

The Hong Kong University of Science and Technology

Seminar on Business Data Science

Department of ISOM

**Dissecting Double Dipping in
Statistical Tests After Clustering**

by

Prof. Jingyi Jessica LI

University of California, Los Angeles

Date: 1 November 2024 (Friday)

Time: 11:00am – 12:00noon

Venue: 4/F Meeting Room (LSK Business Building)

Abstract

Motivated by the widespread use of clustering followed by statistical testing in single-cell and spatial omics data analysis, this talk will address the issue of double dipping. We aim to explore whether double dipping is a significant concern and investigate how various data-splitting and data-simulation strategies can mitigate its impact on inflated false discovery rates (FDR). We will also discuss different perspectives on whether the inference should be conditional on the clustering step or not. In particular, we will highlight the influence of feature correlations on FDR inflation. Through simulation and real-data examples, we will demonstrate how our simulation-based strategy for correcting double dipping can lead to more reliable and insightful discoveries.

Bio

Jingyi Jessica Li, Professor of Statistics and Data Science (also affiliated with Biostatistics, Computational Medicine, and Human Genetics), leads a research group titled the Junction of Statistics and Biology at UCLA. With Ph.D. from UC Berkeley and B.S. from Tsinghua University, Dr. Li focuses on developing interpretable statistical methods for biomedical data. Her research delves into quantifying the central dogma, extracting hidden information from transcriptomics data, and ensuring statistical rigor in data analysis by employing synthetic negative controls. Recipient of multiple awards including the NSF CAREER Award, Sloan Research Fellowship, ISCB Overton Prize, and COPSS Emerging Leaders Award, her contributions have gained recognition in the fields of computational biology and statistics.

**All interested are welcome!
Enquiries: Dept of ISOM**