Statistics and Machine Learning with Python

Module 3: Statistics, Machine Learning, and Model Evaluation

Dr. Yves J. Hilpisch[1]

March 27, 2026

[1] Get in touch: https://linktr.ee/dyjh. Web page: https://thedatascientist.dev. Research, structuring, drafting, and visualizations were assisted by GPT 5.x as a co-writing tool under human direction. Comments and feedback are welcome.

Preface

This preface explains how Module 3 builds on the earlier work in `The Data Scientist`. The focus now shifts from manipulating data to reasoning about uncertainty, training baseline machine learning models, and evaluating them with calm judgment.

Module 1, *Python Programming Foundations for Data Science* [2], teaches Python as a working language. Module 2, *Python Data Analysis, Visualization, and Storytelling* [3], turns that fluency into data fluency: arrays, tables, and figures. This volume, *Statistics, Machine Learning, and Model Evaluation* [4], adds a new layer: statistical thinking and machine learning judgment. The aim is not to cover every algorithm. It is to help the reader make better decisions when building, evaluating, and explaining simple models.

The module is therefore not a full statistics or machine learning reference. Instead, it gives the reader enough practical intuition to:

- describe variability and uncertainty in data,
- frame supervised learning problems clearly,
- train and compare simple models using `scikit-learn`,
- and explain results and limitations without hype.

The chapters assume that the reader is comfortable with Python, NumPy, pandas, and basic visualization from the earlier modules. When those tools appear here, they serve supporting roles for statistical and modeling ideas rather than being new subjects of their own.

The working style remains the same:

- short reading sections,
- small, executable code examples,
- guided exercises and review questions,
- and one coherent capstone notebook.

Readers are encouraged to treat this book less as a reference and more as a coaching guide. The best way to study it is to:

- run and slightly modify every example,
- keep a running notebook of questions and insights,
- and connect each new idea to a small, self-chosen project.

By the end of the module, the goal is not just that the reader can fit models. It is that they can say, calmly and clearly, why a particular model and evaluation setup make sense for a given problem.

Contents

List of Figures

Contact

Statistics and Machine Learning with Python

Module 3: Statistics, Machine Learning, and Model Evaluation

Get in touch:

linktr.ee/dyjh

thedatascientist.dev