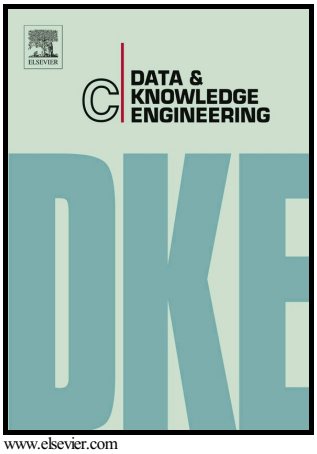
Author's Accepted Manuscript



www.elsevier.com

Disclosure Risk Reduction for Generalized Linear Model Output in a Remote Analysis System

Atikur R. Khan, Christine M. O'Keefe

PII: S0169-023X(16)30148-3
DOI: http://dx.doi.org/10.1016/j.datak.2017.07.009
Reference: DATAK1606

To appear in: *Data & Knowledge Engineering*

Received date: 13 August 2016
Revised date: 16 June 2017
Accepted date: 25 July 2017

Cite this article as: Atikur R. Khan and Christine M. O'Keefe, Disclosure Risk Reduction for Generalized Linear Model Output in a Remote Analysis System *Data & Knowledge Engineering,* http://dx.doi.org/10.1016/j.datak.2017.07.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Disclosure Risk Reduction for Generalized Linear Model Output in a Remote Analysis System

Atikur R. Khan, Christine M. O'Keefe*

CSIRO, Australia

Abstract

Remote analysis systems allow analysts to obtain statistical results without providing direct access to confidential data stored in a secure server system. An attacking analyst could send queries to a remote server to obtain outputs of statistical analyses and use those outputs for a disclosure attack. Statistical disclosure control (SDC) methods are used to modify remote analysis system (RAS) outputs in the protection of confidential information. Confidentiality protection through perturbation is one of the most commonly adopted SDC methods. In the case of generalized linear modelling, random noise is added to the estimated coefficients or to the associated estimating equation prior to getting estimates. This inflates the variances of estimators, and some efficiency and utility of estimators are lost. Thus the application of any perturbation based SDC method could result in an inefficient estimator, with the danger of producing worthless inferences. To date, little attention has been given to systematically controlling the disclosure risk and utility in SDC methods for RAS. In this paper, we develop a framework for the perturbation of estimating equations that enables an RAS to release modified generalized linear model output in such a way that the disclosure risk is not only reduced but also a good utility is maintained. Finally, we present some empirical results demonstrating the application of our framework for obtaining estimates from perturbed estimating

Accepted manuscript

*Corresponding author: Christine M. O'Keefe, Commonwealth Scientific and Industrial Research Organisation (CSIRO), GPO Box 664, Canberra, ACT 2601, Australia

Email addresses: Atikur.Khan@csiro.au (Atikur R. Khan), Christine.Okeefe@csiro.au (Christine M. O'Keefe)

Download English Version:

https://daneshyari.com/en/article/4942423

Download Persian Version:

https://daneshyari.com/article/4942423

Daneshyari.com