

Dissecting Double Dipping in Statistical Tests After Clustering: Mean vs. Variance



Jessica Li, PhD

Professor of Statistics and Data Science at UCLA

Motivated by the widespread use of clustering followed by statistical testing in single-cell and spatial omics data analysis, I will discuss the issue of double dipping in this talk. Our objective is to explore whether double dipping is a significant concern and to investigate if various data-splitting strategies can mitigate this issue. We employ a hierarchical model suited for single-cell and spatial omics data to analyze the double dipping issue. Specifically, we will discuss the implications of double dipping in the context of mean versus variance, addressing whether it arises from distortion of the mean (true signal) or is a consequence of data variance (unwanted variation). We will examine different perspectives, considering whether our inferences should be conditional on the data or pertain to the population from which the data are drawn. Using real-data examples, we will demonstrate how our methods for correcting double dipping can yield more reliable and insightful discoveries.

Thursday, September 19, 2024, 3:30-4:30PM Eastern

133 Rosenau Hall

Zoom Link:

<https://unc.zoom.us/j/93457075747?pwd=hsg9eT4nlw9b6CIAVwoiL4z6ZcXYOa.1>

Meeting ID: 934 5707 5747

Passcode: 135152