

Statistical Learning and Prediction from Data

Direzione e Consulenza d'Impresa – a.a. 2024/25

Fabio Demaria

Descrizione del corso e obiettivi

Molte aree dell'economia, dell'industria e della finanza si occupano dello studio di fenomeni attraverso l'analisi di grandi insiemi di dati, la cui misurazione è caratterizzata da incertezza. Questo corso introduce i principi fondamentali dell'analisi dei dati da un punto di vista statistico e introduce lo studente alla pratica di tali principi.

L'enfasi è posta sull'analisi esplorativa dei dati, sulla costruzione di modelli statistici bivariati e multivariati e sulla corretta interpretazione dei risultati. I principali concetti vengono presentati con rigore durante il corso. Inoltre, particolare attenzione sarà posta sullo sviluppo della capacità di risolvere problemi applicati a dati reali.

Prerequisiti

Per accedere al corso, occorre avere superato un corso di base di analisi matematica ed uno di statistica a livello di laurea triennale.

Contenuti del corso

Per ogni Credito Formativo Universitario (CFU) sono previste 7 ore di didattica frontale e 18 ore di studio autonomo di studenti e studentesse. Pertanto, 1 CFU equivale ad un impegno standard di studio pari a 25 ore.

1 CFU (7 ore)

1. Concetti preliminari
2. Misurazione e scale
3. Raccolta di dati primari e progettazione di sondaggi

1 CFU (7 ore)

4. Indagine campionaria
5. Tecniche di campionamento e pesi campionari

1 CFU (7 ore)

6. Preparazione dei dati, z-score e outlier
7. Test di confronto tra due medie (Test t) e ANOVA a una via

1 CFU (7 ore)

8. Misure di associazione (Chi-quadrato)
9. Covarianza, correlazione e correlazione per ranghi

1 CFU (7 ore)

10. Regressione lineare bivariata e multipla
11. Regressione logistica

1 CFU (7 ore)

12. Analisi fattoriale e Analisi delle Componenti Principali
13. Tecniche di clustering: gerarchico e non gerarchico

Metodi didattici

L'insegnamento si svolge in presenza ed è erogato in italiano. Il metodo didattico si basa su: (a) lezioni frontali con contenuti sia teorici che applicati, supportati da materiale didattico (slide, esercizi, ecc.) e registrazioni delle lezioni; (b) esercitazioni e dimostrazioni dell'uso del software statistico SPSS per l'analisi di dataset, finalizzate allo sviluppo della capacità di applicare le conoscenze acquisite.

Materiali didattici integrativi (slide, esercizi, esempi di compiti d'esame, ecc.) sono reperibili sulle pagine MS Teams/Moodle del corso.

In linea con le decisioni del Corso di Studio, il materiale relativo alle lezioni e le videoregistrazioni saranno resi disponibili a metà del corso. Lo streaming non è invece consentito, in conformità con le disposizioni relative all'erogazione della didattica di UNIMORE.

Testi di riferimento

- [ENG] Mario Mazzocchi (2008). Statistics for Marketing and Consumer Research. London: SAGE. Chapters: 1, 3, 4, 5, 7, 8, 9, 10, 12.
- [ITA] Zani S., Cerioli A. (2007). Analisi dei dati e data mining per le decisioni aziendali. Giuffrè Editore (ISBN: 8814136955). Chapters: 1, 2, 3, 4, 5, 6, 8, 9.
- Slides ed esercizi scaricabili dalle pagine MS Teams/Moodle del corso.

Esame finale

L'esame è una prova scritta che si svolgerà secondo il calendario ufficiale e sarà articolato come segue:

- Prova **intermedia** (facoltativa) della durata di 45 minuti, sugli argomenti della prima parte del corso.
- Prova **finale** (nella sessione estiva ordinaria):
 - Per coloro che **superano** la prova intermedia, la prova finale riguarderà gli argomenti della seconda parte del corso e avrà una durata di 45 minuti.
Gli studenti **frequentanti** che hanno **superato** la prova intermedia possono scegliere di svolgere la prova finale con le seguenti modalità: 20% valutazioni in itinere nella seconda parte del corso ¹; 30% progetto finale di gruppo (max. 3 persone), da esporre il giorno della prova finale.
 - Per coloro che **non sostengono** o **non superano** la prova intermedia, l'esame consisterà in un'unica prova su tutti gli argomenti del corso, della durata di 90 minuti.

¹I test si svolgeranno in aula su base settimanale. Il punteggio finale corrisponderà alla media aritmetica dei punti totalizzati nei test, escludendo il punteggio più basso.

Le prove d'esame sono scritte e prevedono domande a risposta multipla, domande aperte ed esercizi pratici di analisi empirica. Gli indicatori di valutazione sono i seguenti: capacità di applicare le conoscenze teoriche alla risoluzione dei problemi; padronanza del linguaggio statistico; capacità di discutere gli argomenti e i risultati dell'analisi empirica; capacità di approfondire gli argomenti; pensiero critico. Il voto finale sarà calcolato come la media delle valutazioni della prova intermedia e della prova finale per chi ha sostenuto entrambe, oppure come una valutazione unica per chi svolge l'esame interamente in un'unica sessione. Il voto è espresso in trentesimi e l'esame si considera superato con un punteggio minimo di 18/30 (equivalente a una preparazione sufficiente). Gli esiti saranno comunicati entro e non oltre due settimane dalla prova scritta e pubblicati sul portale Esse3.

Materiale necessario per sostenere l'esame: calcolatrice scientifica, tavole statistiche e formulario.

Obiettivi di apprendimento

1. Conoscenza e capacità di comprensione: tramite lezioni in aula e studio individuale Identificazione delle fasi del processo di ricerca, degli strumenti di misura e di raccolta dei dati. Conoscenza delle tecniche di campionamento. Conoscenza dei metodi statistici utili a misurare l'eventuale relazione esistente tra due o più variabili. Conoscenza dei principali metodi statistici multivariati per la riduzione dimensionale di una matrice di dati (analisi delle componenti principali e tecniche di clustering)
2. Capacità di applicare conoscenza e comprensione: Tramite svolgimento di homeworks e analisi di dati empirici, sviluppo di capacità: Capacità di effettuare un'indagine statistica in tutte le sue fasi e utilizzo di queste conoscenze in problemi reali in ambito economico, finanziario e sociale. Capacità di usare un linguaggio statistico per studiare e formalizzare tali problemi. Capacità di scelta della metodologia appropriata a seconda del problema, identificazione delle variabili e delle informazioni che si vogliono ottenere. Capacità di interpretare i risultati delle elaborazioni in maniera critica e trarre conclusioni. Capacità di comunicare i risultati, sotto forma di report o grafici.
3. Autonomia di giudizio: Attitudine ad un approccio metodologico che conduca a verificare tramite argomentazioni rigorose di teoria statistica ed economica le affermazioni e i metodi presentati. Capacità di autovalutazione delle proprie competenze ed abilità.
4. Abilità comunicative: Capacità di affrontare in modo puntuale e coerente la teoria statistica e di argomentare con precisione l'analisi empirica. A conclusione del corso lo studente presenterà in forma scritta i risultati e le interpretazioni di analisi empiriche ottenute anche con l'ausilio di supporto digitale.
5. Capacità di apprendimento: Acquisizione delle conoscenze di tipo statistico ed economico come proprio patrimonio, da poter utilizzare per analisi individuali su dati empirici.

Obiettivi sviluppo sostenibile

Questo insegnamento concorre alla realizzazione degli obiettivi ONU dell'Agenda 2030 per lo Sviluppo Sostenibile.

- 4 – Istruzione di qualità
- 8 – Lavoro dignitoso e crescita economica
- 9 – Industria, innovazione e infrastrutture