

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

(Università degli Studi)

## Insegnamento Statistica e matematica per la data science - Mod. II

GenCod A007244

Docente titolare SERENA ARIMA

**Insegnamento** Statistica e matematica per la data science - Mod. II

**Insegnamento in inglese** Statistics and mathematics for data science - Mod. II

**Settore disciplinare** SECS-S/02

**Corso di studi di riferimento** DATA SCIENCE 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**

**Percorso** PERCORSI COMUNE/GENERICO

**Sede**

**Periodo**

**Tipo esame**

**Valutazione**

**Orario dell'insegnamento**

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

### BREVE DESCRIZIONE DEL CORSO

Il corso introduce gli studenti agli strumenti fondamentali per l'analisi statistica e matematica relativi alla data science. Dopo aver richiamato le nozioni di base, si approfondiranno i principali metodi inferenziali basati sull'approccio di verosimiglianza e gli strumenti matematici per la stima. Il metodo Bayesiano verrà introdotto e verranno illustrate le principali procedure inferenziali.

### PREREQUISITI

Conoscenze di statistica e matematica di base.

### OBIETTIVI FORMATIVI

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

#### **Conoscenze e comprensione.**

Acquisire una conoscenza dei principali metodi matematici e statistici per l'inferenza mediante l'utilizzo di testi di riferimento, materiale didattico ed esercitazioni in aula. Inoltre saranno illustrati temi di avanguardia quali l'uso di metodi quantitativi nell'ambito del data science per le scienze umane e sociali.

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

Applicare e discutere gli strumenti più avanzati della statistica e della data science nell'ambito dell'analisi dei dati sperimentali.

#### **Autonomia di giudizio.**

Interpretare i dati in modo da formulare conclusioni autonome in ambito quantitativo.

#### **Abilità comunicative.**

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

#### **Capacità di apprendimento.**

Aver sviluppato le competenze necessarie per intraprendere studi successivi con un alto grado di autonomia.

### METODI DIDATTICI

Le lezioni tradizionali di tipo frontale verranno accompagnate da esercitazioni su argomenti che saranno oggetto della prova d'esame.

---

## MODALITA' D'ESAME

Esame scritto con quesiti a risposta multipla ed esercizi. Eventuale progetto.

---

## PROGRAMMA ESTESO

### **Introduzione alla scienza statistica**

Statistica: descrizioni e inferenze;  
Tipi di dati e variabili;  
Raccolta dei dati e casualizzazione;  
Statistiche descrittive: sintetizzare i dati univariati e multivariati.

### **Distribuzioni di probabilità**

Introduzione alla probabilità  
Variabili aleatorie e distribuzioni di probabilità  
Valori attesi delle variabili aleatorie  
Distribuzioni di probabilità discrete  
Distribuzioni di probabilità continue  
Generazione numeri casuali  
Distribuzioni congiunte, condizionate e indipendenza  
Correlazione tra variabili aleatorie

### **Distribuzioni campionarie**

Le distribuzioni campionarie: distribuzioni di probabilità delle statistiche  
Teorema del limite centrale: distribuzione campionaria normale per campioni di grandi dimensioni

### **Inferenza statistica: stima**

Stime puntuali e intervalli di confidenza  
Funzione di verosimiglianza e stima di massima verosimiglianza  
Determinare la dimensione del campione prima di raccogliere i dati  
Ricampionamento computazionale e intervalli di confidenza bootstrap  
L'approccio bayesiano all'inferenza statistica

### **Inferenza statistica: test di significatività**

Test del rapporto di verosimiglianza e intervalli di confidenza  
Test non parametrici

### **Classificazione e clustering**

Classificazione: analisi discriminante lineare e alberi  
Analisi dei cluster

---

## TESTI DI RIFERIMENTO

A. Agresti, M. Kateri (2022) Statistica per data scientist, EGEA