Jill Lundell, PhD, PSTAT

Dana-Farber Cancer Institute
Department of Data Sciences
3 Blackfan Circle
Boston MA 02115
jlundell@hsph.harvard.edu
www.jilllundell.com

Education

PhD, **Mathematical Sciences**, Utah State University, Logan, Utah 2019
Dissertation: Tuning Hyperparameters in Supervised Learning Models and Applications of Statistical Learning in Genome-Wide Association Studies with Emphasis on Heritability.
Advisor: Dr. D. Richard Cutler

M.S., Statistics, Utah State University, Logan, Utah 1998
Thesis: On the Model Selection in a Frailty Setting
Advisor: Dr. Olcay Akman

B.S., Mathematics, Utah State University, Logan, Utah 1996

Experience

Postdoctoral Research Fellow, Harvard T.H. Chan School of Public Health and Dana-Farber Cancer Institute Department of Data Science 2019-Present
Boston, Massachusetts
Mentor: Dr. Rafael Irizarry

- Develop and implement data science technologies in collaborative research at Dana-Farber Cancer Institute.
- Develop new data science methods in omics and machine learning.
- Organize the Quantitative Issues in Cancer Research Working Seminar.
- Assist in grant writing for collaborative work and seek funding for own research.
- Write and promote Bioconductor packages for research projects.
- Publish articles.

Senior Statistician, Water and Environmental Technologies 2019-Present
Butte, Montana
Water and Environmental Technologies is a small geoscience, environmental, and engineering firm located in Montana.

- Provide statistical support for projects throughout the company including all of the statistical design and analysis for groundwater monitoring of five large coal powerplants.
- Write R packages to streamline analysis and ensure reproducibility for statistical projects that are ongoing at Department of Energy laboratories and other environmental sites.

Senior Statistician, North Wind, Inc. 2017–2019
Idaho Falls, Idaho
North Wind specializes in environmental management, scientific consulting, and construction.

- Managed all statistical activities at North Wind Group and affiliated North Wind companies.
- Created sampling designs, identified data needs, performed all statistical analyses, and wrote analytical reports.
- Wrote proposals to obtain external funding from private companies and government agencies.
- Provided statistical expertise on projects in many disciplines including nuclear energy, ecology, environmental remediation, and environmental monitoring.
- Managed all non-routine sampling activities for Idaho National Laboratories.

Senior Statistician, Portage, Inc. 2000–2017
Idaho Falls, Idaho
Portage, Inc. specialized in environmental management, engineering and scientific consulting, and information technology.

- Performed the same duties as at North Wind, Inc.
- Created and taught short courses to clients and employees on introductory statistics, including sampling design and hypothesis testing.
- Wrote proposals to obtain new statistical work for Portage and was a key contributor to proposals for large company contracts. Funding was secured from agencies such as the Environmental Protection Agency, U.S. Army Corps of Engineers, Bureau of Land Management, and private companies.
- Wrote business plans to propose and develop new company services and software related to statistics and data technologies.

Faculty, Brigham Young University Idaho 2011-2014
Rexburg, Idaho

- Taught undergraduate mathematics and statistics to classes of up to 50 students.
- Assisted in the creation of a statistics minor for the university with two other instructors.
- Developed a new introduction to statistics curriculum with a team of four instructors and curriculum designers.
- Created new upper division statistics courses.
- Created a training module for faculty and students to teach them how to use R.
- Developed new educational strategies for teaching collaboration skills and computational statistics.

Adjunct Faculty, Idaho State University
Pocatello, Idaho

- Taught undergraduate level classes in mathematics and statistics. 2001–2003

Temporary Lecturer and Consultant, Utah State University 1998–2000
Logan, Utah

- Taught undergraduate and graduate level classes in mathematics and statistics.
- Ran the university statistical consulting center with Dr. Richard Cutler. The center provided statistical consulting to all faculty and graduate students doing research at the university.

Graduate Research Assistant, Los Alamos National Laboratories 1997
Los Alamos, New Mexico

- Provided statistical analysis for a Hepatitis C phylogenetic project in the Theoretical Biology group. I was the only statistician in the group so I had to research the methods that were needed for the project. Methods included bootstrapping genetic code, classifying phylogenies, and bootscanning to look for mosaicism in viral RNA.
- Discovered that a type of hepatitis C that had been identified as a separate species was a short mutation in the RNA for a small group of people in Indonesia. That group was reclassified as a subtype of a different species because of this discovery.

Honors and Awards

Presidential Doctoral Research Fellow at Utah State University 2015-2019

Second place winner of the 2018 Data Expo at the Joint Statistical Meetings (JSM) in Vancouver, British Columbia. This competition is sponsored by the Section on Computational Statistics and Graphics of the ASA and is held approximately every 3 years. The competition requires innovative and comprehensive analysis of a provided data set with results presented at a special session of JSM. My entry was titled "Let's talk about the weather". 2018

PhD Recipient of the Ellis R. Ott Scholarship for Applied Statistics and Quality sponsored by the ASQ Statistics Division. One scholarship is awarded to a Ph D student each year. 2018

Exemplary Faculty Award at Brigham Young University Idaho 2013

Programming Languages and Computer Skills

Proficient	R, Python, SAS, machine learning, cluster computing, LaTeX, Linux/Unix, git, bash, GATK, samtools, plink, other bioinformatics software
Working Knowledge	Deep learning
Exposure	C, C++

Funding Awards

Award amounts are considered a company proprietary information so I cannot identify the contract amounts for the projects. Awards ranged from approximately $20,000 to $250 million dollars.

Statistical analysis for groundwater monitoring at the Idaho CERCLA Disposal Facility at Idaho National Laboratory. CWI and Fluor (PI for Portage, North Wind, and Water and Environmental Technologies). 2010-2022

Statistical analysis and sampling design for the Idaho National Laboratory environmental monitoring program for radiation, soil, and air contamination. (PI for Water and Environmental Technologies, subject matter expert for North Wind). 2019-2022

2018. Butte Area One Parrot tailings removal, Water and Environmental Technologies (PI for North Wind). 2018

Statistical analysis of coal combustion residuals for PacifiCorp coal power plants, Water and Environmental Technologies (PI for North Wind). 2017-2022

Laboratory quality software database with data analytics. EZ Analytics (PI for Portage). 2015

Luckey, Ohio, remediation. United States Army Corps of Engineers (Subject matter expert for Portage). 2014

Data life cycle software. Portage principal owners (PI for Portage). 2014

Libby Asbestos OU-3 Tubb Gulch Drainage Site investigation and engineering evaluation and cost analysis development. USDA-Forest Service, Northern Regional Forester's Office Missoula, Montana (Subject matter expert for Portage). 2014

Expert statistical modeling support for data evaluation and risk assessment. United States Environmental Protection Agency Region 8 (PI for Portage). 2010-2011

Montana dioxin background study. Montana Department of Environmental Quality (Co-PI Portage). 2009-2011

Travel award for the Graybill Conference on Statistical Genomics and Genetics, Fort Collins, Colorado. 2017

Published R Packages

cytofQC, a package for labeling CyTOF data for cleaning. 2022
cytofQC can be downloaded from Github and is currently under review with Bioconductor. The package can import an fcs file and identify events that are beads, debris, aggregates, or permeable using a statistical modeling approach. Events are labeled with the most likely event type and metrics are provided that

indicate how much each event resembles the different event types. Plotting and reporting functions are also included.

VDJdive, a package for analyzing immune receptor repertoire data. 2022
VDJdive can be downloaded from Bioconductor. The package contains functions for reading immune receptor repertoire data, merging it with paired single cell data, quantifying clonotype abundances, calculating diversity metrics, and producing graphs. It implements an E-M algorithm to make use of ambiguous cells for improved quantification.

EZtune: a package for auto tuning supervised learning models. 2018
EZtune can be downloaded from CRAN. The package contains functions for automatically tuning and validating support vector machines, gradient boosting machines, adaboost, and elastic net. Functions are designed to be easy to use for novice R users and can find a well-tuned model with minimal computation time.

Submitted Publications

Lundell JF. Tuning support vector machines and boosted trees using optimization algorithms. *Submitted to Electronic Journal of Statistics March 2023 and awaiting response*. ArXiv:2303.07400. March 2023. https://doi.org/10.48550/arxiv.2303.07400. 2023

Lundell JF. EZtune: An R package for simple tuning of supervised learning models. S*ubmitted to the R Journal March 2023 and awaiting a response.* www.jilllundell.com. 2023

Lundell JF, Bean B, and Symanzik J. Let's talk about the weather: a cluster-based approach to weather forecast accuracy. S*ubmitted to Computational Statistics April 2019 and accepted February 2023*. arXiv:2303.05484. https://doi.org/10.48550/arXiv.2303.05484. 2023

Publications in Preparation

Lundell JF, Street K. cytofQC: A package for customized data cleaning of CyTOF data.

Street K, Movassagh M, **Lundell JF**. VDJdive: A package for clonotype diversity in T-cell receptors.

Nonrefereed Publications

Haney T, **Lundell JF**. Historical data analysis supporting the data quality objectives for the INL site environmental soil monitoring program. INL/INT-15-37431 Revision No. 0. February 2021.v. https://doi.org/10.2172/1769951. 2021

Lundell JF. Am-241, Pu-238, Pu-239/240, and Sr-90 decision levels for the environmental air monitoring program for the Idaho National Laboratory. https://doi.org/10.2172/1875115. 2021

Lundell JF. Tuning hyperparameters in supervised learning models and applications of statistical learning in genome-wide association studies with emphasis on heritability. Unpublished Doctoral dissertation. Utah State University. August 2019. https://doi.org/10.26076/a37e-facb. 2019

Lundell JF, Bean B, Symanzik J. Let's talk about the weather. *JSM Proceedings*, Statistical Computing Section. Alexandria, VA: American Statistical Association. 2018. www.jilllundell.com. 2018

Lundell JF. There has to be an easier way: a simple alternative for parameter tuning of supervised learning methods. *JSM Proceedings*, Statistical Computing Section. Alexandria, VA: American Statistical Association. 2017. www.jilllundell.com. 2017

Lundell JF, Magnuson SO, Scherbinske P, Case MJ. Data quality objectives supporting the environmental direct radiation monitoring program for the Idaho National Laboratory. INL/EXT-15-34803. June 2015. https://doi.org/10.2172/1194017. 2015

Lundell JF. On the model selection in a frailty setting, Unpublished Master's Thesis. Utah State University Department of Mathematics and Statistics. August 1998. https://doi.org/10.26076/c6cf-ceeb. 1998

Invited Talks

Street K, Movassagh M, **Lundell JF.** Single-cell TCR-seq data analysis. Dana-Farber Cancer Institute Neovax Scientific Meeting, Boston, Massachusetts. November 2021. Virtual. 2021

Lundell JF. Industry, academic research, and education. I've done it all. Ask me anything about data science jobs. Data Science Club, University of Wisconsin Milwaukee, Milwaukee, Wisconsin. October 2021. Virtual. 2021

Lundell JF. How I got my job. Utah State University Mathematics and Statistics Department Professional Development Seminar, Logan, Utah. March 2021. Virtual. 2021

Lundell JF. Industry, academic research, and education. I've done it all. Ask me anything about data science jobs," Utah State University Data Science Club, Logan, Utah. March 2021. Virtual. 2021

JF Lundell. Better implementation of supervised learning models and how they can be used to analyze complex data. Job talk, Clemson School of Mathematical and Statistical Sciences, Clemson University, Clemson, South Carolina. January 2019. 2019

Lundell JF, Bean B. I hate group projects! A solution to group work angst using interdisciplinary consulting as a guide. TA Teaching Seminar, Utah State University Department of Mathematics and Statistics, Logan, Utah. November 2018. 2018

Lundell JF. Why statistics? BYU-Idaho Women in STEM Conference, Rexburg, Idaho. March 2018. 2018

Presentations

Lundell JF, Street K. Where did my cancer cells go? cytofQC for better CyTOF data cleaning. Bioconductor Conference, Seattle, Washington. July, 2022. 2022

Lundell JF, Street K. Where did my tumor cells go? A better way to clean CyTOF data using cytofQC. Current Topics in Bioinformatics Series, Harvard Bioinformatics Core, Boston, Massachusetts. December 2021. Virtual. 2021

Lundell JF, Irizarry R. Using Wavelets to Identify Differences in Methylation Data. Joint Statistical Meetings, Virtual. August 2020. 2020

Lundell JF, Bean B. I hate group projects! A solution to group work angst using interdisciplinary consulting as a guide. Together We Teach Conference, Logan, Utah. August 2018. 2018

Lundell JF, Bean B, Symanzik J. Let's talk about the weather. Joint Statistical Meetings, Vancouver, British Columbia. August 2018. 2018

Lundell JF. Which genes are really causing my problems? Filtering with LASSO and elastic net to find the signal in ultra-high dimensional data. The Conference on Statistical Learning and Data Science / Nonparametric Statistics, New York City, New York. June 2018. 2018

Lundell JF. There has to be an easier way: a simple alternative for parameter tuning of supervised learning methods. Joint Statistical Meetings, Baltimore, Maryland. August 2017. 2017

Lyons R, **Lundell JF**, Peralta R. Groundwater modeling of the Uinta Basin, Utah, as a boundary condition of the Birds Nest Aquifer. Utah State University Spring Run-Off Conference, Logan, Utah. March 2017. 2017

Lundell JF, Fu G. Analysis of ultra-high-dimensional polycystic ovary syndrome genome using DC-RR. Joint Statistical Meetings, Chicago, Illinois. August 2016. 2016

Lundell JF, Oates B. Why I should stick my nose in other people's business or why I should participate in all phases of the data life cycle. Radiobioassay and Radiochemical Measurements Conference, Knoxville, Tennessee. October 2014. 2014

Lundell JF, Oates B. How far can you drive with three flat tires and one good tire or why the data life cycle matters to me. Radiobioassay and Radiochemical Measurements Conference, Knoxville, Tennessee. October 2014. (Workshop) 2014

Lundell JF, LaCroix D, Oates B. Data quality assessment: what is it, why use it, and what's in it for me? EPA Quality Management Conference, San Antonio, Texas. May 2009. 2009

Teaching Experience

Harvard T.H. Chan School of Public Health

- BST 260 Introduction to Data Science 2022

University of Puerto Rico, IQ-Bio REU 2020-2022

- Ran replicathon for three summers with a group of postdocs and PhD students.
- Mentored undergraduate students from all over the world in learning R and completing projects with R.
- Taught students the importance and principles of replicability
- Replicathon was virtual 2020-2021, in Puerto Rico in 2022.

Utah State University

- STAT 1040 (formally STAT 201) Introduction to Statistics 2016-2017, 1998-2000
- STAT 5080 Sampling 2000
- MATH 1210 Calculus Techniques 1999
- STAT 2000 Statistics for Life Sciences 1998
- STAT 502 Intermediate Statistics 1998
- MATH 101 Algebra 1997

Brigham Young University Idaho

- MATH 325 Intermediate Statistics 2012-2014
- MATH 221X Introduction to Statistics (Business, Social Statistics, and Biostatistics) 2011-2014
- MATH 108 Math for the Real World 2011-2013

Portage, Inc.

- Introduction to Data Quality Assessment 2007
- Introduction to Sampling Design 2004

Idaho State University

- MATH 1153 Introduction to Statistics 2002-2003
- MATH 1108 Intermediate Algebra 2002-2003
- MATH 0015 Arithmetic and Pre-Algebra 2002

Committee Service

Organizer for R/Bioconductor meetup group in the Boston area 2022

Diversity, Inclusion, and Equity Committee. Department of Biostatistics, Harvard University 2020-2021

Benefits committee. Portage, Inc. Idaho Falls, Idaho 2014-2016

Wellness committee, Portage, Inc. Idaho Falls, Idaho 2014-2016

Committee for restructuring Introduction to Statistics courses. Department of Mathematics, BYU-Idaho 2011-2014

Committee for creating a statistics minor. Department of Mathematics, BYU-Idaho 2011-2013

Journal Reviewer

Journal of Statistics and Data Science Education 2021-2022
Stat 2021
American Journal of Epidemiology 2020-2021
Nature Communications 2020

Credentials

Accredited Professional Statistician™ (PSTAT) 2014-Present
American Statistical Association

Affiliations

American Statistical Association 2005-Present
Society of Industrial and Applied Mathematics 2016–2020
Western North American Region of The International Biometric Society 2016-2020

Websites

Github: https://github.com/jillbo1000
Google Scholar: https://scholar.google.com/citations?hl=en&user=ShrO3-AAAAAJ
LinkedIn: https://www.linkedin.com/in/jill-lundell-00894256
Professional: www.jilllundell.com