
Approches modernes à l'imputation multiple pour les problèmes de données manquantes

Responsable et président: Russell Steele (McGill University)

BEN GOODRICH, Columbia University

Évaluation de l'hypothèse de « données manquantes au hasard » après l'imputation multiple de données manquantes

La plupart des algorithmes d'imputation multiple de valeurs manquantes reposent sur l'hypothèse de « données manquantes au hasard » (MAH), qui affirme grosso modo que la probabilité qu'une donnée soit manquante est conditionnellement indépendante des valeurs de celle-ci. Par conséquent, plusieurs chercheurs croient que les données complétées résultantes ne peuvent pas falsifier l'hypothèse MAH. Cependant, la présence de données observées parmi les données complétées peut constituer un point d'appui. Nous étudions des manières de falsifier l'hypothèse MAH en appliquant des modèles de sélection et une analyse de la structure de covariance aux données complétées. L'efficacité de ces techniques est évaluée à l'aide de simulations.

NATHANIEL SCHENKER, National Center for Health Statistics and Centers for Disease Control and Prevention

Diverses utilisations de l'imputation multiple au National Center for Health Statistics des États-Unis

Cet exposé décrit trois applications récentes ou potentielles de l'imputation multiple au National Center for Health Statistics des États-Unis visant à illustrer différents types de problèmes pour lesquels cette technique peut être employée. L'une des applications concerne des données manquantes sur des images de scanner corporel du National Health and Nutrition Examination Survey (NHANES). La seconde cherche à améliorer les analyses des données autodéclarées du National Health Interview Survey par l'imputation de valeurs cliniques en utilisant des modèles ajustés au plus petit NHANES. La troisième utilise les poids observés à la naissance et des modèles de mélange pour identifier les âges gestationnels incertains dans les données de natalité aux États-Unis et cherche à imputer des âges gestationnels plus plausibles.

YAJUAN SI, Duke University

Imputation multiple dans des études de panel avec érosion et échantillons ajoutés

Les données de panel permettent une meilleure compréhension du comportement de masse au fil du temps à l'aide de vagues de suivi multiples. Cependant, les études de panel souffrent d'érosion, ce qui diminue la taille effective de l'échantillon et peut entraîner un biais dans les estimations si le désistement est systématiquement lié au résultat d'intérêt. L'échantillon ajouté en tant que source externe de données comprend de nouveaux répondants sélectionnés aléatoirement à qui le même questionnaire est administré conjointement avec une vague subséquente du panel. Ces échantillons ajoutés sont employés pour évaluer les effets de l'érosion du panel et pour corriger les biais au moyen de l'imputation multiple.