Salvador Balkus

Curriculum Vitae

sbalkus@g.harvard.edu
salbalkus.github.io
salbalkus
salvador-balkus
0000-0003-4695-833X

Research Interests

Statistical — Causal inference, debiased machine learning, spatial and network data

Scientific — Environmental health, pollution, public policy

Education

2022—2027 **Doctorate of Philosophy in Biostatistics**, *Harvard University*
Advised by Professor Nima Hejazi

2018—2022 **Bachelor of Science in Data Science**, *University of Massachusetts Dartmouth*
Minors: Mathematics, Economics. *GPA*: 4.0/4.0; Summa Cum Laude

Experience

2022—2024 **Graduate Student Researcher**, *Environmental Statistics Training Grant*,
Harvard University, Cambridge, MA
Analyzed spatial and environmental data, developed and critiqued statistical methods, studied effects of electric vehicle uptake on air pollution in California

2020—2022 **Research Assistant**, *Computational Statistics & Data Science Lab*,
University of Massachusetts Dartmouth, North Dartmouth, MA
Harmonized dietary data in SAS, developed R package for COVID-19 modeling

Jun—Aug 2021 **Research Project Manager**, *Research in Industrial Projects for Students*,
Institute for Pure and Applied Mathematics, Los Angeles, CA
Led team to develop object tracking algorithms and visualizations for The Aerospace Corporation

Jun—Aug 2020 **Student Researcher**, *Ecological Modeling REU*,
University of Wisconsin, La Crosse, Remote
Led team to develop forest cover classification model for U.S. Geological Survey

2019—2020 **Research Assistant**, *Public Policy Center*,
University of Massachusetts Dartmouth, North Dartmouth, MA
Designed infographics and reports communicating data to local governments and organizations

Publications

[1] **S. V. Balkus**, S. W. Delaney, and N. S. Hejazi, "The causal effects of modified treatment policies under interference", *arXiv:2412.02105*, 2024.

[2] **S. V. Balkus** and N. S. Hejazi, "`CausalTables.jl`: Simulating and storing data for statistical causal inference in julia", *Under review at Journal of Open Source Software*, 2024.

[3] **S. V. Balkus** and D. Yan, "Improving short text classification with augmented data using GPT-3", *Natural Language Engineering*, pp. 1–30, 2023. DOI: 10.1017/s1351324923000438.

[4] **S. V. Balkus**, H. Fang, and H. Wang, "Federated fuzzy clustering for longitudinal health data", *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 128–132, 2022.

[5] **S. V. Balkus**, H. Wang, B. D. Cornet, C. Mahabal, H. Ngo, and H. Fang, "A survey of collaborative machine learning using 5G vehicular communications", *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1280–1303, 2022. DOI: 10.1109/comst.2022.3149714.

[6] V. S. Gurugubelli, H. Fang, J. M. Shikany, **S. V. Balkus**, J. Rumbut, H. Ngo, H. Wang, J. J. Allison, and L. M. Steffen, "A review of harmonization methods for studying dietary patterns", *Smart Health*, vol. 23, p. 100 263, 2022. DOI: 10.1016/j.smhl.2021.100263.

[7] **S. V. Balkus**, H. Fang, J. Rumbut, A. Moormann, and E. Boyer, "A multi-level biosensor-based epidemic simulation model for COVID-19", *IEEE Internet of Things Journal*, pp. 1–1, 2021. DOI: 10.1109/jiot.2021.3127804.

[8] **S. V. Balkus**, J. Rumbut, H. Wang, and H. Fang, "An adaptive and dynamic biosensor epidemic model for COVID-19", in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, IEEE, Aug. 2020. DOI: 10.1109/iri49571.2020.00051.

Presentations

[1] **S. V. Balkus**, "Statistics in Julia: Is it right for you?", in *Harvard Biostatistics Student Seminar*, Dec. 2024.

[2] **S. V. Balkus**, "Nonparametric network causal inference for continuous exposures in mobile source air pollution", in *American Causal Inference Conference*, May 2024.

[3] **S. V. Balkus**, "Assumption-lean causal inference for mobile source air pollution", in *ASA Boston Chapter Student Research Symposium on Statistics and Data Science*, Apr. 2024.

[4] **S. V. Balkus**, "Improving natural language classification with augmented data from GPT-3", University of Massachusetts Dartmouth, Apr. 2022.

[5] **S. V. Balkus**, "Language models that teach themselves: Augmenting training data for topic classification using GPT-3", in *ASA Boston Chapter Student Research Symposium on Statistics and Data Science*, Apr. 2022.

[6] N. Pai, **S. V. Balkus**, and T. Zeng, "Multi-hypothesis tracking of space objects and targets", in *AMS Joint Mathematics Meetings (JMM) Poster Session*, Apr. 2022.

[7] N. Pai, **S. V. Balkus**, and T. Zeng, "Institute for pure and applied mathematics", in *Multi-Hypothesis Tracking of Space Objects and Targets*, Aug. 2021.

[8] **S. V. Balkus**, "Multi-level biosensor-based epidemic forecasting in small areas", in *Joint Statistical Meetings*, American Statistical Association, Aug. 2021.

[9] **S. V. Balkus**, M. McDevitt, and N. Dean, "A classification system for characterizing diversity across floodplain forests of the Upper Mississippi River System", University of Wisconsin La Crosse, Aug. 2020.

[10] **S. V. Balkus**, "Lunchtime computing: Basics of AWS sagemaker", in *Center for Science Computing and Visualization Research*, University of Massachusetts Dartmouth, Feb. 2020.

Awards

2024 **Certificate of Distinction in Teaching**, *BST231: Methods*, *Harvard University*

2022 **Graduate Research Fellowship**, *National Science Foundation*

2022 **Best Overall + Visualization**, *SEMASS DataFest*, American Statistical Association

2021 **Best Visualization**, *SEMASS DataFest*, American Statistical Association

2021 **Travel Award**, *Joint Mathematics Meetings*, American Mathematical Society

Teaching

Fall 2024 **Teaching Fellow**, *BST 232: Methods I*, Harvard University

Summer 2024 **Curriculum Fellow**, *BST 232: Methods I*, Harvard University

Summer 2024 **Qualifying Exam Tutor**, *Topic: Optimization*, Harvard University

Summer 2024 **Workshop Instructor**, *Replicathon, IQ BIO REU*, University of Puerto Rico Río Piedras

Fall 2023 **Teaching Fellow**, *BST 232: Methods*, Harvard University

Service and Extracurricular

Ad-hoc Reviewer
Journal of Causal Inference (2x), Journal of Open Source Software, Natural Language Processing

2024 —Present **Co-Chair**, *Biostatistics Student Committee*, Harvard University
Organized events and peer mentoring for graduate students

2023 —2024 **Commitee Chair**, *Biostatistics Peer Mentoring Program*, Harvard University
Supported incoming students by serving as peer mentor, creating instructional material, acquiring event funding, and planning seminar series

2022 —2024 **Graphic Designer and Blog Contributor**, *Science in the News*, Harvard University
Designed infographics and wrote blog posts explaining scientific topics to a lay audience

2022 **Biostatistics Session Chair**, *Student Research Symposium on Statistics and Data Science*, ASA Boston Chapter

2020—2022 **President**, *Big Data Club*, University of Massachusetts Dartmouth
Organized data science workshops and networked with local clients for consulting projects

Skills

Systems Linux (Fedora), Windows

Languages `Julia`, `R`, `Python`, LaTeX, markdown, `HTML/CSS`

Tools Git, GitHub, Quarto, RStudio, VSCode, AWS, SLURM

Open-Source Contributor To...

- The Book of Statistical Proofs, `https://statproofbook.github.io/`
- `TMLE.jl`
- `DensityRatioEstimation.jl`

last updated 2024-09-16