**Statistica (Statistical learning)**

Prof. Silvia Angela Osmetti

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

La società moderna è caratterizzata da una massa crescente di informazioni e dati (*big data*). Il corso si propone di fornire agli studenti strumenti quantitativi moderni di *Statistica* e *Data Science* per *modellizzare* e *comprendere* dati di natura complessa sempre più presenti non solo in campo scientifico ma anche in ambito economico-aziendale. Particolare enfasi sarà rivolta agli aspetti *computazionali* - mediante l’uso del linguaggio gratuito ed open-source *R* - al fine di sviluppare concrete capacità di analisi empirica su casi concreti.

Al termine dell’insegnamento, lo studente dovrà:

1. aver acquisito la conoscenza e la capacità di comprensione delle principali tecniche statistiche presentate;
2. essere in grado di applicare i metodi statistici presentati alla risoluzione di problemi reali in ambito economico, finanziario e sociale anche attraverso opportuni software statistici;
3. aver acquisito capacità di pensiero critico che gli consenta di interpretare i risultati ottenuti estrapolando informazioni qualitative dai dati quantitativi, di fornire giudizi indipendenti sulle performance dei modelli statistici nelle applicazioni a problemi reali e di selezionare il modello più adeguato tra quelli studiati;
4. essere in grado di comunicare in modo chiaro e accurato i risultati di un'analisi statistica e motivare quantitativamente le scelte che sono state fatte ad un pubblico specializzato e non specializzato;
5. aver sviluppato buone capacità di apprendimento che permettano di individuare collegamenti tra gli argomenti sviluppati nel corso e altri argomenti affrontati nell’ambito del proprio piano di studi; aver acquisito buone competenze richieste dalle posizioni professionali coinvolte nella gestione dei dati, capacità di ragionamento rigoroso e assunzione di decisioni data-driven.

***PROGRAMMA DEL CORSO***

1. Introduzione a *Statistical learning* e al linguaggio di programmazione R.

2. Regressione lineare semplice e multipla.

3. Classificazione.

4. Metodi bootstrap

5. Metodi di ricampionamento e scelta del modello.

6. Metodi statistici basati su alberi decisionali.

7. Metodi non supervisionati: analisi delle componenti principali e cluster analysis.

***BIBLIOGRAFIA[[1]](#footnote-1)***

J. Gareth-D. Witten-T. Hastie-R. Tibshirani, *An Introduction to Statistical Learning,* Springer, 2013.

Materiale del corso sulla piattaforma *Blackboard*.

Software statistico R *http://www.r-project.org/*

***DIDATTICA DEL CORSO***

Lezioni frontali (60 ore) che prevedono la presentazione dei metodi statistici, l’analisi di programmi in *R* e esercitazioni pratiche in aula informatica.

***METODO E CRITERI DI VALUTAZIONE***

L’esame è basato su una prova scritta da svolgersi in 2 ore: l’esame contiene domande a risposta aperta su metodi statistici presentati a lezione (punti dall’1 al 7 del programma), interpretazione di programmi in *R*, interpretazione e lettura di risultati.

Per gli studenti frequentanti la prova scritta può essere sostituita da due prove parziali che contribuiscono equamente alla determinazione del voto finale (la prova intermedia svolta durante la settimana di sospensione delle lezioni del secondo semestre e la prova di completamento nella sessione estiva). La prova intermedia riguarda i punti 1,2,3 e 4 del programma. Chi supera la prova intermedia deve sostenere il completamento sui punti 5,6 e 7 del programma in uno degli appelli ufficiali della sessione estiva. Indicazioni dettagliate riguardo alle modalità di svolgimento delle suddette prove saranno rese disponibili in Blackboard.

L’esame è volto a valutare innanzitutto capacità di ragionamento e rigore analitico sui temi oggetto del corso, nonché proprietà di linguaggio e abilità comunicative.

***AVVERTENZE E PREREQUISITI***

Gli studenti devono possedere nozioni di analisi dei dati, probabilità e inferenza al livello dei corsi di Statistica di base impartiti in una facoltà di Economia: si cosigliano ad esempio gli argomenti dei corsi di *Statistica (analisi dei dati e probabilità) e Statistica Applicata*, oppure di *Statistics* *e* *Statistical inference* impartiti presso questo Ateneo.

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*.*

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)