

**Institute for Mathematics and its Applications**  
**2011 Seminar Series: 8**  
University of Wollongong

**Title:** Comparison of Remote Analysis with Statistical Disclosure Control for Protecting the Confidentiality of Business Data

**Speaker:** Christine O'Keefe

**Time and Date:** 3:30pm, Thursday, July 28, 2011

**Location:** Room 24.104

**Abstract:** This talk addresses the challenge of allowing statistical analysis of confidential business data while maintaining confidentiality. The most widely-used approach to date is statistical disclosure control, which involves modifying or confidentialising data before releasing it to users. Newer proposed approaches include the release of multiply imputed synthetic data in place of the original data, and the use of a remote analysis system enabling users to submit statistical queries and receive output without direct access to data. Most implementations of statistical disclosure control methods to date involve census or survey microdata on individual persons, because existing methods are generally acknowledged to provide inadequate confidentiality protection to business (or enterprise) data.

In this talk we compare the statistical disclosure control approach with the remote analysis approach, in the context of protecting the confidentiality of business data in statistical analysis. We provide an example which enables a side-by-side comparison of the outputs of exploratory data analysis and linear regression analysis conducted on a sample business dataset under these two approaches, and provide traditional unconfidentialised results as a standard for comparison. Our example supports the conclusion that a remote analysis server may provide more accurate exploratory data analysis and regression results than traditional statistical disclosure control, provided the analyst understands the output confidentialisation methods and their potential impact.