



Doctoral School of Molecular Medicine
Medical University of Lodz

SMM Autumn School 2026

From Data to Evidence in Molecular Medicine:
Statistics, AI, Genomics, and Scientific Writing and
Publishing

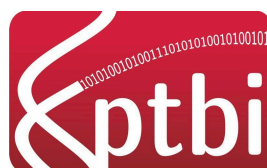
Strategic Sponsor



Sponsors & Partners



bioingen



Overall description of SMM autumn Doctoral School entitled “From Data to Evidence in Molecular Medicine: Statistics, AI, Genomics, and Scientific Writing and Publishing”

25.11.2026

1. 10:30-11:00 Where angels fear to tread, Statisticians calculate: An introduction to statistical measures and Testing.

In this short lecture we will have a rigorous yet accessible introduction to the foundation of statistical concepts. We will begin by deconstructing core descriptive metrics, examining the critical differences between the arithmetic mean and the median, especially with outliers in data. Moreover, we will briefly explore advanced measures of tendency and deviation. Next, with this knowledge the session will concentrate on probability distributions, emphasizing the difference between discrete and continuous problems. Finally, the lecture will culminate in a practical, decision-based roadmap for statistical inference. By systematically linking specific data topologies and distribution models to their corresponding methodological choices, we will learn exactly how and when to select the appropriate statistical test.

Lecturer: dr inż. Joanna Żyła

No laptop needed

2. 11:00-12:30 group A; First steps in R and RStudio.

During the course, we will become familiar with the server-based data analysis environment; together we will learn the basic functionalities of RStudio and selected default R functions. We will focus on loading various data types, working with vectors, matrices, and data frames. We will spend time cleaning the data and selecting subsets by selecting columns, rows, and their specific values. We will save the data subsets to individual variables. During this part, everyone will be able to upload their own data to RStudio, so please save a sample of your data (several dozen columns and rows) in .csv, .xls, or other tabular format beforehand.

Workshop lead: dr Michał J. Dąbrowski

Teaching assistants: dr inż. Michał Damiński; dr Marta Jordanowska-Kotuniak; dr Aleksandra Cabaj

Laptop needed!



biouigen



3. 11:00- 17:00 group B; Hackathon - Single-cell multi-omic cancer data analysis.

During this hackathon, participants will leverage a cutting-edge, multimodal dataset from a recent *Nature Communications* study on human gliomas to design and execute a mini scientific project. Teams will work in parallel, exploring diverse computational tasks that range from single-cell transcriptomics and epigenomics (scRNA/snATAC-seq) to spatial cellular architecture. Using state-of-the-art bioinformatics toolkits in Python or R, students will reconstruct intercellular niches, model transcription factor activity, and map functional states within the tumor microenvironment. Advanced tracks will also offer the opportunity to integrate high-resolution histological imaging with molecular profiles using machine learning.

Workshop lead: dr hab. Bartosz Wojtas

Teaching assistant: dr inż. Joanna Żyła

Laptop needed!

4. 13:00-14:45 group A; Introduction to Statistical Testing — From Hypotheses to Adequate Tests. A Practical Guide to Choosing and Applying Statistical Tests in R.

During this session, participants will learn the fundamentals of statistical hypothesis testing and how to select an appropriate statistical test for a given research question and dataset. We will use a practical decision-tree approach to guide the choice of test, taking into account the type of variables, distribution of data, and assumptions underlying different statistical methods. Using real-world datasets and practical examples in R, participants will work through the process of formulating hypotheses, selecting appropriate statistical tests, and interpreting the results.

Workshop lead: dr inż. Michał Damiński

Teaching assistants: dr Michał J. Dąbrowski; dr Marta Jordanowska-Kotuniak; dr Aleksandra Cabaj.

Laptop needed!

5. 15:30-17:00 group A; Introduction to data visualisation in R.

In this workshop, we will learn the fundamental principles of data visualisation in R. We will explore how to select appropriate visualisation methods for different types of data and become familiar with commonly used plot types. We will learn the basic guidelines for creating clear, informative and reliable figures, with a focus on matching the choice of visualisation to the data and the research question. Good practices in data presentation, as well as common mistakes and how to avoid them, will also be discussed. Through practical examples, we will develop the basic skills needed to create effective figures for data exploration and scientific communication.

Workshop lead: dr Marta Jordanowska-Kotuniak

Teaching assistants: dr Michał J. Dąbrowski; dr inż. Michał Damiński; dr Aleksandra Cabaj.



Laptop needed!

6. From Calculating the Mean to Large Language Models — The Algorithms Behind AI.

This lecture explores the mathematical and algorithmic foundations underlying modern Artificial Intelligence, starting with seemingly simple concepts such as calculating the mean, variance, and basic statistics. We will follow the evolution from classical statistical methods through ML (machine learning) algorithms to today's complex AI models. The lecture will explain the key ideas behind how these algorithms learn from data, make predictions, recognize patterns, and generate new information. Finally, we will connect these foundations to modern Large Language Models (LLMs) and discuss how relatively simple mathematical operations can build up to surprisingly powerful AI systems.

Lecturer: dr inż. Michał Damiński

No laptop needed

26.11.2026

1. How Not to Fall into the Traps of Experimental Design, Data Analysis and Scientific Reasoning.

This workshop introduces the most common pitfalls encountered in experimental design, statistical analysis, and interpretation of biomedical data. Using practical examples from molecular biology, genomics, and bioinformatics, participants will learn how inappropriate study design, misuse of statistical methods, and cognitive biases can lead to misleading conclusions. Topics include effect size interpretation (fold change, odds ratio, hazard ratio), statistical power, multiple testing, p-hacking, data dredging, correlation versus causation, survival bias, and the influence of sample size on statistical significance. The workshop emphasizes critical thinking and good research practices rather than mathematical formalism.

Lecturer: dr hab. Bartosz Wojtas

No laptop needed

2. In search of important signals - statistical significance, noise elimination, selection of informative features.

This workshop will combine hands-on exercises with short lecture components. We will explore the concept of statistical significance and examine how the relationship between the number of features and the number of samples affects our ability to identify meaningful signals in high-dimensional data.

Participants will perform hypothesis testing with multiple-testing correction, apply feature-selection methods to reduce data dimensionality, and learn how to distinguish informative features from noise. Finally, we will identify a set of genes showing significant differential expression between two groups and carry out a functional analysis to interpret their biological relevance.

Lecturer: dr Michał J. Dąbrowski

Laptop needed!

3. Scientific Evidence and Scholarly Publishing.

This session will introduce participants to the principles of evaluating and interpreting scientific evidence, with a particular focus on its reliability, classification, and hierarchy. Participants will explore the criteria used to assess different types of evidence and learn how these influence research evaluation and inform decision-making.

The second part of the session will focus on effective scientific writing, providing practical guidance on how to present research findings clearly, convincingly, and strategically to maximise their scientific impact and increase the likelihood of publication in high-quality, peer-reviewed journals.

Lecturer: prof. Magdalena Chechlińska

No laptop needed

4. Predictive Models: How Not to Fool Yourself with Data Analysis.

During this session, we will explore how to build, evaluate, and validate predictive models using statistical and machine learning approaches. We will discuss the problem of false positive results, using *DESeq2* and *MCFS-ID* as practical examples of how different analytical approaches can lead to different conclusions. Particular attention will be given to large datasets, where statistically significant results do not necessarily imply meaningful or useful effects. We will discuss strategies for detecting and reducing false positive findings, including appropriate statistical testing, multiple testing correction, independent validation, and avoiding overfitting and data leakage. We will also examine the role of effect size in distinguishing statistical significance from practical or biological significance.

Lecturer: dr inż. Michał Damiński

Laptop needed!

27.11.2026

1. AI for Practical Statistical Analysis and Scientific Writing

Participants will learn how to use ChatGPT to support statistical analysis and scientific writing in molecular medicine. The session will focus on practical use: selecting methods, generating and checking analyses, interpreting results, and preparing publication-ready outputs. It will also cover data security, differences between web-based and local models, publication ethics, and responsible use of AI in research.



biouigen



Lecturer: dr Przemysław Tomczyk

Laptop and free ChatGPT account needed!

2. From Next-Generation Sequencing to Precision Oncology: Modern Genomics in Cancer Research.

This lecture provides an overview of modern sequencing technologies and their applications in cancer research. Starting from the evolution of next-generation sequencing platforms, participants will learn how genomic, transcriptomic, epigenomic and single-cell approaches are used to investigate tumour biology, identify therapeutic targets, and monitor disease progression. The course will introduce major international resources such as The Cancer Genome Atlas (TCGA), discuss mutational signatures and tumour evolution, and present examples of how multi-omics data support precision oncology. Practical examples from recent biomedical research will illustrate both the opportunities and the current limitations of genomic technologies in basic and translational cancer research.

Lecturer: dr hab. Bartosz Wojtas

No laptop needed

3. Cracking the Black Box: The mechanics of PCA, UMAP, and pathway enrichment in Single-Cell Analysis.

Serving as a methodological synthesis for the entire Autumn School curriculum, this lecture systematically cracks the analytical "black box" that so often obscures unsupervised machine learning in single-cell transcriptomics. While modern computational biology is full of complex algorithmic nomenclature, the mathematics behind these methods may not be so complicated and lead to Pythagora's theorem. We will rigorously deconstruct the underlying mechanics of dimensionality reduction, demonstrating how frameworks such as Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) are ultimately governed by foundational geometry and linear algebra. The second part will be devoted to pathway enrichment analysis methods including a single-sample approach. Here, again it may turn out that the simple mean serves as a powerful tool to crack cell heterogeneity.

Lecturer: dr inż. Joanna Żyła

No laptop needed