# Statistica applicata e Big Data

## Prof. Enrico Lazzaretto; Prof. Maria Lucia Pace

***OBIETTIVI DEL CORSO E RISULTATI DI APPRENDIMENTO ATTESI***

L’obiettivo della prima parte del corso è fornire agli studenti gli strumenti statistici fondamentali per l’analisi dei dati con il supporto del software R per l’analisi quantitativa.

Al termine di questa prima parte del corso lo studente sarà in grado di:

* applicare i metodi fondamentali dell'inferenza statistica (stima puntuale ed intervallare e verifica di ipotesi parametrica e non);
* gestire ed organizzare i dati mediante il software R.
* analisi descrittiva e bivariata attraverso l’impiego di R.

Obiettivo della seconda parte del corso è fornire agli studenti la capacità di utilizzare strumenti statistici per affrontare e risolvere problemi reali, approfondendo alcuni argomenti che trovano applicazione nel contesto dei Big Data.

Al termine di questa seconda parte del corso lo studente sarà in grado di:

* stimare modelli di regressione lineare e logistica tramite l’utilizzo del software statistico R
* effettuare l’analisi dei residui per valutare la bontà del modello di regressione applicato
* utilizzare i metodi classici di riduzione sintesi e semplificazione dei dati (analisi delle componenti principali analisi fattoriale).

***PROGRAMMA DEL CORSO***

Il programma del corso è suddiviso in 6 moduli che affrontano le seguenti tematiche:

1. **Accenni di inferenza e probabilità**: Metodi fondamentali dell'inferenza statistica: variabili casuali; stima e test statistici
2. **Introduzione ad R come ambiente e strumento di analisi ed elaborazione dei dati**
3. **Regressione lineare semplice:** Introduzione alle tecniche statistiche utilizzate per modellare la dipendenza tra due variabili e i test sui residui
4. **Regressione lineare multipla:** Introduzione alle tecniche statistiche utilizzate per modellare la dipendenza tra più variabili
5. **R avanzato con focus su regressioni multiple**
6. **Regressione logistica:** Comprendere la specificazione di modelli che impiegano le variabili qualitative come variabile di risposta

***BIBLIOGRAFIA***

Espa, G. Micciolo, R. Analisi esplorativa dei dati con R Apogeo, 2012.

S. Borra-A. Di Ciaccio, Statistica 3/ed - Metodologie per le scienze economiche e sociali, McGraw-Hill, 2014.

D.M. Levine-T.C. Krehbiel-M.L. Berenson, Statistica, Pearson Italia, 7a Edizione, Con Mylab E Etext, Milano, 2018. (cap.3, 5-9).

***DIDATTICA DEL CORSO***

Il corso prevede l’alternanza di lezioni in aula e attività da svolgere individualmente e/o in gruppo online.

A distanza è richiesto di studiare gli argomenti presentati nelle video-lezioni ed il materiale di approfondimento. La partecipazione ai webinar (esercitativi e di feedback) risulta importante ai fini dell’apprendimento.

***AVVERTENZE E PREREQUISITI***

Il corso presuppone una conoscenza delle nozioni di base di statistica (descrizione dei dati e teoria della probabilità). Si consigliano al riguardo il seguente testo:

- D.M. Levine-T.C. Krehbiel-M.L. Berenson, Statistica, Pearson Italia, 7a Edizione, Con Mylab E Etext, Milano, 2018. (cap. 2-4)

- S. Borra - A. Di Ciaccio, Statistica. Metodologie per le scienze economiche e sociali, Mc-Graw-Hill, Milano, 2014, 3a ed. [Capp. 1-4, 6, 16]

Le lezioni introduttive sono comunque dedicate al consolidamento di alcuni concetti fondamentali.

*Nel caso in cui la situazione sanitaria relativa alla pandemia di Covid-19 non dovesse consentire la didattica in presenza, sarà garantita l’erogazione dell’insegnamento in distance learning con modalità che verranno comunicate in tempo utile agli studenti*