

Statistica e Analisi dei Dati

Docente: Matteo Caruso

Programma a.a. 2022/2023

Struttura del corso e nozioni introduttive

Giovedì 30 e venerdì 31 marzo

Il corso verrà svolto su 8 settimane, con ultima lezione venerdì 19 maggio e una prova intermedia verosimilmente nell'ultima settimana di aprile (quinta settimana). Verranno alternate lezioni frontali con esercitazioni tramite il software Stata. La prima lezione sarà concentrata un'introduzione alla statistica in maniera molto generale, definizione di statistica descrittiva e inferenza e come esse siano applicate alle scienze sociali (Capitolo 1).

Distribuzioni statistiche e rappresentazioni grafiche

Giovedì 6 e venerdì 7 aprile

Nella seconda settimana, il corso introduce la statistica descrittiva, parlando di osservazioni, frequenza, variabili discrete e continue, e serie storiche (Capitolo 2). Focus particolare sarà sulla rappresentazione grafica dei dati e la loro esposizione in maniera formale (Capitolo 3).

Media e Varianza

Giovedì 13 e venerdì 14 aprile

Nella terza settimana, si definiranno concetti base della statistica quali media, mediana, moda, quartili, varianza, deviazione standard, ed eterogeneità, utili per fare inferenza e proporre dei primi concetti fattuali sulla realtà che ci si appresta ad analizzare (Capitoli 4 e 5). Un focus particolare si darà all'indice di concentrazione di Gini e al suo utilizzo empirico nelle scienze sociali.

Distribuzioni di frequenza e numeri indici

Giovedì 20 e venerdì 21 aprile

La quarta settimana si focalizzerà sulla rappresentazione delle distribuzioni di frequenza (asimmetria, curtosi), sui grafici a scatola (c.d. *box-plot*) e ulteriori

forme di rappresentazione di dati tramite grafici (Capitolo 7). Oltre ciò, le lezioni forniranno gli strumenti per capire la funzione dei numeri indici e come utilizzarli nello studio dei dati e delle serie storiche (Capitolo 8).

Analisi distribuzioni doppie: dipendenza, regressione e correlazione

Giovedì 27 e venerdì 28 aprile

Nella quinta settimana, si partirà dallo studio delle tabelle di contingenza per arrivare a definire il concetto di dipendenza statistica e come misurarla (Capitolo 9). Successivamente, viene introdotto il concetto di regressione lineare semplice e di correlazione, con alcuni esempi pratici (Capitoli 10 e 11).

Probabilità, variabili casuali e distribuzioni di probabilità

Giovedì 4 e venerdì 5 maggio

La sesta settimana si focalizzerà sulla probabilità e sul concetto di variabile casuale e che rilevanza ha nella statistica (Capitoli 12 e 13). Oltre ciò, verranno brevemente introdotte alcune importanti distribuzioni di probabilità (e.g. Bernoulli, Binomiale, Poisson, ecc.) rilevanti per lo studio e l'analisi dei dati (Capitolo 14). Un breve accenno verrà dato alla legge dei grandi numeri e il teorema del limite centrale, con esempi pratici di approssimazione verso distribuzioni Normali (Capitolo 16).

Inferenza: popolazione, campione, stima puntuale e per intervalli

Giovedì 11 e venerdì 12 maggio

Nella settimana settimana, verrà introdotto il concetto di inferenza e il suo utilizzo nello studio di caratteristiche di una popolazione e l'importanza delle indagini tramite campioni casuali (Capitolo 17). Oltre ciò, parte della lezione parlerà dell'utilizzo di stime puntuali e intervallari per fare inferenza su parametri della popolazione (Capitoli 18 e 19).

Test di ipotesi e inferenza sul modello di regressione lineare

Giovedì 18 e venerdì 19 maggio

L'ottava e ultima settimana di corso introdurrà il concetto di test di ipotesi per valutare la significatività dei dati campionari (Capitoli 20 e 21), studiando anche come applicare il modello di regressione lineare nel campione casuale per fare inferenza su parametri della popolazione (Capitolo 23).