UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

# SYLLABUS DEL CORSO

# Foundations of Probability and Statistics

**2122-1-F9101Q002**

## Obiettivi formativi

Il corso si propone di introdurre i concetti e i metodi di statistica descrittiva, calcolo delle probabilità ed inferenza statistica (stima, test, modelli) sia dal punto di vista teorico che applicativo attraverso l'utilizzo di software (R), con particolare attenzione agli argomenti rilevanti per i corsi più avanzati di datamining e machine learning.

Alla fine del corso lo studente ha la possibilità di capire l'induzione statistica e le implicazioni sulla popolazione derivanti dallo studio di un campione di dati, essendo in grado di sperimentare ed applicare le conoscenze acquisite su dataset reali.

## Contenuti sintetici

Metodi di statistica descrittiva, calcolo delle probabilità ed inferenza statistica (stima, test, modelli)

## Programma esteso

- Introduzione al trattamento dati con R
- Analisi descrittiva: distribuzioni, rappresentazioni grafiche, indici di posizione e di variabilità
- Calcolo delle probabilità: concezioni probabilistiche, probabilità sugli eventi, teorema di Bayes, variabili aleatorie e distribuzioni di probabilità, distribuzioni notevoli, enunciazioni LLN e CLT
- Inferenza statistica: la logica del campionamento probabilistico. Stimatori e loro proprietà. Stima puntuale

(Media varianza e proporzione). Cenni sugli stimatori di massima verosimiglianza.
- Stima intervallare: concetto di confidenza, intervalli di confidenza, casi particolari sulla media e la varianza
- Verifica di ipotesi: Il concetto di statistica test. La significatività e la potenza del test. Test sulla media, varianza, proporzione, sulla differenza tra medie, test di indipendenza.
- Regressione lineare semplice: metodo di stima dei minimi quadrati, misure di adeguatezza del modello, distribuzione campionaria degli stimatori OLS, test d'ipotesi e intervalli di confidenza per i coefficienti di regressione, analisi della varianza, outliers e osservazioni influenti

## Prerequisiti

Nessuno.

## Metodi didattici

Lezioni frontali e laboratorio.

## Modalità di verifica dell'apprendimento

PROVA SCRITTA: consiste in una batteria di domande a risposta multipla e/ domande aperte sulla TEORIA affrontata a lezione. (punteggio massimo 28)

PROGETTO: Sviluppo di un progetto originale a partire da una semplice idea o dall'analisi di un caso esistente. Lavoro applicativo da svolgere autonomamente o in gruppo (al max 3 studenti) su un dataset scelto dallo studente (con R) su cui applicare i principali argomenti svolti a lezione (analisi descrittive, test, modello lineare). Il progetto va inviato al docente almeno 10 gg prima della prova, sarà valido per l'intero anno accademico. (punteggio massimo 6).

Non sono previste prove in itinere

## Testi di riferimento

Materiale del docente a disposizione degli studenti sull'e-learning

Testo a scelta:

- A.M. Mood, F.A. Graybill, D.C. Boes, Introduzione alla statistica
- G. Cicchitelli, P. D'Urso M. Minozzo, Statistica: principi e metodi
- P.S. Mann, Introductory Statistics
- M. Lavine, Introduction to Statistical Thought

## Periodo di erogazione dell'insegnamento

Primo semestre (Settembre - Novembre)

## Lingua di insegnamento

Italiano