

Statistica I

Introduzione al corso

Tommaso Rigon

Università Milano-Bicocca

Anno Accademico 2020-2021

- **Email:** `tommaso.rigon@unimib.it`
- **Ricevimento:**
- **Pagina personale:** `https://tommasorigon.github.io`
- **Tutor:**

Organizzazione del corso

- **Pagina web del corso:** <https://tommasorigon.github.io/StatI/>
- Nella pagina potete trovare:
 - Ultimi avvisi
 - Calendario delle lezioni
 - Materiale didattico da scaricare (slide, esercizi, etc.)
 - Modalità d'esame
 - Altro

Modalità d'esame

Statistica: cenni storici

- Sebbene oggi abbia un significato molto diverso, la parola **statistica** deriva dal vocabolo italiano **Stato**.
- **Ghislini** nel 1589 indica la statistica come *“descrizione delle qualità che caratterizzano e degli elementi che compongono uno Stato”*.
- La **collezione di dati** circa la popolazione e l'economia cominciò nelle città di Venezia e Firenze durante il Rinascimento. Venivano ad esempio registrate le **nascite**, i **decessi**, i **matrimoni**, soprattutto in occasioni delle pestilenze.
- Ancora oggi alcuni enti (ad esempio ISTAT ed EUROSTAT), intendono la statistica come “disciplina dello Stato”. Nel linguaggio moderno, chiameremmo questa disciplina “**statistica ufficiale**” oppure anche “**statistica economica**”.

- Il **pestifero e contagioso morbo** di **Carlo Cipolla** contiene una ricostruzione storica di ciò che avvenne a Pistoia durante l'epidemia di peste del 1630–1631.



- Il marchese **Luigi Vettori**, mandato a Pistoia da Firenze, scrisse un meticoloso rapporto sull'epidemia, conteggiando i morti, gli infetti, ed i ricoverati al lazzaretto.
- Il rapporto del marchese è senz'altro da considerarsi un'analisi di **statistica descrittiva** ante litteram!

Verso la statistica moderna

- Il ruolo della statistica cambiò notevolmente a partire dalla fine del 1800, anche grazie allo sviluppo del **calcolo delle probabilità** e del **gioco d'azzardo**.
- La **Royal Statistical Society** (RSS) venne fondata nel 1834, mentre l'**American Statistical Association** (ASA) nel 1839. Il primo dipartimento di statistica al mondo venne fondato nel 1911 all'University College di Londra (UCL).
- Nella prima metà del '900 avvenne la **prima rivoluzione** della statistica. Furono gettate le fondamenta della teoria della statistica moderna.
- A partire dagli anni '70 e '80 del secolo scorso è cominciata una **seconda rivoluzione**, grazie all'immensa potenza di calcolo e disponibilità di dati derivanti dallo **sviluppo informatico**.
- Questa fase rivoluzionaria è ancora in atto. Nel contesto moderno, la statistica viene anche chiamata **data science**.

La percezione della statistica

- *“Ci sono tre tipologie di menzogne: le bugie, le maledette bugie e le statistiche”.* Pronunciata da **Mark Twain**.
- *“Come i sogni, le statistiche sono una forma di appagamento”.* **Jean Baudrillard**.
- A volte purtroppo la statistica è accolta con sospetto.
- La statistica e i dati, di per sé, non mentono mai. Chi mente sono due categorie di statistici: quelli **disonesti** e quelli **incompetenti**.
- Sfortunatamente, anche quando un bravo statistico svolge correttamente il suo lavoro, questo potrebbe venire **percepito** come “falso” se mancano le competenze quantitative di base.

La statistica, Trilussa (Carlo Alberto Salustri)

*Sai ched'è la statistica? È na' cosa
che serve pe fà un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.*

*Ma pè me la statistica curiosa
è dove c'entra la percentuale,
pè via che, lì, la media è sempre eguale
puro co' la persona bisognosa.*

*Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:*

*e, se nun entra nelle spese tue,
t'entra ne la statistica lo stesso
perch'è c'è un antro che ne magna due.*

La statistica nella cultura popolare



- La figura qui sopra illustra un esempio di **statistico incompetente**.
- La poesia di Trilussa evidenzia invece lo scetticismo di chi è sprovvisto delle nozioni statistiche di base.

In risposta alle precedenti provocazioni

- Fortunatamente, non tutti sono dello stesso stesso parere. La statistica è in effetti **alla base del metodo scientifico**.
- *"E' semplice mentire con la statistica, ma è molto più semplice mentire senza di essa"*. **Frederick Mosteller**.
- *"La statistica è la grammatica della scienza"*. **Karl Pearson**.
- *"Coloro che non conoscono la statistica sono condannati a reinventarla"*. **Bradley Efron**.
- *"Il mestiere sexy dei prossimi 10 anni sarà lo statistico. Le persone pensano che stia scherzando, ma chi avrebbe immaginato che gli informatici sarebbero stati sexy negli anni '90?"* **Hal Varian**, chief economist di Google.

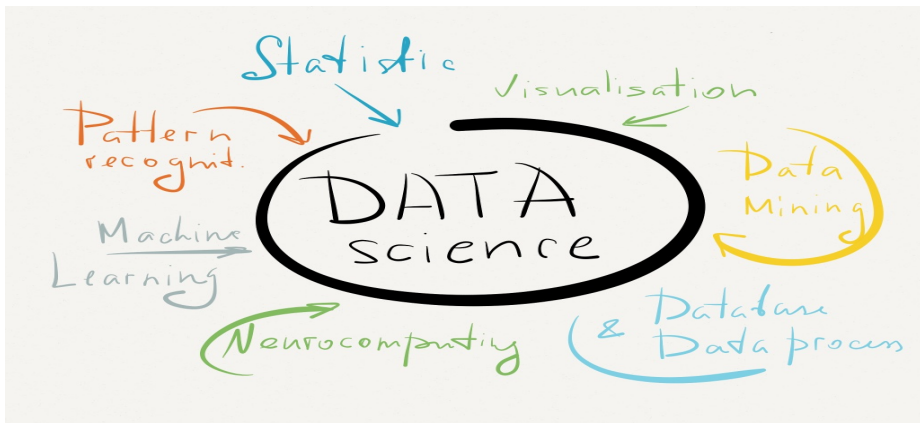
La definizione di statistica

- La statistica non è più la “scienza dello Stato”. Non esiste una definizione univocamente accettata, ma le varie proposte hanno spesso un ampio respiro.

Alcune definizioni

- “La statistica riguarda **ciò che si può imparare dai dati**”. **Anthony Davison**.
- “La statistica è essenzialmente lo **studio dell'incertezza**. Il ruolo degli statistici è di **assistere** i lavoratori di altri campi, o i propri clienti, i quali incontrano l'incertezza nel loro lavoro.” **Dennis Lindley**.
- “La statistica è la tecnologia necessaria per **estrarre significato dai dati** e per **gestire l'incertezza**”. **David Hand**.
- “La statistica è la **scienza delle decisioni** in condizioni di incertezza”. **Domenico Piccolo**.

Data science: la statistica moderna



Una scienza ancillare

- *“L'aspetto migliore dell'essere uno statistico è che puoi giocare nel giardino di chiunque”.* John Tukey.
- La statistica è una scienza **di supporto** a tutte le altre discipline, ad esempio:
 - Biologia
 - Medicina
 - Epidemiologia
 - Scienze sociali & Demografia
 - Psicologia
 - Marketing
 - Finanza
- La statistica viene spesso chiamata una **scienza ancillare**, perché è presente trasversalmente in quasi tutte le discipline scientifiche.
- Al tempo stesso, la statistica è una **disciplina a sé stante**, con le sue regole ed il suo linguaggio.

Statistica descrittiva ed inferenziale

- La statistica viene spesso suddivisa in due macro-argomenti.

Statistica descrittiva

- **Descrive** e **riassume** la realtà.
 - Creazione di indici, tra cui: media aritmetica, mediana, moda, varianza, etc.
 - Visualizzazione dei dati tramite ad esempio istogrammi, boxplot, etc.
- **Quantifica** l'incertezza tramite strumenti algebrici, geometrici e grafici.
- **Esplora** i dati.

Statistica inferenziale

- **Quantifica** l'incertezza tramite il **calcolo delle probabilità**.
- **Stima** le quantità ignote di un modello statistico.
- **Prevede** fenomeni futuri fornendo margini di incertezza.

Contenuto del corso

- Statistica I è un corso di **statistica descrittiva**. Non faremo quindi (quasi) nessuna menzione alla probabilità.
- La statistica descrittiva e quella inferenziale **non sono argomenti separati**. Molti concetti sono comuni, sebbene vengano a volte presentati in modi diversi.
- Durante il corso i concetti teorici di statistica descrittiva verranno sempre accompagnati da un'**applicazione**.
- I contesti applicativi saranno molto diversi tra loro, a sottolineare il fatto che la statistica sia una **scienza trasversale**.

Il ruolo dello statistico

- Ogni problema ha caratteristiche peculiari, con conoscenze pregresse diverse. In ogni problema statistico ci sono **scelte** da fare.
- Uno **statistico** è quindi come un **vecchio artigiano**: tutti i pezzi sono unici, i risultati dipendono dalla sua capacità. Successi ed insuccessi sono frutto delle sue scelte.
- **Non esistono ricette** e procedure da applicare in automatico. Guardate ai contenuti dei corsi (di Statistica) solo come ad esempi di quello che si può fare.
- Del resto, perché dovrebbero esistere delle lauree in Statistica se gli statistici fossero sostituibili con delle macchine?
- Questo corso mostra idee, tecniche e strategie che si sono rivelate utili in una gamma di problemi reali. Ma i problemi che voi affronterete in futuro non saranno gli stessi!