy

- What methods can we use to gain insights?

Topics SMD-A

- Data analysis fundamentals in Python
- Numerical basics
- 1D-distributions
- Random number generation
- nD-distributions
- Monte Carlo simulations
- Supervised Machine Learning
- Unsupervised Machine Learning

Lectures

- Wednesdays 10:15 in HG2 HSII
- “Classic” lectures
- Live hands-on sessions

Exam

- Medical Physics
  - SMD-A with 4 CP
  - SMD-B (winter term) with 5 CP
- Physics
  - SMD-A with 5 CP
  - SMD-B (winter term) with 5 CP

There will be oral exams in two slots, right after the semester and in September.

Exam Requirements

Successful participation in the tutoriums.
That means:

- Do not be absent unexcused more than two times
- **> 50%** of the points in graded sheets
- Present your solutions at least two times

Sign Up in the LSF

Tutoriums

Thursday	12–13	CP-03-150	Lukas Beiske, lukas.beiske@tu-dortmund.de Tristan Gradetzke, tristan.gradetzke@tu-dortmund.de
Friday	9–10	CP-03-150	Stefan Fröse, stefan.froese@tu-dortmund.de Vincent Latko, vincent.latko@tu-dortmund.de
Friday	10–11	CP-03-150	Jonas Hackfeld, jonas.hackfeld@ruhr-uni-bochum.de Ludwig Neste, ludwig.neste@tu-dortmund.de

Choose your tutorium

Indicate your preference in the poll by Friday (April 7th)

Please specify at least one second preference.

If you already know your submission partners, let us now by using the comment field.
`https://terminplaner6.dfn.de/p/d1a752f78feb0c002a8f32e65402a1d4-196195`

Tutoriums

- First tutorium: April 13. / 14.
- We will work together on exercises in the first tutorium
- Exercise Sheets:
  - Handout: Wednesdays
  - Submission: Tuesdays 23:59
  - First graded exercise:
    - Handout: April 19
    - Submission: April 25

Submission of Exercises

- Submission due on Tuesdays at 23:59 (hard deadline set in Moodle)
- Submit a ZIP-file `Sheet{number:02d}_{Surname1}_{Surname2}_{Surname3}.zip` including:
  - PDF with all calculations, results, comments
    - Make clear where you get your results
    - Include all plots you created
  - Code for programming exercises
    - Talking variable names, comments (trains good coding practices)
    - Code has to run without errors without further ado
    - Use python standard library functions, as well as `numpy`, `pandas`, `matplotlib`
    - If explicitly mentioned, you are allowed to use other packages, e.g. `scikit-learn`
- If you cannot solve an exercise, give detailed explanations what you tried and what you wanted to achieve

Submission of Coding Exercises

- Python ≥ 3.10
- One python file per exercise
- Meaningful name, e.g. `week01_exercise1.py`
- The solutions must be produced by executing the file once:

```
python week01_exercise1.py
```

- Very long runtimes indicate inefficient code
  - → Code should run a few minutes at most and usually in a couple of seconds
- Discuss your results in the PDF file according to the exercise
- Hint: use a workflow automation (`Makefile` or snakemake) to produce all results

Submission with Jupyter Notebooks

- Alternative to the submission of python text-files
- Meaningful name, e.g. `week01_exercise1.ipynb`
- Needs to run from top to bottom
  - → Check with `Restart Kernel and Run All` before submission
- Discuss results and answer exercise tasks in markdown cells
- Export PDF from jupyter notebook, and include that in your submission

```
\usepackage{pdfpages} % in preamble
\includepdf{...}      % in document
```

```
pdftk week01_exercise1.pdf week01_exercise2.pdf cat output
 ↪  week01_merged.pdf
```

- Include jupyter notebook in ZIP file

Software for Tutoriums

- All coding exercises with python ≥ 3.10
- Setup your computer before the first tutorium
- We recommend installing mambaforge:
  `https://github.com/conda-forge/miniforge#mambaforge`
- Our tip: use a dedicated `conda` environment

```
mamba env create -f environment.yml  # find in moodle
```

- There will be a project exercise spanning both semesters and some exercise sheets, details later

https://physikon.pep-dortmund.org

TU Dortmund Help Programs

DoBuS

Behinderung und Studium

https://dobus.zhb.tu-dortmund.de

Psychologische Studienberatung

https://www.tu-dortmund.de/psychologischeberatung

Please feel free talk to us if you need any support!