

## OBIETTIVO DEL CORSO

Il corso si propone di fornire ai partecipanti un'introduzione ai metodi per l'analisi dei dati multivariata attraverso l'impiego di Stata. Grazie all'enorme quantità di dati ormai disponibile in ogni settore industriale e commerciale, le tecniche di analisi statistica multivariata ricoprono oggi più che mai un ruolo fondamentale per l'estrazione di utili informazioni dai dati stessi. Durante il corso saranno illustrate le principali metodologie di analisi multivariata (analisi dei cluster, analisi delle componenti principali, analisi fattoriale) attraverso esempi e casi concreti.

## REQUISITI RICHIESTI:

Conoscenze di base di Statistica e del Software Stata.

## DESTINATARI

Il corso è di interesse per ricercatori e analisti che si occupano di economia, finanza, sanità pubblica e di scienze sociali, e che desiderino condurre ricerche empiriche utilizzando dati multivariati.

## PROGRAMMA

### PARTE I: I DATI MULTIVARIATI: PRIMI INDICATORI DI SINTESI E RAPPRESENTAZIONI GRAFICHE

1. Tipi di variabili e il problema dei dati mancanti
2. Covarianze, correlazioni e misure di distanza
3. La distribuzione normale multivariata
4. Grafici per la visualizzazione di dati multivariati

### PARTE II: ANALISI DELLE COMPONENTI PRINCIPALI E ANALISI FATTORIALE

1. Analisi delle componenti principali
  - Introduzione
  - Calcolo delle componenti principali
  - Calcolo degli *scores* delle componenti principali
2. Analisi dei fattori
  - Introduzione
  - Stima dei fattori
  - Scelta del numero di fattori
  - Rotazione dei fattori

### PARTE III – ANALISI DEI CLUSTER

1. Introduzione agli algoritmi di *clustering*
2. Principali metodi agglomerativi di *clustering*
  - Il dendrogramma
  - *Single linkage*
  - *Complete linkage*
  - *Average linkage*
  - Metodo di *Ward*
3. Principali metodi divisivi di *clustering*
  - *K-means*
4. Profilazione dei cluster

### PARTE IV – ALTRI METODI DI ANALISI DEI DATI MULTIVARIATI

1. Estensioni degli approcci presentati
2. Altre tecniche di analisi multivariata (analisi delle corrispondenze, *scaling* multidimensionale, etc.)

## SVOLGIMENTO

Il corso si articola in quattro parti. La prima parte si concentrerà su come sintetizzare le informazioni contenute in un set di dati multivariati. Particolare enfasi sarà posta sulla rappresentazione grafica dei dati, sull'individuazione di *outlier* e sulla gestione dei dati mancanti.

La seconda parte si focalizzerà su due tecniche multivariate per la riduzione della complessità di un data set: l'analisi delle componenti principali e l'analisi fattoriale. Dopo averne illustrato gli elementi fondanti e il loro utilizzo operativo, la presentazione si concentrerà sulla disamina delle differenze tra queste due metodologie, spesso confuse nella pratica.

La terza parte sarà dedicata all'analisi dei cluster: dopo aver presentato i principali algoritmi di *clustering*, ci si concentrerà sulla profilazione dei cluster ottenuti.

Infine nella quarta parte verrà fatto un cenno ad altre importanti tecniche per l'analisi multivariata quali l'analisi delle corrispondenze e il *multidimensional scaling*.

Le lezioni saranno di tipo interattivo e avranno contenuto prevalentemente applicato, sperimentando di volta in volta le tecniche apprese su dati reali.

**MATERIALI:** I materiali del corso includono i lucidi con la parte teorica, i *do-file* e le banche dati per l'implementazione di tutte le applicazioni empiriche. Questo consentirà ad ogni partecipante di esercitarsi sui contenuti del corso, eseguendo autonomamente i *file* distribuiti.

## TESTI E REFERENZE UTILI

- Anderson, T. W. (2003), An Introduction to Multivariate Statistical Analysis, 3rd edition, Wiley.
- Everitt, B. S., Landau, S., Leese, M. and Stahl, D. (2011), Cluster Analysis, 5th edition, Wiley.
- Flury, B. (1997), A First Course in Multivariate Statistics, Springer
- Hastie, T., Tibshirani, R. and Friedman, J. (2009), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, Springer
- Izenman, A. J. (2008), Modern Multivariate Statistical Techniques, Springer.
- Johnson, R. A. and Wichern, D. W. (2007), Applied Multivariate Statistical Analysis, 6th edition, Pearson/Prentice Hall.
- Mardia, K. V., Kent, J. T. and Bibby J. M. (1980), Multivariate Analysis, Academic Press.
- Timm, N. H. (2002), Applied Multivariate Analysis, Springer

## DATA E ISCRIZIONE

Il corso è previsto a Milano nei giorni 10-11 Dicembre 2015, la durata è di 13 ore distribuite in due giornate con inizio delle lezioni alle ore 9.00 e termine alle ore 17.00.

La quota di iscrizione è di Euro 1095,00 + IVA ed include le due pause caffè giornaliere ed il light lunch, oltre al necessario materiale didattico.

L'iscrizione al corso dovrà avvenire tramite lo specifico modulo di registrazione e pervenire a TStat S.r.l. almeno 15 giorni prima dell'inizio del corso stesso. E' possibile scaricare il modulo di registrazione dal nostro sito [www.tstat.it](http://www.tstat.it) oppure può essere richiesto alla segreteria organizzativa TStat a [corsi@tstat.it](mailto:corsi@tstat.it).



PER ULTERIORI INFORMAZIONI RIVOLGERSI A:

TStat S.r.l. - Via Rettangolo, 12-14 - 67039 Sulmona – AQ - Tel. 0864 210101 - Fax 0864 206014

Email: [corsi@tstat.it](mailto:corsi@tstat.it) - Web: <http://www.tstat.it>