



McGill

Department of
Epidemiology, Biostatistics
and Occupational Health

Biostatistics Seminars Winter 2021



Daniela M. Witten, PhD

Professor of Statistics and Biostatistics at
University of Washington, and the Dorothy Gilford
Endowed Chair in Mathematical Statistics

Valid Hypothesis Testing after Hierarchical Clustering

Wednesday, January 20, 2021
3:30 pm – 4:30 pm – [Zoom Link](#)

Abstract: As datasets continue to grow in size, in many settings the focus of data collection has shifted away from testing pre-specified hypotheses, and towards hypothesis generation. Researchers are often interested in performing an exploratory data analysis in order to generate hypotheses, and then testing those hypotheses on the same data; I will refer to this as 'double dipping'. Unfortunately, double dipping can lead to highly-inflated Type 1 errors. In this talk, I will consider the special case of hierarchical clustering. First, I will show that sample-splitting does not solve the 'double dipping' problem for clustering. Then, I will propose a test for a difference in means between estimated clusters that accounts for the cluster estimation process, using a selective inference framework. I will also show an application of this approach to single-cell RNA-sequencing data. This is joint work with Lucy Gao (University of Waterloo) and Jacob Bien (University of Southern California).

Bio: Daniela Witten is a professor of Statistics and Biostatistics at University of Washington, and the Dorothy Gilford Endowed Chair in Mathematical Statistics. She develops statistical machine learning methods for high-dimensional data, with a focus on unsupervised learning. Daniela is the recipient of an NIH Director's Early Independence Award, a Sloan Research Fellowship, an NSF CAREER Award, a Simons Investigator Award in Mathematical Modeling of Living Systems, a David Byar Award, a Gertrude Cox Scholarship, and an NDSEG Research Fellowship. She is also the recipient of the Spiegelman Award from the American Public Health Association for a statistician under age 40 who has made outstanding contributions to statistics for public health, as well as the Leo Breiman Award for contributions to the field of statistical machine learning. She is a Fellow of the American Statistical Association, and an Elected Member of the International Statistical Institute. Daniela's work has been featured in the popular media: among other forums, in Forbes Magazine (three times) and Elle Magazine, on KUOW radio (Seattle's local NPR affiliate station), in a NOVA documentary, and as a PopTech Science Fellow. Daniela is a co-author (with Gareth James, Trevor Hastie, and Rob Tibshirani) of the very popular textbook "Introduction to Statistical Learning". She was a member of the National Academy of Medicine (formerly the Institute of Medicine) committee that released the report "Evolution of Translational Omics". Daniela completed a BS in Math and Biology with Honors and Distinction at Stanford University in 2005, and a PhD in Statistics at Stanford University in 2010. For more info, please visit: <https://www.danielawitten.com/>