NOVEL APPROACHES TO CREATING SYNTHETIC DATA FROM MULTIVARIATE SURVEY DATA FOR STATISTICAL DISCLOSURE CONTROL

ABSTRACT:

Statistical disclosure is the unauthorized identification of individuals, organizations, or key items through the release of individual-level data or summary statistics. The risks that follow disclosure may cause significant damage when the information is used for malicious purposes, resulting in monetary, emotional, or reputational harm. Data syntheses of partially or fully-synthetic datasets can help guard against disclosure risks when sharing or releasing data. In this dissertation, we propose two new methods to guard against statistical disclosure through the creation of fully-synthetic data to release in place of the originals.