



UNIVERSITÀ DEGLI STUDI DI FIRENZE  
DIPARTIMENTO DI STATISTICA  
"Giuseppe Parenti"

Corso presso il Dipartimento di Statistica "G. Parenti" dell'Università di Firenze  
12 – 13 Febbraio 2008

su

## **Approcci moderni all'analisi robusta di dati multidimensionali**

**Anthony C. Atkinson**

London School of Economics, Londra

**Andrea Cerioli**

Sezione di Statistica e Informatica, Dipartimento di Economia, Università di Parma

**Marco Riani**

Sezione di Statistica e Informatica, Dipartimento di Economia, Università di Parma

### **Abstract**

L'obiettivo del ciclo di seminari è quello di rivisitare le tradizionali tecniche statistiche per l'analisi di dati multidimensionali ed autocorrelati alla luce di alcune tendenze innovative nella letteratura internazionale, che pongono l'accento sull'importanza della proprietà di robustezza dei metodi, cioè sulla loro capacità di fornire risultati stabili e scientificamente attendibili anche in presenza di valori anomali. I metodi tradizionali (*deletion diagnostics*) cercano di identificare i valori anomali operando in maniera "backwards" valutando l'anomalia o l'influenza d'una osservazione su un modello statistico solo dopo che esso è stato adattato all'intero data set. Così facendo, però, i metodi tradizionali soffrono d'un grave inconveniente, l'*effetto di mascheramento*, che si manifesta in presenza di gruppi di outliers e che rende l'influenza individuale di ciascuno di essi molto limitata e dunque non identificabile. Inoltre, in presenza di outlier multipli c'è un aumento esponenziale del numero di modelli da considerare.

Un'alternativa ai metodi *backward* è costituita dall'utilizzo di stimatori robusti (LMS, MAD, TRIMMED MEAN). Generalmente tali metodi permettono di rilevare gli outlier, però presentano l'inconveniente di sottopesare o tralasciare alcune osservazioni. Inoltre, essi possono fallire completamente se le osservazioni non provengono da un'unica popolazione, ma da più popolazioni distinte.

Lo scopo dei seminari è quello di presentare una nuova metodologia statistica nota con il nome di *forward search* (Atkinson and Riani 2000, Atkinson, Riani and Cerioli, 2004) che combina la robustezza e l'efficienza e consente di cogliere l'effetto inferenziale delle diverse unità statistiche sul modello proposto. Attraverso una combinazione efficace di modellazione statistica e di grafici diagnostici, la *forward search* costituisce uno strumento potente al fine di individuare la presenza di valori anomali, sia individuali che raggruppati, e di valutare il loro effetto sui risultati delle analisi tradizionali.

Il primo seminario avrà per oggetto l'applicazione della *forward search* ai modelli di regressione lineare multipla ed ai modelli lineari generalizzati. L'obiettivo di questo seminario è



UNIVERSITÀ DEGLI STUDI DI FIRENZE  
DIPARTIMENTO DI STATISTICA  
“Giuseppe Parenti”

quello di introdurre la nuova metodologia, inquadrarla nel *framework* degli stimatori esistenti e di metterne in luce i vantaggi rispetto alle procedure diagnostiche tradizionali.

Il secondo seminario avrà per oggetto l'applicazione della *forward search* ai modelli per dati multivariati (componenti principali, analisi discriminante, trasformazioni multivariate, cluster).

## Struttura dei seminari

Mattino: lezione frontale

Pomeriggio: analisi dei dati tramite SPLUS ed R

## Bibliografia

Atkinson A.C. and Riani M. (2000), *Robust Diagnostic Regression Analysis*, Springer Verlag, New York.

Atkinson A.C., Riani, M. and Cerioli A. (2004), *Exploring Multivariate Data With the Forward search*, Springer Verlag, New York.