

1135-VT-138

Jenkin Tsui* (jenkin.tsui@utoronto.ca), Dept. of Computer and Mathematical Sciences, 1065 Military Trail, Toronto, Ontario M1C1A4, Canada, and **Abel Dasyuva** and **Kenneth Chu**.

Optimal Estimating Equation for Logistic Regression with Linked Data.

Record Linkage is the act of bringing together records from two files, say file X and file Y, that relate to the same individual or entity. A record is unlinked if it has no link to any outgoing record in the other file. A link between two unmatched records is called a false positive. Chipperfield et al. (2011) considered the situation where logistic regression is applied, using linked pairs and a sample of such pairs, which are each known to be matched or unmatched through clerical reviews. Their proposed solution is inspired by the maximum-likelihood framework incorporating separate features for false positives and unlinked records. In the maximum-likelihood framework, under general regularity conditions, the Fisher information is equal to the variance-covariance matrix of the score function. This key property leads to the asymptotic efficiency of maximum likelihood estimators through the Cramer-Rao bound. In what follows, we show that this property is not satisfied by Chipperfield's score function. Then we refine the related estimator to decrease its variance in large samples by proposing an optimal estimating equation for logistic regression with linked data while accounting for false positives, built upon the Quasi-likelihood framework. (Received September 27, 2017)