KAREN A KOPCIUK, Cancer Epidemiology and Prevention Research, AHS/University of Calgary

Distribution-Based Imputation for Left-Censored Metabolomics Data

Mass spectrometry instruments used to measure metabolomics features have left limits of detection that can result in substantial missing data. Imputing a constant value results in both bias and reduced variability in their distributions while dropping them altogether could distort relationships with other metabolites. Since estimation in projection-based methods such as principal component and partial least squares regression models is based on variation, misleading results are likely. A distribution-based imputation method is proposed to recover the missing data values and compared with imputed constant values using simulated data. Performance is evaluated using the area under the ROC curve.