Anna Neufeld

Contact Information

Department of Mathematics and Statistics
Williams College
18 Hosxey St.
Williamstown, MA, 01267

acn2@williams.edu
https://anna-neufeld.github.io
781-392-4894

Employment

Assistant Professor of Statistics 2024 - Present
Department of Mathematics and Statistics
Williams College, Williamstown, MA

Postdoctoral researcher 2023-2024
Biostatistics Program (Public Health Sciences Division)
Fred Hutchinson Cancer Center, Seattle, WA
- Advisor: Dr. Jeffery Leek

Education

University of Washington, Seattle, Washington 2018-2023
PhD in Statistics
- Advisor: Dr. Daniela Witten
- Thesis: Addressing double dipping via selective inference and data thinning

Williams College, Williamstown, MA 2014-2018
BA in Mathematics (Highest Honors) and Computer Science. *Summa Cum Laude*
- Advisor: Dr. Brianna Heggeseth
- Thesis: Regression trees for longitudinal data

Teaching Experience

At Williams College
- STAT 202, Introduction to Statistical Modeling
  - Fall 2024, 2 sections, 68 total students.
  - Spring 2025, 1 section, 45 students.
- STAT 442, Statistical Learning and Data Mining
  - Spring 2025, 12 students.
- CSCI 374T, Machine Learning
  - Spring 2021, 10 students.
  - Co-instructed this tutorial with Professor Andrea Danyluk, while serving as a remote visiting lecturer in the Department of Computer Science.

At University of Washington
- Instructor of Record:
  - Stat 311: Elements of Statistical Methods (Summer 2019). Fully responsible for course of 60 students.
  - Stat 499C: Undergraduate Research (Autumn 2021-Spring 2023, 6 consecutive quarters). Managed the logistics, enrollment, and course credit associated with the Directed Reading Program. There were 5-12 undergraduate students enrolled for credit each quarter.

- Head teaching assistant:
  - Stat 311: Elements of Statistical Methods (Autumn 2019, Winter 2020).
  - Developed lab assignments, maintained course website, helped write assignments and exams, served as liaison between professor and other TAs. There were 180 total students in the course. Also led two lab sections of 30 undergraduates each.
- Teaching assistant:
  - Stat 570: Regression Methods for Independent Data (Autumn 2021)
  - Stat 527: Nonparametric Regression (Spring 2021)
  - Stat 311: Elements of Statistical Methods (Autumn 2018, Spring 2020)
  - Stat 423: Applied Regression and Analysis of Variance (Winter 2019)
  - CSE/Stat 416: Introduction to Machine Learning (Spring 2019)
  - Responsible for lab sections of around 30 undergraduates each. Also held office hours and graded assignments. Courses listed above range from undergraduate introductory courses (Stat 311 and Stat 416) to core courses for the statistics PhD program (Stat 570).

Preprints

Ameer Dharamshi, **Anna Neufeld**, Lucy L. Gao, Jacob Bien, and Daniela Witten (2024). Thinning a Wishart Random Matrix. https://arxiv.org/abs/2502.09957.

Anna Neufeld, Joshua Popp, Lucy L. Gao, Alexis Battle, and Daniela Witten (2023) Negative binomial count splitting for single-cell RNA sequencing data. https://arxiv.org/pdf/2307.12985.

Ameer Dharamshi, **Anna Neufeld**, Lucy L. Gao, Jacob Bien, and Daniela Witten (2024). Decomposing Gaussians with unknown covariance. *Revise and resubmit at Biometrika.* https://arxiv.org/pdf/2409.11497

Stephen Salerno, Kentaro Hoffman, Awan Afiaz, **Anna Neufeld**, Tyler H. McCormick, and Jeffrey T. Leek (2024) Sample Size Considerations for Post-Prediction Inference.

Publications

Colleen E. O'Connor, **Anna Neufeld**, Chelsea L. Fortin, Fredrik Johansson, Jonathan Mene, Sarah H. Saxton, Susana P. Simmonds, Irina Kopyeva, Nicole E. Gregorio, Cole A. DeForest, Daniela M. Witten, Kelly R. Stevens (2025) Bioprinted Platform for Parallelized Screening of Engineered Microtissues In Vivo. To appear in *Cell Stem Cell.*

Stephen Salerno, Jiacheng Miao, Awan Afiaz, Kentaro Hoffman, **Anna Neufeld**, Qiongshi Lu, Tyler H. McCormick, and Jeffrey T. Leek (2025) ipd: An R Package for Conducting Inference on Predicted Data. To appear in *Bioinformatics.*

Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, Daniela Witten, and Jacob Bien (2024) Discussion of "data fission: splitting a single data point". *The Journal of the American Statistical Association. Invited discussion.*

Ameer Dharamshi, **Anna Neufeld**, Keshav Motwani, Lucy L. Gao, Daniela Witten, and Jacob Bien (2024) Generalized data thinning using sufficient statistics. *Journal of the American Statistical Association. SLDS student paper award awarded to Ameer Dharamshi at JSM, 2024.*

Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, and Daniela Witten (2024) Data thinning for convolution-closed distributions. *Journal of Machine Learning Research.*

Anna Neufeld, Lucy L. Gao, Joshua Popp, Alexis Battle, and Daniela Witten (2022) Inference after latent variable estimation for single cell RNA-sequencing data. *Biostatistics. Winner of best student paper award at WNAR, 2022.*

Anna C. Neufeld, Lucy L. Gao, and Daniela M. Witten (2022) Tree-Values: selective inference for regression trees. *Journal of Machine Learning Research.*

Anna Neufeld and Daniela Witten (2021). Discussion of Breimans Two Cultures: From Two Cultures to One. *Observational Studies. Invited Discussion.*

Maxian, O*., **Neufeld, A.***, Talis, E. J.*, Childs, L. M., & Blackwood, J. C. (2017). Zika virus dynamics: when does sexual transmission matter?. *Epidemics.*
(** denotes equal contribution*)

Software

datathin: splitting a random variable into independent training and test components. R package available on github, with tutorials available on our website.

countsplit: splitting integer-valued count matrices. R package available on CRAN, with tutorials available on our website.

treevalues: selective inference for regression trees. R package available on github, with tutorials available on our website.

splinetree: longitudinal trees and forests using a spline projection method. R package available on CRAN. Tutorials available on our website.

Invited Presentations

- International Seminar on Selective Inference, March 2025, Virtual. *Invited discussant for a talk on Selective Inference for Randomized Regression Trees.*
- Statistics Empowering Data Science (SEEDS) Conference, January 2025, University of Southern California, Los Angeles, CA. *Data thinning to avoid double dipping.*
- Williams College Science Lunch, October 2024, Williamstown, MA. *What is double dipping and how can we avoid it?*
- Williams College Statistics Colloquium, September 2024, Williamstown, MA. *What is double dipping and how can we avoid it?*
- Joint Statistical Meetings, JASA Theory & Methods Invited Session, Portland, OR. *Discussion of "Data Fission".*
- ASA Section on Statistical Genetics and Genomics monthly webinar, April 2024, virtual.
- Banff International Research Station Workshop on Statistical Aspects of Trustworthy Machine Learning. February, 2024. Banff, AB. *Data thinning to avoid double dipping.* [slides].
- University of Washington Department of Genome Sciences and Computational Molecular Biology Program Combi seminar, January 2024, Seattle, WA. *Avoiding double dipping in the analysis of scRNA-seq data.*

- Channing Division Methods Meeting, Brigham and Women's Hospital and Harvard Medical School, December 2023, virtual. *Avoiding double dipping in the analysis of scRNA-seq data.*
- Williams College Statistics Colloquium, November 2023, Williamstown, MA. *Data thinning to avoid double dipping.* [slides].
- Reed College Mathematics Colloquium, October 2023, Portland, OR. *Data thinning to avoid double dipping*
- North American Machine Learning, Optimization, and Statistics Symposium (NAMOS), June 2023, Vancouver, BC. *Data thinning.*
- Private Brands Casual Science Seminar at Amazon, April 2023, virtual. *Data thinning.* [slides].
- Fred Hutchinson Cancer Center, group of Dr. Mike Wu, April 2023, virtual. *Data thinning.*
- Williams College Statistics Colloquium, October 2022, virtual. *Avoiding "double dipping" in the analysis of single cell RNA-sequencing data.*
- Joint Statistical Meetings, August 2022. Panelist for: *Leo Breiman's Two Cultures: Introspection, Debate, and Discussion 20 Years Later.*
- International Seminar on Selective Inference, June 2022, virtual. *Inference after latent variable estimation for single cell RNA-sequencing data.* [recording], [slides].
- Boston Children's Hospital Vascular Anomalies Group Meeting, June 2022, virtual. *Statistical issues when analyzing single cell RNA-sequencing data.*

Contributed Presentations

- Institute of Mathematical Statistics New Researcher's Conference, August 2024, Corvalis, OR. Speed talk on *data thinning.*
- Joint Statistical Meetings, topic contributed session on "Harnessing multiple genomic features to identify disease biomarkers, August 2024, Portland, OR. *Avoiding double dipping in the analysis of single-cell RNA sequencing data.*
- WNAR 2022 Student Paper Competition Session, June 20202, virtual. *Inference after latent variable estimation for single cell RNA-sequencing data.*

Mentoring

Williams College senior theses

- Matthew Wu, 2025

NSF INSPIRE U2 REU, Spelman College Summer 2023

(Increasing Statistical Preparation in Research Education for Underrepresented Undergraduates)

- Virtual RStudio mentor for two undergraduate researchers; Allyanna Lewis and Sasha Villefranche. Both students will be presenting at the Annual Biomedical Research Conference for Minoritized Scientists (ABRCMS).

University of Washington Statistics Directed Reading Program (DRP)

- Co-founded a program that pairs undergraduates with PhD student mentors for independent studies. Modeled after successful Directed Reading Programs (DRPs) in mathematics departments at several universities. More information can be found at our website.
- Served as the graduate student coordinator from 2020-2023. Managed admissions, recruitment, and scheduling for 95 undergraduate independent studies.

Several undergraduates have gone on to graduate school in statistics or related fields.

- Served as the instructor of record for Stat 499 (the associated course that undergraduate participants can register for) from 2021-2023.
- Served as a mentor for the following eight undergraduates:
  - Winter 2020, Christina Nick, Statistical Natural Language Processing
  - Spring 2020, Rachael Ren, Infectious Disease Modeling with Differential Equations. See project writeup and presentation. Now a PhD student in statistics at the University of Texas.
  - Autumn 2020, Harper Zhu, Infectious Disease Modeling on a Network. See presentation and shiny app. Now a MS studnent in data science at Cornell Tech.
  - Spring 2021, Kayla Kenyon, Infectious Disease Modeling. See presentation.
  - Autumn 2021, Cathy Qi, Multiple Testing.
  - Winter 2022, Hisham Bhatti, Statistical Concepts in Clinical Trials.
  - Spring 2022, Wei Jun Tan [presentation] and Iris Zhao [presentation], Introduction to Computational Biology, co-mentored with Alan Min.

Professional Service

Manuscript reviewer: *Annals of Statistics*, *Journal of the Royal Statistical Society: Series B*, *Journal of the American Statistical Association*, *Annals of Applied Statistics*, *Journal of Machine Learning Research*, *Cell Systems*, *Statistical Science*, *Journal of Computational and Graphical Statistics*, *Japanese Journal of Statistics and Data Science*, *Electronic Journal of Statistics*, *Biometrics*, *Bernoulli*, and *Genetic Epidemiology*.

Judge: Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) Undergraduate Statistics Class Project Competition (USCLAP): Spring 2023 and Fall 2024.

Invited panelist:

- Panel on navigating the academic job market, hosted by the Institute for Mathematical Statistics (IMS) New Researchers Group, November 2024, virtual.
- Panel on statistics graduate school, hosted by the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) at the Electronic Undergraduate Statistics Research Conference, November 2022, virtual.

Department or University Service

Williams College

- Statistics hiring committee, Fall 2024.

University of Washington

- Lead consultant and volunteer recruitment coordinator for UW StatCom (statistics in the community non-profit consulting program), 2022-2023.
- Member of UW Statistics Undergraduate Curriculum Committee, 2021-2022.
- UW Statistics Graduate Student Representative (elected and paid position that involves attending faculty meetings, serving on committees, and planning orientation for new students). 2020-2021.
- Student interviewer for tenure track assistant professor search (2020) and teaching professor search (2021).
- First round reviewer for PhD Admissions, 2020-2023.

- UW Statistics Diversity, Inclusion, Community, and Equity (DICE) Committee, 2019-2023.
- Chair of the UW Statistics Fun Committee, 2019-2020.
- Organizer and Founder: Statistics Education Reading Group, 2019-2020.

Outreach

- Lesson leader at UW Math Day, March, 2023.
- Lesson leader at UW GEAR UP's (Gaining Early Awareness and Readiness of Undergraduate Education) after school program (May 2021) and statistics summer camp (July 2022).