

# DATA SCIENCE PER LE SCIENZE UMANE E SOCIALI (LM81)

(Università degli Studi)

## Insegnamento Modelli multidimensionali per l'analisi dei dati

GenCod A007246

Docente titolare Enrico CIAVOLINO

**Insegnamento** Modelli multidimensionali per l'analisi dei dati

**Insegnamento in inglese** Multidimensional models for data

**Settore disciplinare** M-PSI/03

**Corso di studi di riferimento** DATA SCIENZE PER LE SCIENZE UMANE E

**Tipo corso di studi** Laurea Magistrale

**Crediti** 8.0

**Ripartizione oraria** Ore Attività frontale: 48.0

**Per immatricolati nel** 2024/2025

**Erogato nel** 2024/2025

**Anno di corso** 1

**Lingua** ITALIANO

**Percorso** PERCORSI COMUNE/GENERICO

**Sede**

**Periodo** Primo Semestre

**Tipo esame** Orale

**Valutazione** Voto Finale

**Orario dell'insegnamento**

<https://easyroom.unisalento.it/Orario>

### BREVE DESCRIZIONE DEL CORSO

Il corso presenta le tecniche statistiche di analisi multidimensionale per dati quantitativi (ACP) e qualitativi (CA) nonché la Cluster Analysis per lo studio di dati sperimentali in ambito psicologico e sociale. Il modello di Analisi della Varianza completerà la parte teorica, affrontando gli aspetti parametrici relativi agli esperimenti. Gli argomenti verranno affrontati sia dal punto di vista teorico che applicativo mediante l'uso del software open source R.

### PREREQUISITI

Nozioni di matematica e statistica di base

---

## OBIETTIVI FORMATIVI

Il corso, pur mantenendo un'impostazione metodologica di fondo rivolta allo studio dell'analisi multidimensionale, si propone di avvalersi del supporto di software statistici dedicati, in modo da accompagnare all'apparato teorico un consistente risvolto pratico, sia dal punto di vista della risoluzione di casi studio reali o simulati, sia nell'acquisizione di familiarità con i software per l'elaborazione dei dati.

Il corso intende fornire agli studenti i principali strumenti statistici per l'analisi di fenomeni reali di tipo psicologico e sociale.

### **Conoscenze e comprensione.**

Acquisire una conoscenza dei metodi e delle tecniche quantitative ad un livello post secondario mediante l'utilizzo di testi di riferimento, materiale didattico e software opensource. Saranno illustrati temi di avanguardia, come tecniche di data science per le scienze umane e sociali, inoltre verranno trasmesse nozioni che consentiranno lo sviluppo di analisi quantitative in contesti originali.

### **Capacità di applicare conoscenze e comprensione.**

Applicare e discutere metodi quantitativi avanzati nell'ambito dell'analisi dei dati sperimentali, in modo da risolvere problemi in ambiti nuovi o non familiari.

### **Autonomia di giudizio.**

Interpretare i dati in modo da formulare conclusioni autonome nell'ambito dell'analisi quantitativa multidimensionale.

### **Abilità comunicative.**

Comunicare metodi e tecniche multidimensionale per la risoluzione di problemi quali-quantitativi a interlocutori specialisti e non specialisti.

### **Capacità di apprendimento.**

Aver sviluppato le competenze necessarie per intraprendere studi successivi, oppure studi autonomi.

---

## METODI DIDATTICI

Le lezioni tradizionali di tipo frontale verranno accompagnate da esercitazioni svolte nel laboratorio informatico, in modo da rendere autonomi gli studenti nell'utilizzo del software statistico R per l'analisi dei modelli statistici.

---

## MODALITA' D'ESAME

Gli **esami** si terranno **in presenza** (salvo diverse disposizioni) con le seguenti modalità:

La valutazione avverrà attraverso la somministrazione di un **Test di 20 domande** a risposta multipla con quattro modalità:

- 10 domande avranno un punteggio pari a 2.
- 10 domande avranno un punteggio pari ad 1.
- Il Test verrà integrato con una verifica orale su una o due domande a risposta multipla.

**Il prerequisito per accedere al test** è lo sviluppo di un report di ricerca (max 10 pages) in cui siano analizzati dati reali e che mostri l'apprendimento delle tecniche quantitative di analisi dei dati (ACP, AC, Cluster, AnOVA) e l'uso del software R. Dal sito: <https://formazioneonline.unisalento.it> è possibile scaricare il materiale didattico, il Template per il report e il link a dataset da utilizzare. L'elaborato può essere sviluppato in gruppo (massimo 5 persone) e singolarmente. **Il report deve essere consegnato 5 giorni prima dell'esame**

**Di seguito le modalità nel caso in cui dovesse riprendere la modalità online:**

**Il link al TEAM dell'appello verrà comunicato sulla pagina ufficiale del docente** il giorno dopo la chiusura delle iscrizioni all'esame. Il **link del TEST** sarà attivato sulla piattaforma <https://formazioneonline.unisalento.it> nella pagina di riferimento dell'insegnamento, cioè dove viene reso disponibile il materiale didattico. La compilazione delle risposte avverrà online. Si prega inoltre di verificare il **corretto funzionamento della webcam e del microfono**.

Gli studenti dovranno seguire la procedura indicata dalla guida per lo studente.

---

## ALTRE INFORMAZIONI UTILI

Gli studenti devono prenotarsi per l'esame finale esclusivamente utilizzando le modalità previste dal sistema VOL.

---

## PROGRAMMA ESTESO

### ▪ Elementi essenziali di algebra lineare

▪ Definizione di Matrici e Vettori. Somma e Prodotto Interno tra matrici. Alcune proprietà delle matrici. Matrici di Codevarianza e Covarianza. Standardizzazione e Matrice di Correlazione.

### ▪ Analisi in Componenti Principali (ACP)

▪ Definizione e obiettivi dell'ACP. Rappresentazione dei dati nello spazio degli individui: Ricerca della prima e seconda componente principale. Rappresentazione dei dati nello spazio delle variabili: Ricerca della prima e seconda componente principale. Scelta del numero delle componenti da estrarre. Rappresentazioni grafiche. Proprietà.

### ▪ Analisi delle Corrispondenze (AC)

▪ Introduzione e cenni storici. Matrice dei profili riga. Matrice dei profili colonna. Spazio delle righe. Spazio delle colonne. Formule di transizione. Contributo assoluto CA e contributo relativo CR. Rappresentazioni grafiche. Proprietà.

### ▪ Cluster Analysis

▪ Introduzione e cenni storici. Indici di dissimilarità. Distanza Euclidea e di Manhattan. Indici di similarità. Indice di Zubin e di Jaccard. Algoritmi di CA gerarchici. Legami tra unità.

### ▪ Analisi della Varianza

▪ Introduzione alla statistica inferenziale, Regressione, Anova ad una via.

Gli aspetti teorici vanno integrati con l'uso del software R. Dal sito: <https://formazioneonline.unisalento.it> è possibile scaricare il materiale didattico, il Template per il report e il link a dataset da utilizzare.

---

## TESTI DI RIFERIMENTO

Alla pagina <https://formazioneonline.unisalento.it> sono rese disponibili dispense, data set ed esercizi di apprendimento.