Luben M. C. Cabezas

São Carlos - SP | lucruz45.cab@gmail.com | +55 16 99233-9433 | Luben Miguel Cruz Cabezas | Monoxido45

Education

Phd — Aug 2022 – Present

Federal University of São Carlos, Statistics
Institute of Mathematics and Computer Science - USP, Statistics

- **GPA:** 4.0/4.0
- **Relevant Coursework:** Advanced Probability, Advanced Inference, Statistical Learning, Advanced Topics in Statistical Learning, Stochastic Simulation
- **Fellowship:** Funded by FAPESP (Grant 2022/08579-7)

BS — Mar 2018 – Apr 2022

Federal University of São Carlos, Statistics

- GPA: 9.24/10
- **Relevant Coursework:** Probability, Statistical Inference, Bayesian Inference, Machine Learning, Regression analysis, Stochastic Processes, Computationally Intensive Methods

Working Experience

Statistical Machine Learning Lab - SMaLL (UFSCar), Researcher — São Carlos, SP, Aug 2022 – Present

- Ph.D. researcher at SMaLL, supervised by Rafael Izbicki.
- Research focused on Conformal Prediction, Likelihood-Free Inference, Non-Parametric Inference, and Hypothesis Testing.
- Developed *LOCART*, a novel locally adaptive conformal prediction method, published as a full-length paper in Information Sciences.
- Designed and maintained accessible Python and R packages to implement research frameworks.

Terranova (Brazilian Jurimetrics Association), Statistics Intern — São Paulo, SP, Nov 2021 – June 2022

- Developed and maintained internal R-based packages for data analysis and modeling.
- Worked on proof-of-concept (POC) development for predictive models.
- Contributed to the implementation and debugging of Shiny dashboards.

Statistical Machine Learning Lab - SMaLL (UFSCar), Research Assistant — São Carlos, SP, Nov 2020 – Nov 2021

- Undergraduate researcher at SMaLL, supervised by Rafael Izbicki.
- Developed new visualization techniques, feature importance metrics, and scoring methods for hierarchical clustering using phylogenetic probabilistic models. This work resulted in a full-length paper published in Applied Soft Computing (2023).
- Research funded by FAPESP (Grant 2020/10861-7).

Publications

Regression trees for fast and adaptive prediction intervals — Jan 2025

Luben M. C. Cabezas, Mateus P. Otto, Rafael Izbicki, Rafael B. Stern
10.1016/j.ins.2024.121369

Hierarchical Clustering: visualization, feature importance and model selection — July 2023

Luben M. C. Cabezas, Rafael Izbicki, Rafael B. Stern
10.1016/j.asoc.2023.110303

Epistemic Uncertainty in Conformal Scores: A Unified Approach (pre-print) Feb 2025
Luben M. C. Cabezas, Vagner S. Santos, Thiago R. Ramos, Rafael Izbicki
arxiv:2502.06995

Distribution-Free Calibration of Statistical Confidence Sets (pre-print) Nov 2024
Luben M. C. Cabezas, Guilherme P. Soares, Thiago R. Ramos, Rafael B. Stern, Rafael Izbicki
arxiv:2411.19368

REACT to NHST: Sensible conclusions to meaningful hypotheses (pre-print) Aug 2023
Rafael Izbicki, **Luben M. C. Cabezas**, Fernando A. B. Colugnatti, Rodrigo F. L. Lassance, Altay A. L. de Souza, Rafael B. Stern
arxiv:2308.09112

Skills

Coding Languages

- Advanced: Python, R, Shell Script
- Intermediate: SQL

Developing tools:

- Advanced: Git, Markdown, Linux, Latex

Languages

- Portuguese: Native
- English: Fluent
- Spanish: Advanced
- French: Beginner

Teaching Experience

Teaching Assistant, Department of Statistics, UFSCar. Assisted in the course "Statistical Inference" for second-year Statistics students (90 hours total) during the second semester of 2022 (class 1001736). Responsible professor: Rafael Izbicki.

Teaching Assistant, Department of Statistics, UFSCar. Assisted in the course "Data Mining" for third-year Statistics students (60 hours total) during the first semester of 2022 (class 158518). Responsible professor: Rafael Izbicki.

Work Presentations

VII Latin American Meeting on Bayesian Statistics (COBAL) and XVII Brazilian Meeting of Bayesian Statistics (EBEB): Pragmatic variable selection in nonparametric binary regression, 2024.

X Workshop on Probabilistic and Statistical Methods (WPSM): Regression Trees for Fast and Adaptive Prediction Intervals, 2024

67ª RBras e 20º SEAGRO: Toward Local and Valid Uncertainty Estimation in ML, 2023

XVI CLAPEM (Latin America Congress of Probability and Mathematical Statistics) Towards Local and Valid Uncertainty Estimation in ML, 2023.

XXVIII Congresso de Iniciação Científica e XIII Congresso de Iniciação em Desenvolvimento Tecnológico e Inovação (UFSCar) A data-splitting approach for comparing hierarchical clustering algorithms, 2022.

XXVII Congresso de Iniciação Científica e XII Congresso de Iniciação em Desenvolvimento Tecnológico e Inovação (UFSCar): Previsão em modelos filogenéticos, 2020